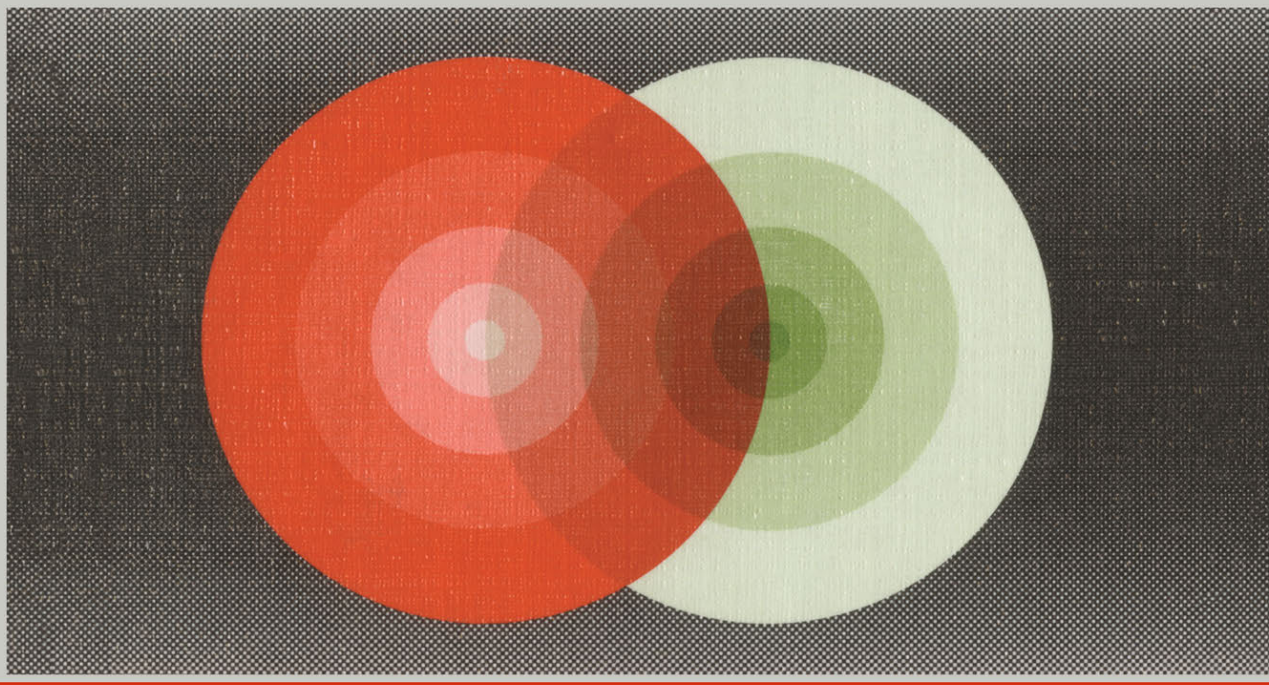
Maximum Entropy and Bayesian Methods

Edited by

J. Skilling

Springer-Science+Business Media, B.V.



Fundamental Theories of Physics

Maximum Entropy and Bayesian Methods

Fundamental Theories of Physics

An International Book Series on The Fundamental Theories of Physics: Their Clarification, Development and Application

Maximum Entropy and Bayesian Methods

Cambridge, England, 1988

edited by

J. Skilling
Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Cambridge, U.K.

Springer-Science+Business Media, B.V.

Library of Congress Cataloging in Publication Data

Maximum Entropy Workshop (8th : 1988 : St. John's College)
Maximum entropy and Bayesian methods, Cambridge, U.K., 1988 /
edited by J. Skilling.
p. cm. -- (Fundamental theories of physics)
Includes index.
Proceedings of the 8th MaxEnt Workshop held at St. John's College,
Cambridge, England, August 1-5, 1988.

1. Entropy (Information theory)--Congresses. 2. Bayesian
statistical decision theory--Congresses. I. Skilling, J. (John)
II. Title. III. Series.
Q370.M385 1988
001.53'9--dc19 89-2480

---

printed on acid free paper

This work relates to Department of Navy Grant N00014–88–J–1126 issued by the Office of Naval Research. The United States Government has a royalty-free license throughout the world in all copyrightable material contained herein."

ISBN 978-90-481-4044-2 ISBN 978-94-015-7860-8 (eBook)
DOI 10.1007/978-94-015-7860-8
Softcover reprint of the hardcover 1st edition 1989

Originally published by Kluwer Academic Publishers in 1989.

Dedication

To the ideal of rational inference

PREFACE

The "8th MaxEnt Workshop", to give it its short name, was held in St. John's College, Cambridge, England, on August 1 - 5, 1988, and this Volume of 55 papers records the Proceedings.

History repeats itself in many ways. All ancient civilisations evolved some core of basic practical mathematics, but it was the Greeks who insisted upon superior intellectual standards. It was the Greeks who invented the world of axioms and theorems. They invented and formalised the central idea of logical proof, so that assent to axioms perforce requires assent to their consequences, however long the chain of reasoning involved. Conversely, any attack on the consequences becomes an attack upon the axioms, which are usually much simpler to discuss. This power and beauty swept cruder mathematics aside for ever.

An echo of this occurs in our own century. We live in a complicated world, and our procedures for learning from observations are codified as the subject of statistics. Often enough, we are concerned with difficult problems, and a variety of more or less *ad hoc* practical techniques has evolved to deal with them. Yet there is an inner logic to the practice of inference, which leads inevitably to the use of quantified probabilities, manipulated by Bayes' theorem and assigned by the principle of maximum entropy (MaxEnt). Those who live by this logic are called Bayesians. Because of their inner certainty of methodology, they are sometimes perceived as religious fundamentalists - but that does not in itself mean that they are wrong.

The Bayesian/MaxEnt church is alive and well, and has its own calendar of saints (and devils). Foreshadowed by the MIT meeting of 1978, the first formal assembly was held in Laramie, Wyoming in June 1981. It was my particular pleasure to attend that first "Workshop on Maximum Entropy and Bayesian Methods in Applied Statistics", organised by Ray Smith and Tom Grandy, who have since become lasting friends of mine. MaxEnt was then beginning to grow beyond the confines of statistical thermodynamics, where it had enjoyed a certain degree of protection afforded by the abstract nature of the subject, by the conveniently large value of Avogadro's number, and by the demonstrable success of its predictions. Bayes' theorem, likewise, was beginning to break free of the suffocating weight of "orthodox" statistics, substantially aided by the brilliant logic and polemic of Edwin Jaynes. Each summer since 1981 has seen a further meeting in the series. Each has been notable for some new inspiration and application, for the particular strength of rational thought is that it works. More and more quickly, inference problems in all sorts of disciplines are being brought within the purview of Bayesian/MaxEnt analysis.

The 1988 meeting, held in St. John's College, Cambridge, had a particularly appropriate venue. Sir Harold Jeffreys, who

cared deeply about rational inference throughout a long working life, is the Senior Fellow. Prof. Edwin T. Jaynes, whose influence on the subject has been so profound, is also connected with St. John's, having been Overseas Fellow in 1983/4. On a lesser plane, my own introduction to MaxEnt was a lunch-time conversation in College with my mentor, friend and colleague, Steve Gull.

Being the first of the workshops to be held in Europe, this meeting attracted over 100 delegates, from industry and from defence establishments as well as from academia. A central topic such as inference can be expected to touch a number of other subjects, but even the organisers were surprised by the variety of topics which were offered and presented, from philosophy to floods, from biology to astronomy, and with references ranging from New Left Publications to Acta Crystallographica.

Profound thanks are due to our financial sponsors, who provided the funds needed to invite distinguished overseas speakers whilst keeping the fees low enough for the academic pocket. The United States Navy Office of Naval Research maintained its valued connection with the workshop series through its grant N00014-88-J-1126, and industrial support was provided by E.I. DuPont Company Central Research and Development, ICI Chemicals and Polymers Group, Glaxo Group Research Limited, British Petroleum plc, and Maximum Entropy Data Consultants Limited. Thanks are also due to St. John's College, which provided such appropriate and attractive facilities, and whose staff were unfailingly generous with their time and effort. Not least, I wish to thank in particular my wife and son, Jennifer and Martin Skilling, for their secretarial and organisational help, which contributed so much to the smooth running of the meeting. Thank you, all.

The authors of the papers published here also deserve my editorial thanks for producing their papers so well and so promptly. In the interests of quick publication, the workshop is continuing the recent practice of using camera-ready copy. Because interest continues to grow, the workshops are currently being formally organised on a continuing basis, with a permanent organising committee, and their Proceedings are henceforward to be published by Kluwer Academic Publishers under the generic title

"Maximum Entropy and Bayesian Methods (location) (year)".

It is hoped that each successive volume will continue to capture something of the excitement and vitality of current research.

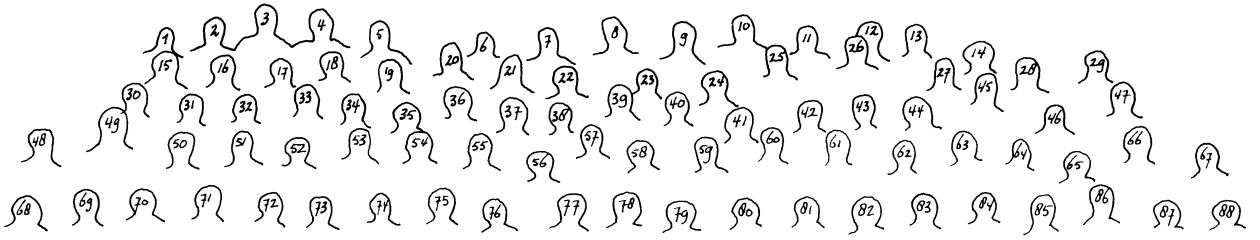
Lastly, I wish as Editor to dedicate this Volume, not to any particular individual, but to that transcending ideal to which we try to aspire - the ideal of rational inference.

St. John's College
Cambridge
January 1989

John Skilling

CONTENTS

Photograph courtesy of W. Eaden Lilley, Cambridge



1.Robert Papoular 2. Alastair Livesey 3. Tom Marsh 4. David Morrison 5. Paul Jowitt
6. Richard Bryan 7. F.Fröhner 8. Chris Gilmore 9. Keith Henderson 10. R.B.Hicks
11. Louis Roemer 12. Colin Bannister 13. Robert Collins 14. Juana Sanchez.
15. Carlos Rodriguez 16. Stephen Wilkins 17. Andrew Templeman 18. Stig Steenstrup
19. Randall Barron 20. Martin Barth 21. Alan M.Thompson 22. C.J.Shelton
23. D.van Ormondt 24. Monika Sebold 25. John Close 26. John Richardson
27. David Drabold 28. E.R.Podolyak 29. Ernest Laue.
30. John Karkheck 31. Myron Tribus 32. Ken Packer 33. Ofer Lahav 34. Marc A.Delsuc
35. Birgit Meyer 36. H.A.Mayer-Hasselwander 37. Nailong Wu 38. E.L.Kosarev
39. David Hestenes 40. J.K.Elder 41. Michael K.-S.Tso 42. David Larner 43. Odet Pols
44. Dave Wilkinson 45. Mark Charter 46. Dursun Ustundag 47. John Burg.
48. John Dougherty 49. Chris Burrows 50. Ali Mohammad-Djafari 51. Sibusiso Sibisi
52. Andrew Strong 53. Jacob Bekenstein 54. Ron Canterna 55. Do Kester 56. A.Baruya
57. Romke Bontekoe 58. A.D.McLachlan 59. Colin Fox 60. R.T.Constable 61. N.A.Farrow
62. R.D.Levine 63. Y.Tikochinsky 64. G.J.Daniell 65. R.Huis 66. John Pendrell
67. S.Hildebrand.
68. Keith Horne 69. Steve Luttrell 70. Lee Schick 71. Gary Erickson
72. Larry Bretthorst 73. Tom Grandy 74. Ray Smith 75. Steve Gull 76. Martin Skilling
77. Jennifer Skilling 78 John Skilling 79. Edwin Jaynes 80. Rabinder N.Madan
81. Paul F.Fougere 82. George Klir 83. Peter Cheeseman 84. Anthony Garrett
85. Andrew Gelman 86. Matthew Self 87. Roger Balian 88. Lucien Preuss

CLEARING UP MYSTERIES - THE ORIGINAL GOAL

E. T. Jaynes
Wayman Crow Professor of Physics
Washington University, St. Louis MO 63130, USA

Abstract. We show how the character of a scientific theory depends on one's attitude toward probability. Many circumstances seem mysterious or paradoxical to one who thinks that probabilities are physically real things. But when we adopt the "Bayesian Inference" viewpoint of Harold Jeffreys, paradoxes often become simple platitudes and we have a more powerful tool for useful calculations. This is illustrated by three examples from widely different fields: diffusion in kinetic theory, the Einstein-Podolsky-Rosen (EPR) paradox in quantum theory, and the second law of thermodynamics in biology.

INTRODUCTORY REMARKS

Our group has the honour to be among the first to use this splendid new Fisher building with its 300 seat auditorium. But perhaps, at a meeting concerned with Bayesian inference, we should clarify which Fisher inspired that name.

St. John's College was founded in the year 1511, its foundress being the Lady Margaret Beaufort. John Fisher was then Chancellor of the University of Cambridge, and after her death he found himself obliged to make heroic efforts to ensure that her wishes were carried out. But for those efforts, made some 480 years ago, St. John's College would not exist today. Historians have suggested that, but for the efforts of John Fisher in holding things together through a turbulent period, the entire University of Cambridge might not exist today.

Although the terms "Bayesian" and "Maximum Entropy" appear prominently in the announcements of our meetings, our efforts are somewhat more general. Stated broadly, we are concerned with this: "What are the theoretically valid, and pragmatically useful, ways of applying probability theory in science?"

The new advances of concern to us flow from the recognition that, in almost all respects that matter, the correct answers were given here in St. John's College some fifty years ago, by Sir Harold Jeffreys. He stated the general philosophy of what scientific inference is, fully and correctly, for the first time; and then proceeded to carry both the mathematical theory and its practical implementation farther than anyone can believe today, who has not studied his works.

The ideas were subtle, and it required a long time for their merit to be appreciated; but we can take satisfaction in knowing that Sir Harold lived to see a younger generation of scientists eagerly reading, and profiting by, his work. In September 1983 I had a long, delightful conversation over tea with Sir Harold and Lady Jeffreys, and know how pleased they both were.

Important progress is now being made in many areas of science by adopting the viewpoint and extending the methods of Harold Jeffreys. Even those of us who were long since convinced of their theoretical merit are often astonished to discover the amount of numerical improvement over "orthodox" statistical methods, that they can yield when programmed into computers. It is hardly ever small except in trivial problems, and nontrivial cases have come up where they yield orders of magnitude better sensitivity and resolution in extracting information from data.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 1–27.

This means that in some areas, such as magnetic resonance, it is now possible to conduct quantitative study of phenomena which were not accessible to observation at all by the previously used Fourier transform methods of data analysis; old data which have been preserved may have a new lease on life. The technical details of this are to appear in the forthcoming book of G. L. Bretthorst (1988).

Even when the numerical improvement is small, the greater computational efficiency of the Jeffreys methods, which can reduce the dimensionality of a search algorithm by eliminating uninteresting parameters at the start, can mean the difference between what is feasible and what is not, with a given computer. As the complexity of our problems increases, so does the relative advantage of the Jeffreys methods; therefore we think that in the future they will become a practical necessity for all workers in the quantitative sciences.

How fitting it is that this meeting is being held back where these advances started. Our thanks to the Master and Council of St. John's College, who made it possible.

THE MOTIVATION

Probability theory is a versatile tool, which can serve many different purposes. The earliest signs of my own interest in the field involved not data analysis, but recognition that the Jeffreys viewpoint can clear up outstanding mysteries in theoretical physics, by raising our standards of logic. As James Clerk Maxwell wrote over 100 years ago and Harold Jeffreys quoted 50 years ago, probability theory is itself the true logic of science.

The recent emphasis on the data analysis aspect stems from the availability of computers and the failure of "orthodox" statistics to keep up with the needs of science. This created many opportunities for us, about which other speakers will have a great deal to say here. Also, as will be noted here by David Drabold, John Skilling, and others, the MAXENT algorithm has proved to be a powerful tool for theoretical calculations. But while pursuing these important applications we should not lose sight of the original goal, which is in a sense even more fundamental to science.

Therefore in this opening talk we want to point out a field ripe for exploration by giving three examples, from widely different areas, of how scientific mysteries are cleared up, and paradoxes become platitudes, when we adopt the Jeffreys viewpoint. Once the logic of it is seen, it becomes evident that there are many other mysteries, in all sciences, calling out for the same treatment.

The first example is a simple exercise in kinetic theory that has puzzled generations of physics students: how does one calculate a diffusion coefficient and not get zero? The second concerns the currently interesting "Einstein-Podolsky-Rosen paradox" and "Bell inequality" mysteries in quantum theory: do physical influences travel faster than light? The third reexamines the old mystery about whether thermodynamics applies to biology: does the high efficiency of our muscles violate the second law?

DIFFUSION

Think, for concreteness, of a solution of sugar in water, so dilute that each sugar molecule interacts constantly with the surrounding water, but almost never encounters another sugar molecule. At time $t = 0$ the sugar concentration varies with position according to a function $n(x,0)$. At a later time we expect that these variations will smooth out, and eventually $n(x,t)$ will tend to a uniform distribution.

Since sugar molecules -- or as we shall call them, "particles" -- are not created or destroyed, it seems natural to think that there must have been a diffusion current, or flux J(x,t) carrying them from the high density regions to the low, so that the change in density with time is accounted for by the conservation law:

$$\frac{\partial n}{\partial t} + \mathrm{div}(J) = 0 \quad . \tag{1}$$

Phenomenologically, Fick's law relates this to the density gradient:

$$J = - D \ \mathrm{grad}(n) \tag{2}$$

In the case of sugars, this is easy to measure by optical rotation. In Maxwell's great Encyclopaedia Brittanica article on diffusion he quotes the experimental result of Voit for sucrose: D = 3.65 E–05 square cm/sec.

Our present problem is: how do we calculate J(x,t) from first principles? Maxwell gave the simple kinetic theory of diffusion in gases, based on the idea of the mean free path. But in a liquid there is no mean free path. Maxwell, who died in 1879, never knew the general theoretical formula for the diffusion coefficient which we now seek, and which applies equally to gases, liquids, and solids.

Only with the work of Einstein in the first decade of this Century were the beginnings made in seeing how to deal with the problem, culminating finally in the correct formula for the diffusion coefficient. But Einstein had to work at it harder than we shall, because he did not have Harold Jeffreys around to show him how to use probability theory.

It would seem that, given where a particle is now, we should find its velocity v, and summing this over all particles in a small region would give the local flux J(x,t). However, the instantaneous velocity of a particle is fluctuating wildly, with a mean-square value given by the Rayleigh-Jeans equipartition law; and that is not the velocity we seek. Superposed on this rapidly fluctuating and reversing thermal velocity, of the order of 100 meters/sec, is a very much slower average drift velocity representing diffusion, which is our present interest.

Given where a particle is now, x(t), its average velocity over a time interval 2τ centered at the present is

$$\overline{v} = \frac{x(t + \tau) - x(t - \tau)}{2\tau} \tag{3}$$

so if we make our best estimate of where the particle will be a time τ in the future that is long on the time scale of thermal fluctuations, and where it was an equal time in the past, we have an estimate of its average slow velocity about the present time. The probability that it will move from x(t) to $y = x(t + \tau)$ in the future is given by some distribution $P(y \mid x,\tau)$. Its motion is the result of a large number of small increments (encounters with individual water molecules). Therefore the Central Limit Theorem, interpreted with the judgment that scientists develop (but cannot always explain to mathematicians, because it draws on extra information that a mathematician would never use in proving the theorem) tells us that this will have a Gaussian form, and from symmetry the mean displacement is zero:

$$P(y \mid x,I) = A \exp[- (y-x)^2/2\sigma^2(\tau)] \tag{4}$$

where I stands for the general prior information stated or implied in our formulation of the problem. All the analysis one could make of the dynamics of sugar-water interactions would, in the end, serve only to determine the spreading function $\sigma^2(\tau) = (\delta x)^2$, the expected square of

the displacement.

But now our trouble begins; the particle is as likely to be battered to the right as to the left; so from symmetry, the expectation of y is $\langle y \rangle = x$. Now all the equations of motion, however complicated, are at least time-reversal invariant. Therefore for the past position $z = x(t-\tau)$ we should have the same probability distribution (4) which is independent of the sign of τ, and again $\langle z \rangle = x(t)$. Therefore the estimated velocity is zero.

Surely, this must be right, for our particle, interacting only with the surrounding water, has no way of knowing that other sugar molecules are present, much less that there is any density gradient. From the standpoint of dynamics alone (i.e., forces and equations of motion) there is nothing that can give it any tendency to drift to regions of lower rather than higher density. Yet diffusion does happen!

In the face of this dilemma, Einstein was forced to invent strange, roundabout arguments -- half theoretical, half phenomenological -- in order to get a formula for diffusion. For example, first estimate how the density $n(x,t)$ would be changed a long time in the future by combining the distributions (4) generated by many different particles, then substitute it into the phenomenological diffusion equation that we get by combining (1) and (2); and from that reason backwards to the present time to see what the diffusion flux must have been.

This kind of indirect reasoning has been followed faithfully ever since in treatments of irreversible processes, because it seems to be the only thing that works. Attempts to calculate a flux directly at the present time give zero from symmetry, so one resorts to "forward integration" followed by backward reasoning. Yet this puzzles every thoughtful student, who thinks that we ought to be able to solve the problem by direct reasoning: calculate the flux $J(x,t)$ here and now, straight out of the physics of the situation.

Furthermore, instead of our having to assume a phenomenological form, a correct analysis ought to give it automatically; i.e. it should tell us from first principles why it is the density gradient, and not some other function of the density, that matters, and also under what conditions this will be true. Evidently, we have a real mystery here.

Why did our first attempt at direct reasoning fail? Because the problem is not one of physical prediction from the dynamics; it is a problem of inference. The question is not "How do the equations of motion require the particles to move about on the average?" The equations of motion do not require them to move about at all. The question is: "What is the best estimate we can make about how the particles are in fact moving in the present instance, based on all the information we have?" The equations of motion are symmetric in past and future; but our information about the particles is not.

Given the present position of a particle, what can we say about its future position? The zero movement answer above was correct; for predicting where it will be in the future, the knowledge of where it is now makes all prior information about where it might have been in the past irrelevant. But estimating where it was in the past is not a time-reversed mirror image of this, for we have prior knowledge of the varying density of particles in the past. Knowledge of where it is now does not make that prior knowledge irrelevant; and sound logic must take both into account.

Let us restate this in different language. Equation (4) expresses an average over the class of all possible motions compatible with the dynamics, in which movements to the right and the left have, from symmetry, equal weight. But of course, our particular particle is in fact executing only one of those motions. Our prior information selects out of the class of all possibilities in (4) a smaller class in which our particle is likely to be, in which movements to

the right and left do not have equal weight. It is not the dynamics, but the prior information, that breaks the symmetry and leads us to predict a non-zero flux.

While $P(x \mid z,t)$ is a direct probability, the same function as (4), the probability we now need is $P(z \mid x,t)$, an inverse probability which requires the use of Bayes' theorem:

$$P(z \mid x,t,I) = AP(z \mid I)P(x \mid z,I) \quad . \tag{5}$$

The prior probability $P(z \mid I)$ is clearly proportional to $n(z)$, and so from (3)

$$\log P(z \mid x,I) = \log n(z) - (z-x)^2/2\sigma^2(\tau) + (\text{const.}) \quad . \tag{6}$$

Differentiating, the most probable value of the past position z is not x, but

$$\hat{z} = x + \sigma^2 \, \text{grad}(\log n) = x + (\delta x)^2 \, \text{grad}(\log n) \tag{7}$$

whereupon, substituting into (3) we estimate the drift velocity to be

$$\overline{v} = -(\delta x)^2/2\tau \, \text{grad}(\log n) \tag{8}$$

and our predicted average diffusion flux over the time interval 2τ is

$$J(x,t) = n\,\overline{v} = -(\delta x)^2/2\tau \, \text{grad}(n) \quad . \tag{9}$$

Bayes' theorem has given us just Einstein's formula for the diffusion coefficient:

$$D = \frac{(\delta x)^2}{2\tau} \tag{10}$$

and a good deal more. We did not assume that grad(n) was the appropriate phenomenological form; Bayes' theorem told us that automatically. At the same time, it told us the condition for validity of that form; unless $(\delta x)^2$ is proportional to τ, there will be no unique diffusion coefficient, but only a sequence of apparent diffusion coefficients $D(\tau)$ for the average drift over different time intervals 2τ. Then the flux $J(x,t)$ will depend on other properties of $n(x,t)$ than its gradient, and in place of (2) a more complete Bayesian analysis will give a different phenomenological relation, involving an average of grad(n) over a short time in the past. Thus (9) is only the beginning of the physical predictions that we can extract by Bayesian analysis.

While (8) is the best estimate of the average velocity that we could make from the assumed information, it does not determine the velocity of any one particle very well. But what matters is the prediction of the observable net flux of N particles. In principle we should have calculated the joint posterior distribution for the velocities of N particles, and estimated their sum. But since that distribution factors, the calculation reduces to N repetitions of the above one, and the relative accuracy of the prediction improves like the square root of N, the usual rule in probability theory.

In practice, with perhaps 0.001M sugar solutions, the relevant values of N are of the order of 1 E+16, and the prediction is highly reliable, in the following sense: for the great majority of the N-particle motions consistent with the information used, the flux is very close to the predicted value.

DISCUSSION

The above example may indicate the price that kinetic theory has paid for its failure to comprehend and use the Bayesian methods that Harold Jeffreys gave us 50 years ago, and how many other puzzles need to be reexamined from that viewpoint. The only reason why the fluxes persisted in being zero was failure to put the obviously necessary prior information into the probabilities. But as long as one thinks that probabilities are physically real things, it seems wrong to modify a probability merely because our state of knowledge has changed.

The idea that probabilities can be used to represent our own information is still foreign to "orthodox" teaching, although the above example shows what one gains by so doing. Prior information is often highly cogent, and sound reasoning requires that it be taken into account. In other fields this is considered a platitude; what would you think of a physician who looked only at your present symptoms, and refused to take note of your medical history?

In the next talk, Ray Smith will survey the arguments of George Polya and Richard Cox indicating the sense in which Bayesian inference is uniquely determined by simple qualitative desiderata of rationality and logical consistency. Here I want only to indicate something about the rationale of their application in real problems.

Conventional training in the physical sciences concentrates attention 100% on physical prediction; the word "inference" was never uttered once in all the science courses I ever took. Therefore, the above example was chosen because its rationale is clear and the actual calculation is utterly trivial; yet its power to yield not only results that previously required more work but also more details about them, is apparent at once.

To appreciate the distinction between physical prediction and inference it is essential to recognize that propositions at two different levels are involved. In physical prediction we are trying to describe the real world; in inference we are describing only our state of knowledge about the world. A philosopher would say that physical prediction operates at the ontological level, inference at the epistemological level. Failure to see the distinction between reality and our knowledge of reality puts us on the Royal Road to Confusion; this usually takes the form of the Mind Projection Fallacy, discussed below.

The confusion proceeds to the following terminal phase: a Bayesian calculation like the above one operates on the epistemological level and gives us only the best predictions that can be made from the information that was used in the calculation. But it is always possible that in the real world there are extra controlling factors of which we were unaware; so our predictions may be wrong. Then one who confuses inference with physical prediction would reject the calculation and the method; but in so doing he would miss the point entirely.

For one who understands the difference between the epistemological and ontological levels, a wrong prediction is not disconcerting; quite the opposite. For how else could we have learned about those unknown factors? It is only when our epistemological predictions fail that we learn new things about the real world; those are just the cases where probability theory is performing its most valuable function. Therefore, to reject a Bayesian calculation because it has given us an incorrect prediction is like disconnecting a fire alarm because that annoying bell keeps ringing. Probability theory is trying to tell us something important, and it behooves us to listen.

THE MIND PROJECTION FALLACY

It is very difficult to get this point across to those who think that in doing probability calculations their equations are describing the real world. But that is claiming something that one could never know to be true; we call it the Mind Projection Fallacy. The analogy is to a movie projector, whereby things that exist only as marks on a tiny strip of film appear to be real objects moving across a large screen. Similarly, we are all under an ego-driven temptation to project our private thoughts out onto the real world, by supposing that the creations of one's own imagination are real properties of Nature, or that one's own ignorance signifies some kind of indecision on the part of Nature.

The current literature of quantum theory is saturated with the Mind Projection Fallacy. Many of us were first told, as undergraduates, about Bose and Fermi statistics by an argument like this: "You and I cannot distinguish between the particles; *therefore* the particles behave differently than if we could." Or the mysteries of the uncertainty principle were explained to us thus: "The momentum of the particle is unknown; *therefore* it has a high kinetic energy." A standard of logic that would be considered a psychiatric disorder in other fields, is the accepted norm in quantum theory. But this is really a form of arrogance, as if one were claiming to control Nature by psychokinesis.

In our more humble view of things, the probability distributions that we use for inference do not describe any property of the world, only a certain state of information about the world. This is not just a philosophical position; it gives us important technical advantages because of the more flexible way we can then use probability theory. In addition to giving us the means to use prior information, it makes an analytical apparatus available for such things as eliminating nuisance parameters, at which orthodox methods are helpless. This is a major reason for the greater computational efficiency of the Jeffreys methods in data analysis.

In our system, a probability is a theoretical construct, on the epistemological level, which we assign in order to represent a state of knowledge, or that we calculate from other probabilities according to the rules of probability theory. A frequency is a property of the real world, on the ontological level, that we measure or estimate. So for us, probability theory is not an Oracle telling how the world must be: it is a tool for learning (1) Is our state of knowledge adequate to describe the world? or (2) For which aspects of the world is our information adequate to make predictions?

This point comes across much more strongly in our next example, where belief that probabilities are physically real produces a major quandary for quantum theory, in the EPR paradox. It is so bad that some have concluded, with the usual consistency of quantum theory, that (1) there is no real world, after all, and (2) physical influences travel faster than light.

BACKGROUND OF EPR

Quantum Mechanics (QM) is a system of mathematics that was not developed to express any particular physical ideas, in the sense that the mathematics of relativity theory expresses the ideas of Einstein, or that of genetics expresses the ideas of Mendel. Rather, it grew empirically, over about four decades, through a long series of trial-and-error steps. But QM has two difficulties; firstly, like all empirical equations, the process by which it was found gives no clue as to its meaning. QM has the additional difficulty that its predictions are incomplete, since in general it gives only probabilities instead of definite predictions, and it does not indicate what extra information would be required to make definite predictions.

Einstein and Schroedinger saw this incompleteness as a defect calling for correction in some future more complete theory. Niels Bohr tried instead to turn it into a merit by fitting it into his philosophy of complementarity, according to which one can have two different sets of concepts, mutually incompatible, one set meaningful in one situation, the complementary set in another. As several of his early acquaintances have testified (Rozental, 1964), the idea of complementarity had taken control of his mind years before he started to study quantum physics.

Bohr's "Copenhagen Theory" held that, even when the QM state vector gives only probabilities, it is a complete description of reality in the sense that nothing more can ever be known; not because of technological limitations, but as a matter of fundamental principle. It seemed to Einstein that this completeness claim was a gratuitous addition, in no way called for by the facts; and he tried to refute it by inventing thought experiments which would enable one to get more information than Bohr wished us to have. Somehow, the belief has been promulgated that Bohr successfully answered all of Einstein's objections.

But when we examine Bohr's arguments, we find a common logical structure; always they start by postulating that the available measurement apparatus is subject to his "uncertainty" limitations; and then by using only classical physics (essentially, only Liouville's theorem) they come to the conclusion that such an apparatus could not be used for Einstein's purpose. Bohr's foregone conclusion is always assured by his initial postulate, which simply appears out of nowhere. In our view, then, the issue remains open and we must raise our standards of logic before there can be any hope of resolving it.

Leslie Ballentine (1970) analyzed the Bohr and Einstein positions and showed that much of the chanting to the effect that "Bohr won, Einstein lost" is sustained by quoting Einstein's views and attributing them to Bohr. Virtually all physicists who do real quantum-mechanical calculations interpret their results in the sense of Einstein, according to which a pure state represents an ensemble of similarly prepared systems and is thus an incomplete description of an individual system. Bohr's completeness claim has never played any functional role in applications, and in that sense it is indeed gratuitous.

CONFRONTATION OR RECONCILIATION?

Put most briefly, Einstein held that the QM formalism is incomplete and that it is the job of theoretical physics to supply the missing parts; Bohr claimed that there are no missing parts. To most, their positions seemed diametrically opposed; however, if we can understand better what Bohr was trying to say, it is possible to reconcile their positions and believe them both. Each had an important truth to tell us.

But Bohr and Einstein could never understand each other because they were thinking on different levels. When Einstein says QM is incomplete, he means it in the ontological sense; when Bohr says QM is complete, he means it in the epistemological sense. Recognizing this, their statements are no longer contradictory. Indeed, Bohr's vague, puzzling sentences -- always groping for the right word, never finding it -- emerge from the fog and we see their underlying sense, if we keep in mind that Bohr's thinking is never on the ontological level traditional in physics. Always he is discussing not Nature, but our information about Nature. But physics did not have the vocabulary for expressing ideas on that level, hence the groping.

Paul Dirac, who was also living here in St. John's College at the time when he and Harold Jeffreys were doing their most important work side by side, seems never to have realized what Jeffreys had to offer him: probability theory as the vehicle for expressing

epistemological notions quantitatively. It appears to us that, had either Bohr or Dirac understood the work of Jeffreys, the recent history of theoretical physics might have been very different. They would have had the language and technical apparatus with which Bohr's ideas could be stated and worked out precisely without mysticism, and which Einstein would have understood and accepted at once.

Needless to say, we consider all of Einstein's reasoning and conclusions correct on his level; but on the other hand we think that Bohr was equally correct on his level, in saying that the act of measurement perturbs the system being measured, and this places a limitation on the information we can acquire and therefore on the predictions we are able to make. The issue is merely whether this limitation is as great, and has the same quantitative form, as Bohr supposed. This is still an open question, but we may be able to settle it soon in the quantum optics laboratory, thanks to the spectacular recent advances in experimental techniques such as those by H. Walther and coworkers (Rempe et al, 1987) as discussed by Knight (1987) and in the *Scientific American* (June 1987, p. 25).

Bohr had no really cogent reason for his postulate that the limitations on the ability of the QM formalism to predict are also -- in complete, quantitative detail -- limitations on what the experimenter can measure; this seems to us an outstanding example of the Mind Projection Fallacy. We need a more orderly division of labour; it is simply not the proper business of theoretical physics to make pronouncements about what can and what cannot be measured in the laboratory, any more than it would be for an experimenter to issue proclamations about what can and cannot be calculated in the theory.

We believe that to achieve a rational picture of the world it is necessary to set up another clear division of labour within theoretical physics; it is the job of the laws of physics to describe physical causation at the level of ontology, and the job of probability theory to describe human inferences at the level of epistemology. The Copenhagen theory scrambles these very different functions into a nasty omelette in which the distinction between reality and our knowledge of reality is lost.

Although we agree with Bohr that in different circumstances different quantities are predictable, in our view this does not cause the concepts themselves to fade in and out; valid concepts are not mutually incompatible. Therefore, to express precisely the effect of disturbance by measurement, on our information and our ability to predict, is not a philosophical problem calling for complementarity; it is a technical problem calling for probability theory as expounded by Jeffreys, and information theory. Indeed, we know that toward the end of his life, Bohr showed an interest in information theory.

EPR

But to return to the historical account; somehow, many physicists became persuaded that the success of the QM mathematical formalism proved the correctness of Bohr's private philosophy, even though few understood what that philosophy was. All the attempts of Einstein, Schroedinger, and others to point out the patent illogic of this were rejected and sneered at; it is a worthy project for future psychologists to explain why.

The Einstein-Podolsky-Rosen (EPR) article of 1935 is Einstein's major effort to explain his objection to the completeness claim by an example that he thought was so forceful that nobody could miss the point. Two systems, S_1 and S_2, that were in interaction in the past are now separated, but they remain jointly in a pure state. Then EPR showed that according to QM an experimenter can measure a quantity q_1 in S_1, whereupon he can predict with certainty

the value of q_2 in S_2. But he can equally well decide to measure a quantity p_1 that does not commute with q_1; whereupon he can predict with certainty the value of p_2 in S_2.

The systems can be so far apart that no light signal could have traveled between them in the time interval between the S_1 and S_2 measurements. Therefore, by means that could exert no causal influence on S_2 according to relativity theory, one can predict with certainty either of two noncommuting quantities, q_2 and p_2. EPR concluded that both q_2 and p_2 must have had existence as definite physical quantities before the measurements; but since no QM state vector is capable of representing this, the state vector cannot be the whole story.

Since today some think that merely to verify the correlations experimentally is to refute the EPR argument, let us stress that EPR did not question the existence of the correlations, which are to be expected in a classical theory. Indeed, were the correlations absent, their argument against the QM formalism would have failed. Their complaint was that, with physical causation unavailable, only instantaneous psychokinesis (the experimenter's free-will decision which experiment to do) is left to control distant events, the forcing of S_2 into an eigenstate of either q_2 or p_2. Einstein called this "a spooky kind of action at a distance".

To understand this, we must keep in mind that Einstein's thinking is always on the ontological level; the purpose of the EPR argument was to show that the QM state vector cannot be a representation of the "real physical situation" of a system. Bohr had never claimed that it was, although his strange way of expressing himself often led others to think that he was claiming this.

From his reply to EPR, we find that Bohr's position was like this: "You may decide, of your own free will, which experiment to do. If you do experiment E_1 you will get result R_1. If you do E_2 you will get R_2. Since it is fundamentally impossible to do both on the same system, and the present theory correctly predicts the results of either, how can you say that the theory is incomplete? What more can one ask of a theory?"

While it is easy to understand and agree with this on the epistemological level, the answer that I and many others would give is that we expect a physical theory to do more than merely predict experimental results in the manner of an empirical equation; we want to come down to Einstein's ontological level and understand what is happening when an atom emits light, when a spin enters a Stern-Gerlach magnet, etc. The Copenhagen theory, having no answer to any question of the form: "What is really happening when - - - ?", forbids us to ask such questions and tries to persuade us that it is philosophically naive to want to know what is happening. But I do want to know, and I do not think this is naive; and so for me QM is not a physical theory at all, only an empty mathematical shell in which a future theory may, perhaps, be built.

THE BELL INEQUALITIES

John Bell (1964) studied a simple realization of the EPR scenario in which two spin 1/2 particles denoted A and B were jointly in a pure singlet state (like the ground state of the Helium atom) in the past. This is ionized by a spin-independent interaction and they move far apart, but they remain jointly in a pure singlet state, in which their spins are perfectly anticorrelated.

Each of two experimenters, stationed at A and B, has a Stern-Gerlach apparatus, which he can rotate to any angle. Following Bell's notation, we denote by P(A|a) the probability that spin A will be found up in the direction of the unit vector "a"; and likewise P(B|b) refers to spin B being up in the direction b. For a singlet state, these are each equal to 1/2 from

symmetry. The spooky business appears in the joint probability, which QM gives as

$$P(AB\,|\,ab) = \frac{1}{2}\sin^2(\theta/2) \tag{11}$$

where $\cos\theta = a\cdot b$. This does not factor in the form $P(AB\,|\,ab) = P(A\,|\,a)P(B\,|\,b)$ as one might expect for independent measurements. We can measure A in any direction we please; whereupon we can predict with certainty the value of B in the same direction.

From this, EPR would naturally conclude that the results of all possible measurements on B were predetermined by the real physical situation at B; *i.e.*, if we find B up in any direction b, then we would have found the same result whether or not the A measurement was made. Bohr would consider this a meaningless statement, since there is no way to verify it or refute it. Also, he would stress that we can measure B in only one direction, whereupon the perturbation of the measurement destroys whatever might have been seen in any other direction. Note that, as always, Bohr is epistemological; the notion of a "real physical situation" is just not in his vocabulary or his thinking.

EPR will then require some hidden variables in addition to the QM state vector to define that "real physical situation" which is to predetermine the results of all measurements on B. Bell, seeking to accommodate them, defines a class of hidden variable theories -- call them Bell theories -- in which a set of variables denoted collectively by λ also influences the outcomes A and B. It appears to him that the intentions of EPR are expressed in the most general way by writing

$$P(AB\,|\,ab) = \int P(A\,|\,a\lambda)\,P(B\,|\,b\lambda)\,p(\lambda)\,d\lambda \tag{12}$$

and he derives some inequalities that must be satisfied by any probability expressible in this form. But the QM probabilities easily violate these inequalities, and therefore they cannot result from any Bell theory.

Of course, the fundamentally correct relation according to probability theory would be,

$$P(AB\,|\,ab) = \int P(AB\,|\,ab\lambda)\,P(\lambda\,|\,ab)\,d\lambda \quad . \tag{13}$$

But if we grant that knowledge of the experimenters' free choices (a,b) would give us no information about λ: $P(\lambda\,|\,ab) = p(\lambda)$ (and in this verbiage we too are being carefully epistemological), then Bell's interpretation of the EPR intentions lies in the factorization

$$P(AB\,|\,ab\lambda) = P(A\,|\,a\lambda)\,P(B\,|\,b\lambda) \tag{14}$$

whereas the fundamentally correct factorization would read:

$$P(AB\,|\,ab\lambda) = P(A\,|\,Bab\lambda)\,P(B\,|\,ab\lambda) = P(A\,|\,ab\lambda)\,P(B\,|\,Aab\lambda) \tag{15}$$

in which both a,b always appear as conditioning statements. However, Bell thinks that the EPR demand for locality, in which events at A should not influence events at B when the interval is spacelike, require the form (14). In his words, "It would be very remarkable if b proved to be a causal factor for A, or a for B; *i.e.*, if $P(A\,|\,a\lambda)$ depended on b or $P(B\,|\,b\lambda)$ depended on a. But according to quantum mechanics, such a dilemma can happen. Moreover, this peculiar long-range influence in question seems to go faster than light".

Note, however, that merely knowing the direction of the A measurement does not change any predictions at B, although it converts the initial pure singlet state into a mixture. It is easy to verify that according to QM, $P(B\,|\,ab) = P(B\,|\,b) = 1/2$ for all a,b. As we would expect from (15), it is necessary to know also the result of the A measurement before the correlation

affects our predictions; according to QM, $P(B \mid Aab) = (1 - \cos\theta)/2$. Thus while the QM formalism disagrees with the factorization (14), it agrees with what we have called the "fundamentally correct" probability relations (perhaps now it is clearer why we said that some of Bohr's ideas could have been expressed precisely in Bayesian terms).

Regardless, it seemed to everybody twenty years ago that the stage was set for an experimental test of the issue; perform experiments where the predictions of quantum theory violate the Bell inequalities, and see whether the data violate them also. If so, then all possible local causal theories are demolished in a single stroke, and the Universe runs on psychokinesis. At least, that was the reasoning.

The experiments designed to test this, of which the one of Alain Aspect (1985, 1986) is perhaps the most cogent to date, have with only one exception ended with the verdict "quantum theory confirmed", and accordingly there has been quite a parade of physicists jumping on the bandwagon, declaring publicly that they now believe in psychokinesis. Of course, they do not use that word; but at the 1984 Santa Fe Workshop more than one was heard to say: "The experimental evidence now forces us to believe that atoms are not real." and nobody rose to question this, although it made me wonder what they thought Alain's apparatus was made of.

Alain Aspect himself has remained admirably level-headed through all this, quite properly challenging us to produce a classical explanation of his experiment; but at the same time refusing to be stampeded into taking an obviously insane position as did some others.

The dilemma is not that the QM formalism is giving wrong predictions, but that the current attempts at interpreting that formalism from Einstein's ontological viewpoint are giving us just that spooky picture of the world that Einstein anticipated and objected to. Of course, those with a penchant for mysticism are delighted.

How do we get out of this? Just as Bell revealed hidden assumptions in von Neumann's argument, so we need to reveal the hidden assumptions in Bell's argument. There are at least two of them, both of which require the Jeffreys viewpoint about probability to recognize:

(1) Bell took it for granted that a conditional probability $P(X \mid Y)$ expresses a physical causal influence, exerted by Y on X. But we show presently that one cannot even reason correctly in so simple a problem as drawing two balls from Bernoulli's Urn, if he interprets probabilities in this way. Fundamentally, consistency requires that conditional probabilities express logical inferences, just as Harold Jeffreys saw. Indeed, this is also the crucial point that Bohr made in his reply to EPR, in words that Bell quoted. But Bell added: "Indeed I have very little idea what this means."

(2) The class of Bell theories does not include all local causal theories; it appears to us that it excludes just the class of theories that Einstein would have liked most. Again, we need to learn from Jeffreys the distinction between the epistemological probabilities of the QM formalism and the ontological frequencies that we measure in our experiments. A local causal theory need not reproduce the mathematical form of the QM probabilities in the manner of (12); rather, since by definition it operates at the ontological level, it should predict the frequencies observed in well-defined real experiments (not just thought-experiments).

The spooky stuff is a consequence of Hidden Assumption (1), and it disappears if we conclude, with Jeffreys and Bohr, that what is traveling faster than light is not a physical influence, but only a logical inference. To render Bohr's quoted statement into plain English:

> The measurement at A at time t does not change the real physical situation at B; but it changes our state of knowledge about that situation, and therefore it changes the predictions we are able to make about B at some time t′. Since this is a matter of logic rather than physical causation, it does not matter whether t′ is before, equal to, or after t.

Again we see how Bohr's epistemological viewpoint corresponds to Bayesian inference, and could have been expressed precisely in Bayesian terms. However, Bohr could not bring himself to say it as we did, because for him the phrase "real physical situation" was taboo.

But it may seem paradoxical that two different pure states (eigenstates of noncommuting quantities q_2 and p_2) can both represent the same real physical situation; if so, then perhaps the conclusion is that one has learned an important fact about the relation of the QM state vector to reality. This supports the Einstein view of the meaning of a pure state as an ensemble; for in statistical mechanics it is a platitude that the true microstate may appear in two different ensembles, representing two different states of knowledge about the microstate.

BERNOULLI'S URN REVISITED

Define the propositions:

$I \equiv$ "Our urn contains N balls, identical in every respect except that M of them are red, the remaining (N - M) white. We have no information about the location of particular balls in the urn. They are drawn out blindfolded without replacement."

$R_i \equiv$ "Red on the i'th draw, $i = 1, 2, \ldots$"

Successive draws from the urn are a microcosm of the EPR experiment. For the first draw, given only the prior information I, we have

$$P(R_1 \mid I) = M/N \tag{16}$$

Now if we know that red was found on the first draw, then that changes the contents of the urn for the second:

$$P(R_2 \mid R_1, I) = (M-1)/(N-1) \tag{17}$$

and this conditional probability expresses the causal influence of the first draw on the second, in just the way that Bell assumed.

But suppose we are told only that red was drawn on the second draw; what is now our probability for red on the first draw? Whatever happens on the second draw cannot exert any physical influence on the condition of the urn at the first draw; so presumably one who believes that conditional probability must express physical causation would say that $P(R_1 \mid R_2, I) = P(R_1 \mid I)$. But this is patently wrong; probability theory requires that

$$P(R_1 \mid R_2, I) = P(R_2 \mid R_1, I) \tag{18}$$

This is particularly obvious in the case M = 1; for if we know that the one red ball was taken in the second draw, then it is certain that it could not have been taken in the first.

In (18) the probability on the right expresses a physical causation, that on the left only an inference. Nevertheless, the probabilities are necessarily equal because, although a later draw cannot physically affect conditions at an earlier one, *information* about the result of the second draw has precisely the same effect on our *state of knowledge* about what could have been taken in the first draw, as if their order were reversed.

Eq. (18) is only a special case of a much more general result. The probability of drawing any sequence of red and white balls (the hypergeometric distribution) depends only on the number of red and white balls, not on the order in which they appear; i.e. it is an exchangeable distribution. From this it follows by a simple calculation that for all i and j,

$$P(R_i \mid I) = P(R_j \mid I) = M/N \tag{19}$$

That is, just as in QM, merely knowing that other draws have been made does not change our prediction for any specified draw, although it changes the hypothesis space in which the prediction is made; before there is a change in the actual prediction it is necessary to know also the results of other draws. But the joint probability is by the product rule,

$$P(R_i,R_j \mid I) = P(R_i \mid R_j,I)\, P(R_j \mid I) = P(R_j \mid R_i,I)\, P(R_i \mid I) \tag{20}$$

and so we have for all i and j,

$$P(R_i \mid R_j,I) = P(R_j \mid R_i,I) \tag{21}$$

and again a conditional probability which expresses only an inference is necessarily equal to one that expresses a physical causation. This would be true not only for the hypergeometric distribution, but for any exchangeable distribution. We see from this how far Karl Popper would have got with his "propensity" theory of probability, had he tried to apply it to a few simple problems.

It might be thought that this phenomenon is a peculiarity of probability theory. On the contrary, it remains true even in pure deductive logic; for if A implies B, then not-B implies not-A. But if we tried to interpret "A implies B" as meaning "A is the physical cause of B", we could hardly accept that "not-B is the physical cause of not-A". Because of this lack of contraposition, we cannot in general interpret logical implication as physical causation, any more than we can conditional probability. Elementary facts like this are well understood in economics (Simon & Rescher, 1966; Zellner, 1984); it is high time that they were recognized in theoretical physics.

OTHER HIDDEN - VARIABLE THEORIES

Now consider Hidden Assumption (2). Bell theories make no allowance for time variation of the hidden variable λ; but if it is to take over the job formerly performed by the QM state vector ψ, then λ must obey some equations of motion which are to replace the Schroedinger equation.

This is important, because one way for a causal theory to get probability into things is time alternation; for example, in conditions where present QM yields a time independent probability p for spin up, λ would be oscillating in such a way that for a fraction p of the time the result is "up", etc. Indeed, Einstein would have considered this the natural way to obtain the QM probabilities from a causal theory, for in his early papers he defined the "probability of a state" as the fraction of the time in which a system is in that state. But this is a relation between QM and the causal theory of a different nature than is supposed by the form (12).

Time alternation theories have another attractive feature, that they predict new effects that might in principle be observed experimentally, leading to a crucial test. For example, when two spins are perfectly anticorrelated, that would presumably signify that their λ's are oscillating in perfect synchronism so that, for a given result of the A measurement, the exact time interval between the A and B measurements would determine the actual result at B, not

merely its probability. Then we would be penetrating the Copenhagen fog and observing more than Bohr thought possible. The experiments of H. Walther and coworkers on single atom masers are already showing some resemblance to the technology that would be required to perform such an experiment.

We have shown only that some of the conclusions that have been drawn from the Bell-Aspect work were premature because (1) the spooky stuff was due only to the assumption that a conditional probability must signify a physical influence, and (2) the Bell arguments do not consider all possible local causal theories; the Bell inequalities are only limitations on what can be predicted by Bell theories. The Aspect experiment may show that such theories are untenable, but without further analysis it leaves open the status of other local causal theories more to Einstein's liking.

That further analysis is, in fact, already underway. An important part of it has been provided by Steve Gull's "You can't program two independently running computers to emulate the EPR experiment" theorem, which we learned about at this meeting. It seems, at first glance, to be just what we have needed because it could lead to more cogent tests of these issues than did the Bell argument. The suggestion is that some of the QM predictions can be duplicated by local causal theories only by invoking teleological elements as in the Wheeler-Feynman electrodynamics. If so, then a crucial experiment would be to verify the QM predictions in such cases. It is not obvious whether the Aspect experiment serves this purpose.

The implication seems to be that, if the QM predictions continue to be confirmed, we exorcise Bell's superluminal spook only to face Gull's teleological spook. However, we shall not rush to premature judgments. Recalling that it required some 30 years to locate von Neumann's hidden assumptions, and then over 20 years to locate Bell's, it seems reasonable to ask for one year to search for Gull's, before drawing conclusions and possibly suggesting new experiments.

In this discussion we have not found any conflict between Bohr's position and Bayesian probability theory, which are both at the epistemological level. Nevertheless, differences appear on more detailed examination to be reported elsewhere. Of course, the QM formalism also contains fundamentally important and correct ontological elements; for example, there has to be something physically real in the eigenvalues and matrix elements of the operators from which we obtain detailed predictions of spectral lines. It seems that, to unscramble the epistemological probability statements from the ontological elements we need to find a different formalism, isomorphic in some sense but based on different variables; it was only through some weird mathematical accident that it was possible to find a variable ψ which scrambles them up in the present way.

There is clearly a major, fundamentally important mystery still to be cleared up here; but unless you maintain your faith that there is a rational explanation, you will never find that explanation. For 60 years, acceptance of the Copenhagen interpretation has prevented any further progress in basic understanding of physical law. Harold Jeffreys (1957) put it just right: "Science at any moment does not claim to have explanations of everything; and acceptance of an inadequate explanation discourages search for a good one."

Now let us turn to an area that seems about as different as one could imagine, yet the underlying logic of it hangs on the same point: What happens in the real world depends on physical law and is on the level of ontology. What we can predict depends on our state of knowledge and is necessarily on the level of epistemology. He who confuses reality with his knowledge of reality generates needless artificial mysteries.

THE SECOND LAW IN BIOLOGY

As we learn in elementary thermodynamics, Kelvin's formula for the efficiency of a Carnot heat engine operating between upper and lower temperatures T_1, T_2:

$$\eta \leq 1 - T_2/T_1 \quad , \tag{22}$$

with equality if and only if the engine is reversible, expresses a limitation imposed by the second law of thermodynamics. But the world's most universally available source of work -- the animal muscle -- presents us with a seemingly flagrant violation of that formula.

Our muscles deliver useful work when there is no cold reservoir at hand (on a hot day the ambient temperature is at or above body temperature) and a naive application of (22) would lead us to predict zero, or even negative efficiency. The observed efficiency of a muscle, defined as

$$\eta = \frac{\text{(work done)}}{\text{(work done + heat generated)}}$$

is difficult to measure, and it is difficult to find reliable experimental values with accounts of how the experiments were done. We shall use only the latest value we have located, (Alberts, *et al.* 1983). The heat generated that can be attributed to muscle activity appears to be as low as about 3/7 of the work done; which implies that observed muscle efficiencies can be as high as 70% in favourable conditions, although a Carnot engine would require an upper temperature T_1 of about 1000 K to achieve this. Many authors have wondered how this can be.

The obvious first answer is, of course, that a muscle is not a heat engine. It draws its energy, not from any heat reservoir, but from the activated molecules produced by a chemical reaction. Only when we first allow that primary energy to degrade itself into heat at temperature T_1 -- and then extract only that heat for our engine -- does the Kelvin efficiency formula (22) apply in its conventional meaning. It appears that our muscles have learned how to capture the primary energy before it has a chance to degrade; but how do we relate this to the second law?

Basic material on muscle structure and energetics of biochemical reactions is given by Squire (1981) and Lehninger (1982), and profusely illustrated by Alberts, *et al* (1983). The source of energy for muscle contraction (and indeed for almost everything a cell does that requires energy) is believed to be hydrolysis of adenosine triphosphate (ATP), for which the reported heat of reaction is $\Delta H = -9.9$ kcal/mol, or 0.43 ev per molecule. This energy is delivered to some kind of "engine" in a muscle fiber, from whence emerges useful work by contraction. The heat generated by a muscle is carried off by the blood stream, at body temperature, 273 + 37 = 310 K. Thus the data we have to account for are:

Ambient temperature:	310 K
Source energy:	0.43 ev/molecule
Efficiency:	70%.

We do not attempt to analyze all existing biological knowledge in this field about the details of that engine, although in our conclusions we shall be able to offer some tentative comments on it. Our present concern is with the general physical principles that must govern conversion of chemical energy into mechanical work in any system, equilibrium or nonequilibrium, biological or otherwise, whatever the details of the engine. In the known facts of muscle performance we have some uniquely cogent evidence relevant to this problem.

The status of the second law in biology has long been a mystery. Not only was there a seeming numerical contradiction between muscle efficiency and the second law, but also the general self-organizing power of biological systems seemed to conflict with the "tendency to disorder" philosophy that had become attached to the second law (much as Bohr's philosophy of complementarity had become attached to quantum mechanics). This led, predictably, to a reaction in the direction of vitalism.

In our view, whatever happens in a living cell is just as much a real physical phenomenon as what happens in a steam engine; far from violating physical laws, biological systems exhibit the operation of those laws in their full generality and diversity, that physicists had not considered in the historical development of thermodynamics. Therefore, if biological systems seem to violate conventional statements of the second law, our conclusion is only that the second law needs to be restated more carefully. Our present aim is therefore to find a statement of the second law that reduces to the traditional statements of Clausius and Gibbs in the domain where they were valid, but is general enough to include biological phenomena.

The "tendency to disorder" arguments are too vague to be of any use to us, although it is clear that they must be mistaken and it would be interesting to understand why. Muscle efficiency will provide our test case, because here we have some quantitative data to account for. But a muscle operates in a nonequilibrium situation, for which no definite second law is to be found in the current thermodynamic literature. The conventional second law presupposes thermalization because temperature and entropy are defined only for states of thermal equilibrium. How do we circumvent this?

Some have thought that it would be a highly difficult theoretical problem, calling for a generalised ergodic theory to include analysis of ''mixing'' and ''chaos''. Another school of thought holds that we need a modification of the microscopic equations of motion to circumvent Liouville's theorem (conservation of phase volume in classical Hamiltonian systems, or unitarity in quantum theory), which is thought to be in conflict with the second law.

We suggest, on the contrary, that only very simple physical reasoning is required, and all the clues pointing to it can be found already in the writings of James Clerk Maxwell and J. Willard Gibbs over 100 years ago. Both had perceived the epistemological nature of the second law and we think that, had either lived a few years longer, our generalised second law would long since have been familiar to all scientists. We give the argument in three steps: (a) reinterpret the Kelvin formula, (b) make a more general statement of the second law, (c) test it numerically against muscles.

The observed efficiency of muscles may be more cogent for this purpose than one might at first think. Since animals have evolved the senses of sight, sound, and smell to the limiting sensitivity permitted by physical law, it is only to be expected that they would also have evolved muscle efficiency (which must be of equal survival value) correspondingly. If so, then the maximum observed efficiency of muscles should be not merely a lower bound on the maximum theoretical efficiency we seek, but close to it numerically.

GENERALISED EFFICIENCY FORMULA

Consider the problem first in the simplicity of classical physics, where the Rayleigh-Jeans equipartition law holds. If in the Kelvin formula (22) we replace temperature by what it then amounts to -- energy per degree of freedom $E/N = (1/2)kT$, it takes the form

$$\eta = 1 - (E_2/N_2)(N_1/E_1) \tag{23}$$

which does not look like much progress, but by this trivial rewriting we have removed the limitation of thermal equilibrium on our energy source and sink. For "temperature" is defined only for a system in thermal equilibrium, while "energy per degree of freedom" is meaningful not only in thermal equilibrium, but for any small part of a system -- such as those activated molecules -- which might be far from thermal equilibrium with the surroundings.

One might then question whether such a nonequilibrium interpretation of (22) is valid. We may, however, reason as follows. Although conventional thermodynamics defines temperature and entropy only in equilibrium situations where all translational and vibrational degrees of freedom (microscopic coordinates) have the same average energy, it cannot matter to an engine whether all parts of its energy source are in equilibrium with each other.

Only those degrees of freedom with which the engine interacts can be involved in its efficiency; the engine has no way of knowing whether the others are or are not excited to the same average energy. Therefore, since (23) is unquestionably valid when both reservoirs are in thermal equilibrium, it should be correct more generally, if we take E_2/N_2 and E_1/N_1 to be the average energy in those degrees of freedom with which the engine actually interacts. But while a muscle has a small source reservoir, it has a large sink. Therefore for E_2/N_2 we may take $(1/2)kT_2$ at body temperature.

As a check on this reasoning, if the primary energy is concentrated in a single degree of freedom and we can extract it before it spreads at all, then our engine is in effect a "pure mechanism" like a lever. The generalised efficiency (23) then reduces to $1 - kT_2/2E_1$ or, interpreting E_1 as the work delivered to it,

$$\text{(Work out)} = \text{(Work in)} - (1/2)kT_2 \ . \tag{24}$$

The last term is just the mean thermal energy of the lever itself, which cannot be extracted reproducibly by an apparatus that is itself at temperature T_2 or higher. At least, if anyone should succeed in doing this, then he would need only to wait a short time until the lever has absorbed another $(1/2)kT_2$ from its surroundings, extract that, and repeat -- and we would have the perpetual motion machine that the second law holds to be impossible. Thus (24) still expresses a second law limitation, and the simple generalisation (23) of Kelvin's formula appears to have a rather wide range of application.

But although these are interesting hints, we are after something more general, which can replace the second law for all purposes, not merely engines. To achieve this we must understand clearly the basic physical reason why there is a second law limitation on processes. We suggest that the fundamental keyword characterizing the second law is not "disorder", but "reproducibility".

THE REASON FOR IT

The second law arises from a deep interplay between the epistemological macrostate (i.e. the variables like pressure, volume, magnetization that an experimenter measures and which therefore represent our knowledge about the system) and the ontological microstate (the coordinates and momenta of individual atoms, which determine what the system will in fact do). For example, in either a heat engine or a muscle the goal is to recapture energy that is spread in an unknown and uncontrolled way over many microscopic degrees of freedom of the source reservoir, and concentrate it back into a single degree of freedom, the motion of a piston or tendon. The more it has spread, the more difficult it will be to do this.

The basic reason for the "second law" limitation on efficiency is that the engine must work reproducibly; an engine that delivered work only occasionally, by chance (whenever the initial microstate of reservoirs and engine happened to be just right) would be unacceptable in engineering and biology alike.

The initial microstate is unknown because it is not being controlled by any of the imposed macroscopic conditions. The initial microstate might be anywhere in some large phase volume W_i compatible with the initial macrostate M_i; and the engine must still work. It is then Liouville's theorem that places the limitation on what can be done; physical law does not permit us to concentrate the final microstates into a smaller phase volume than W_i and therefore we cannot go reproducibly to any final macrostate M_f whose phase volume W_f is smaller than W_i. The inequality $W_f \geq W_i$ is a necessary condition for any macroscopic process $M_i \rightarrow M_f$ to be reproducible, whatever the initial microstate in W_i.

Of course, something may happen by chance that is not reproducible. As a close analogy, we can pump the water from a tank of volume V_1 into a larger tank of volume $V_2 > V_1$, but not into a smaller one of volume $V_3 < V_1$. Therefore any particular tagged water molecule in one tank can be moved reproducibly into a larger tank but not into a smaller one; the probability of success would be something like V_3 / V_1. Here the tanks correspond to the macrostates M, their volumes V correspond to phase volumes W, the tagged molecule represents the unknown true microstate, and the fact that the water flow is incompressible corresponds to Liouville's theorem.

Now we know that in classical thermodynamics, as was first noted by Boltzmann, the thermodynamic entropy of an equilibrium macrostate M is given to within an additive constant by $S(M) = k \log W(M)$, where k is Boltzmann's constant. This relation was then stressed by Planck and Einstein, who made important use of it in their research. But the above arguments make it clear that there was no need to restrict this to equilibrium macrostates M. Any macrostate -- equilibrium or nonequilibrium -- has an entropy $S(M) = k \log W(M)$, where $W(M)$ is the phase volume compatible with the controlled or observed macrovariables X_i (pressure, volume, magnetization, heat flux, electric current, etc.) that define M. Then a generalised second law

$$S(\text{final}) \geq S(\text{initial}) \tag{25}$$

follows immediately from Liouville's theorem, as a necessary condition for a change of state $M_i \rightarrow M_f$ to be reproducible.

Stated more carefully, we mean "reproducible by an experimenter who can control only the macrovariables $\{X_i\}$ that define the macrostates M". A little thought makes it clear that this proviso was needed already in the classical equilibrium theory, in order to have an air-tight statement of the second law which could not be violated by a clever experimenter. For if

Mr. A defines his thermodynamic states by the n macrovariables $\{X_1, \ldots, X_n\}$ that he is controlling and/or observing, his entropy S_n is a function of those n variables. If now Mr. B, unknown to Mr. A, manipulates a new macrovariable X_{n+1} outside the set that Mr. A is controlling or observing, he can bring about, reproducibly, a change of state for which S_n decreases, although S_{n+1} does not. Thus he will appear to Mr. A as a magician who can produce spontaneous violations of the second law, at will.

But now we must face an ambiguity in the definition and meaning of W; it appears to have different aspects. The phase volume $W(X_1, \ldots, X_n)$ consistent with a given set of extensive macrovariables $\{X_1, \ldots, X_n\}$ is a definite, calculable quantity which represents on the one hand the degree of control of an experimenter over the microstate, when he can manipulate only those macrovariables; thus W appears ontological. On the other hand, W represents equally well our degree of ignorance about the microstate when we know only those macrovariables and nothing else; and thus it appears epistemological. But as illustrated by the scenario of Mr. A and Mr. B above, it is a matter of free choice on our part which set of macrovariables we shall use to define our macrostates; thus it appears also anthropomorphic! Finally, we have been vague about just how many microscopic degrees of freedom are to be included in W. Then what is the meaning of the second law (25)? Is it an ontological law of physics, an epistemological human prediction, or an anthropomorphic art form?

The answer is that Eq. (25) cannot be an ontological statement (i.e. a deductively proved consequence of the laws of physics) because the mere calculation of W makes no use of the equations of motion, which alone determine which macrostate will in fact evolve from a given microstate in W_i. It may be that, because of properties of the equations of motion that we did not use, our experimenter's method of realizing the macrostate M_i would not, in many repetitions, produce all microstates in the volume W_i, only a negligibly small subset of them occupying a phase volume $W' << W_i$. Then the process $M_i \rightarrow M_f$ might still be possible reproducibly even though $S_f < S_i$, if $S_f > S'$. Conversely, because of special circumstances such as unusual constants of the motion, the process $M_i \rightarrow M_f$ might prove to be impossible even though $S_f > S_i$.

On the other hand, (25) is always epistemological because it is always true that W(M) measures our degree of ignorance about the microstate when we know only the macrostate M. Thus the original second law and our generalisation (25) of it have the same logical status as Bayesian inference; they represent the best predictions we can make from the information we have. In fact, by a more sophisticated approach a refined form of (25) can be derived as an example of Bayesian inference. Therefore the second law works functionally like any other Bayesian inference; the predictions are usually right, indicating that the information and assumptions used in the calculation were adequate for the purpose. Only when the predictions are wrong do we learn new things about the ontological laws of physics.

It is greatly to our advantage to recognize this. By getting our logic straight we not only avoid the Mind Projection Fallacy of claiming more than has been proved, we gain an important technical flexibility in using the second law. Instead of committing the error of supposing that a given physical system has one and only one "true" ontological entropy, we recognize that we could have many different states of knowledge about it, leading to many different entropies (as in the scenario of Mr. A and Mr. B above), which can serve many different purposes.

Just as the class of phenomena that an experimenter can evoke from a given system in the laboratory depends on the kind of apparatus he has (i.e. on which of its macrovariables he

can manipulate), so the class of phenomena that we can predict with thermodynamics for a given system depends on the kind of knowledge we have about it. This is not a paradox, but a platitude.

One reason why the second law has had little useful application in biology is failure to recognize that it is not an ontological law of physics; it is only a rule for conducting human inference. If you fail to specify what biological information you propose to take into account, then thermodynamics may not be able to give you any useful answer because you have not asked any well posed question.

Even when it does not lead to different final results, taking prior information into account can affect computational efficiency in applying the second law, because it can help us to make a more parsimonious choice of the microvariables that we shall include in W. For it to be generally valid, the entropy in (25) must be, in principle, the total entropy of all systems that take part in the process. But this does not, in practice, determine exactly how much of the outside world is to be included. In a sense everything in the universe is in constant interaction with everything else, and one must decide when to stop including things. Including more than we need is not harmful in the sense of leading to errors, since this only adds the same quantity to both sides of (25). But it can cost us wasted effort in calculating unnecessary details that cancel out of our final results.

At this point the aforementioned flexibility of our methods becomes important. We have already made use of it in the discussion following Eq. (23); now we want to apply that reasoning to phase volumes and to general processes. In a fast process, that happens in a time so short that thermal equilibrium of the whole system is never reached, only the phase volume belonging to those degrees of freedom actually involved in the interactions could be relevant; the second law may be applied in terms of Liouville's theorem in a relatively small subspace of the full one that we use in equilibrium theory. In the application to muscle efficiency, this means that we need calculate only phase volumes corresponding to degrees of freedom that are directly involved in muscle operation; ones that are affected only later, after the muscle contraction is over, may be relevant for the ultimate fate of the heat generated, but they cannot affect its efficiency.

This corresponds to a familiar procedure in treatment of spin systems. Spin-spin relaxation is often orders of magnitude faster than spin-lattice relaxation, so one can consider the microvariables of the spin system as forming a nearly isolated dynamical system in their own right, with a "private" second law of their own. Slichter (1980) shows that this approach enables one to predict masses of details correctly.

In the above we have supposed the classical equipartition law; but our arguments should need modifying only if the engine (i.e., the piston or tendon) interacts directly with degrees of freedom for which equipartition fails. In the case of muscles, it appears that the direct interactions are with coordinates of low-frequency vibration modes of large protein molecules. How energy gets transferred from an excited electronic state of ATP to such a vibration mode would remain in the province of quantum theory; but this can be virtually 100% efficient.

QUANTITATIVE DERIVATION

Now we are ready for a specific calculation of muscle efficiency using the above principles. The phase volumes W that we calculate are, of course, functions of the macrovariables that define the macrostates. In the case of a muscle, what is happening is just that energy Q_1 is being abstracted from the source reservoir and energy Q_2 is delivered to the sink, the difference appearing as work. Energy is the only macrovariable being manipulated, so our phase volumes will be functions of source and sink energies. We need not consider a phase volume for the engine, because that is the same at the beginning and end (the engine is restored ready to run again). As in conventional statistical mechanics, we introduce the density functions $\rho(E)$, often called structure functions, of source and sink by considering their energies known to some tolerances δE. Thus the phase volumes for source and sink are

$$W_1 = \rho_1(E_1)\, \delta E_1 \tag{26a}$$

$$W_2 = \rho_2(E_2)\, \delta E_2 \tag{26b}$$

Then the initial and final phase volumes are

$$W_i = \rho_1(E_1)\, \rho_2(E_2)\, \delta E_1\, \delta E_2 \tag{27a}$$

$$W_f = \rho_1(E_1 - Q_1)\, \rho_2(E_2 + Q_2)\, \delta E_1\, \delta E_2 \tag{27b}$$

With Q_1 and Q_2 definite quantities, the tolerances δE_1 and δE_2 are the same at the beginning and end, so they cancel out and their values do not matter. The necessary condition of reproducibility $W_i \leq W_f$ when we manipulate only energies now becomes:

$$\rho_1(E_1)\, \rho_2(E_2) \;\leq\; \rho_1(E_1 - Q_1)\, \rho_2(E_2 + Q_2)\ . \tag{28}$$

Let us try to predict the maximum work obtainable, using only this relation (which makes no use of such notions as temperature, equation of state, heat capacity, or reversible operation). Given the energy Q_1 extracted from the source, the maximum work we can get reproducibly is $Q_1 - Q_2$, where from (28), Q_2 is the root of

$$\log \rho_1(E_1) + \log \rho_2(E_2) = \log \rho_1(E_1 - Q_1) + \log \rho_2(E_2 + Q_2)\ . \tag{29}$$

Now vary Q_1; the RHS of (29) remains constant, and $Q_1 - Q_2$ is a maximum when

$$-\frac{\partial}{\partial Q_1}\log \rho_1(E_1 - Q_1) = \frac{\partial}{\partial Q_2}\log \rho_2(E_2 + Q_2) \tag{30}$$

Therefore the maximum efficiency is

$$\eta = \frac{Q_1 - Q_2}{E_1} \tag{31}$$

where Q_1, Q_2 are the simultaneous roots of (29) and (30).

Now we need to decide on the functions $\rho_1(E_1)$ and $\rho_2(E_2)$. Recall some familiar examples of such functions; for an ideal gas of n particles in volume V,

$$\rho(E) = \frac{V^n (2\pi m E)^{\frac{3n}{2} - 1}}{\Gamma(3n/2)}\ . \tag{32}$$

For n classical harmonic oscillators with frequencies $\{\omega_1, \ldots, \omega_n\}$,

$$\rho(E) = \frac{(2\pi)^n}{(\Pi_i \omega_i)\Gamma(n)}\, E^n\ . \tag{33}$$

In both cases, $\rho(E)$ is proportional to $E^{N/2}$, where N is the number of degrees of freedom of the system. This is approximately true for most systems even in quantum statistics, where N may be regarded as a slowly varying function of E, signifying the effective number of degrees of freedom excited at energy E. So let us take

$$\log \rho_1(E_1) = \frac{N_1}{2} \log E_1 + \text{const.} \tag{34a}$$

$$\log \rho_2(E_2) = \frac{N_2}{2} \log E_2 + \text{const.} \tag{34b}$$

which seems quite realistic for the case of muscles. Eliminating Q_2 from (29), (30), Q_1 is determined from

$$(N_1 + N_2) \log \left[\frac{E_1 - Q_1}{E_1} \right] = N_2 \log \left[\frac{N_1 E_2}{N_2 E_1} \right] \tag{35}$$

and then Q_2 is found from (30). But from (23) we recognize the quantity

$$r \equiv \frac{N_1 E_2}{N_2 E_1} \tag{36}$$

as the analog of (T_2/T_1) in equilibrium theory. Then after some algebra, we find that (31) is

$$\eta = 1 + \frac{N_2}{N_1} r - \left[\frac{N_1 + N_2}{N_1} \right] r^{\frac{N_2}{N_1 + N_2}} . \tag{37}$$

In the case $N_1 = N_2$, this is $(1 - \sqrt{r}\,)^2$, contrasted with Kelvin's differential efficiency $(1 - r)$. Appropriate for muscles is the limiting form as $N_2 \to \infty$, $E_2/N_2 \to \frac{1}{2}\, kT_2 = \text{const.}$ (the blood stream is very large compared to a muscle fiber). Some care is needed in taking the limit, and (37) then reduces to

$$\eta = 1 - r + r \log r \tag{38}$$

Now everything boils down to the question: what is r for a muscle? As before, let us take for the large sink reservoir, $E_2 = \frac{1}{2}\, N_2\, kT_2$ where $T_2 = 310$ K. The maximum theoretical efficiency surely corresponds to the maximum concentration of primary energy that seems possible in a muscle; the energy of ATP hydrolysis of one molecule is concentrated into a single vibration mode and is captured before it spreads to others. Therefore for the source, let $E_1 = 0.43$ n ev, the heat of reaction of n ATP molecules, and $N_1 = 2n$, corresponding to one vibration mode per molecule. This gives

$$r = \frac{310 \times 1.36 \times 10^{-16}}{0.43 \times 1.6 \times 10^{-12}} = 0.062, \tag{39}$$

from which (38) gives

$$\eta = 76.5\% \tag{40}$$

Doubtless, the near agreement with the value reported by Alberts et al (1983) is fortuitous; the existing measurements are too uncertain to draw any real conclusions. But one might have hoped that the maximum theoretical efficiency would come out just above the maximum observed efficiency; and at least that much has been realized. It appears that the information we used was adequate for the purpose, and there is no longer any mystery.

A CHECK

We derived the efficiency formula (38) without assuming any slow reversible operation as conventional thermodynamics does. On the other hand, neither did we assume that it is not slow, so if our derivation is correct, the formula ought to remain valid in the limit when the process is so slow that the conventional theory does apply. To check this, let us apply conventional theory to a small source whose temperature T_1 drops slowly as the engine runs, so we have a sequence of infinitesimal reversible Carnot cycles. Suppose that the sink is so large that T_2 remains constant. Then drawing heat Q_1 from the source, the maximum work we can get according to classical thermodynamics is

$$W(Q_1) = \int_0^{Q_1} \left[1 - \frac{T_2}{T_1(Q)}\right] dQ \quad . \tag{41}$$

Now suppose, corresponding to the Rayleigh-Jeans assumptions in our first derivation, that the source has a constant heat capacity C, so that $T_1(Q) = T_1 - Q/C$, where T_1 is the initial source temperature; then $E_1 = CT_1$. The engine will run only as long as $T_1(Q) > T_2$, so the maximum obtainable work is given when the upper limit of integration is $Q_1 = C(T_1 - T_2)$. Making these substitutions, the integral is easily evaluated, with the result

$$W_{max} = C\left[T_1 - T_2 + T_2 \log\frac{T_2}{T_1}\right] \quad . \tag{42}$$

Dividing by $E_1 = CT_1$, we recover the result (38) that we derived previously using only phase volume considerations. This confirms that our generalised second law reduces, as it should, to the conventional one when the latter is applicable.

But this conventional "slow, reversible" second law is not applicable to a muscle, because if a muscle operated slowly enough to make its assumptions valid, other degrees of freedom that we have left out of our calculation would take over and thermalize the primary energy, making the muscle nearly useless. It is just to avoid thermalization that biological processes must take place rapidly, and thus we require a "fast" second law to analyze them.

Our generalisation of the second law not only preserves the dynamics and therefore the Liouville theorem, it preserves the Clausius relation $S_f \geq S_i$ and the Boltzmann entropy formula $S = k \log W$; and it even preserves the intuitive meaning of it that was recognized by Boltzmann, Einstein, and Planck. Therefore we have not changed the basic rationale underlying the second law and the Kelvin efficiency rule in any way; we have only opened our eyes to their full meaning.

Far from being in conflict with the second law, Liouville's theorem is the reason for it. Had Liouville's theorem been discovered before the work of Carnot, it appears to us that the second law, in the full generality we have given it, might have been anticipated theoretically without any reference to heat engines; or indeed to the notions of temperature and thermal equilibrium. Note that we have made no use of the notions of order and disorder. Indeed, as Maxwell noted in the aforementioned article on diffusion, those notions are only expressions of human aesthetic judgments; Nature has no way of knowing what you or I consider "orderly". The second law limitation on macroscopic processes is easily understood, in physically meaningful terms, as the price we pay for *reproducibility*.

CONCLUSION

As those promised tentative comments on biological information, we see the above as evidence that the energy of ATP hydrolysis is confined to a single vibration mode in a muscle; if it spread to two modes, then we would have $r = 2 \times 0.062 = 0.124$, and (38) would predict a theoretical maximum efficiency of only 62%. Had the energy spread to ten vibration modes before being recaptured, the predicted efficiency would be only 8%. It appears that animals have indeed evolved muscle efficiency to the maximum that could be realized in a biochemical environment powered by the ATP hydrolysis reaction, although a reaction with a greater ΔH would permit still higher efficiency.

Finally, what was the effective upper temperature T_1 for the muscle? With two degrees of freedom per ATP molecule, this is given by $kT_1 = 0.43ev$, or

$$T_1 = \frac{0.43\times1.6\times10^{-12}}{1.36\times10^{-16}} = 5060 \text{ K} \tag{117}$$

This is startling because it is about the temperature at the surface of the sun! It appears, then, that a muscle is able to work efficiently not because it violates any laws of thermodynamics, but because it is powered by tiny "hot spots" of molecular size, as hot as the sun.

This shows how far a biological system is from thermal equilibrium in the respects that matter. If one says that the temperature in a living cell is "uniform", he can mean only that it is uniform as registered by a thermometer whose bulb is thousands of times coarser than the units that are performing the essential biological functions.

If we examine the current literature of bioenergetics with this in mind, we are struck by the fact that virtually all treatments begin by stating that biological systems are at uniform temperature and the chemical reactions proceed isothermally; then virtually all the discussion is in terms of reaction free energies ΔF or ΔG.

Now the free energy change of a reaction is only a fictitious kind of energy, that could in principle be observed in very special circumstances. It is the work made available when the temperature and concentrations are uniform and the reaction proceeds so slowly that it remains at equilibrium with respect to the original temperature and concentration; i.e. when heat can flow in or out of the cell rapidly enough, and the reactants and products can diffuse in and out rapidly enough, to maintain the initial uniformity. Conditions in a biological process such as nucleotide synthesis are about as far from this as can be imagined, in several respects:

(1) A cell may have very few (i.e., less than 20) molecules of a given type, and they are not free to diffuse about because of intracellular membranes; thus the uniform concentrations presupposed in the definition of reaction free energies seem not only not realized, but not even meaningful. Lehninger (1982) warns us that this might invalidate conventional thermodynamic treatments.

(2) A reaction is over -- the job is done -- in a time too short to reach equilibrium anyway. For many reactions the situation may be more nearly adiabatic than isothermal; thus the "real" physical energies ΔU, ΔH that have a meaning independently of thermal equilibrium, are the ones most relevant for biological processes.

(3) Hundreds of other reactions are going on simultaneously, and while they may not interfere directly with a reaction of interest, they must modify the environment in which that reaction takes place. On the scale of sizes and times that matter, a living cell is never in a state remotely like thermal equilibrium or uniform concentrations.

Recognizing this, we can understand another reason why biological thermodynamics has been puzzling in the past. Conventional free energy thermodynamics is doubtless adequate to describe slow, gross phenomena such as osmotic effects, but it may be irrelevant for biological functions like muscle contraction and protein synthesis, which necessarily, to avoid thermalization from the surroundings, take place rapidly and on the molecular scale.

As our treatment of muscle efficiency shows, the small scale does not in itself preclude the application of thermodynamics, but attempts to do this could not have succeeded until the above points were recognized and we had a quite different statement of the second law. Of course, muscle performance is only a special case of the general problem, but seeing how to apply the second law to muscle behaviour should give a useful clue for other cases.

In these first crude estimates to illustrate the principle, our reasoning was so general -- concerning only phase volumes -- that we did not need to invoke any particular details of the mechanism of muscle action. However, the myosin bridge mechanism for striated muscle contraction proposed by Sir A. F. Huxley (1957) and described by Squire (1981) and Alberts, *et al.* (1983) appears not only consistent with our speculations; it fits in very nicely with them. The bending of that bridge is a degree of freedom that corresponds to a low-frequency vibration mode for which the classical equipartition law would hold, and the relative stiffness and massiveness of the myosin head makes it seem well adapted to resisting rapid thermalization while transferring its energy into the macroscopic sliding of the actin fiber. We could hardly have asked for a better candidate for our one vibrational mode to receive the ATP hydrolysis energy.

Presumably, our argument could be refined by taking further information of this kind into account, although the observed facts of muscle performance suggest that the final conclusion cannot be very different; i.e. most of that information will be irrelevant for predicting the net efficiency, although it is highly relevant for predicting finer details such as force-velocity curves, fatigue, etc.

Having seen this biological mechanism, it is easy to believe that synthesized or extracted macromolecules could do similar things *in vitro.* Indeed, the first step in this direction has been taken already. In the fascinating "myosin motor" of Shimizu (1979) we have a molecular engine operating *in vitro*; not very efficiently, but nevertheless confirming the idea. In time the design of useful anti-Carnot molecular engines (artificial muscles) might become about as systematic and well understood as the design of dyes, drugs, and antibiotics is now.

REFERENCES

B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson (1983), *Molecular Biology of the Cell*, Garland Publishing Co., New York; pp. 550-609.

A. Aspect and P. Grangier (1985), "Tests of Bell's Inequalities ---", in *Symposium on the Foundations of Modern Physics; 50 years of the EPR Gedankenexperiment*, P. Lahti and P. Mittlestadt, Editors (World Scientific Publishing Co., Singapore).

A. Aspect (1986), "Tests of Bell's Inequalities with Pairs of Low Energy Correlated Photons", in Moore & Scully (1986).

L. E. Ballentine (1970), "The Statistical Interpretation of Quantum Mechanics", Rev. Mod. Phys. **42**, 358-381.

J. S. Bell (1964), "On the EPR Paradox", Physics **1**, 195-200,

J. S. Bell (1966), "On the Problem of Hidden Variables in Quantum Mechanics:, Rev. Mod. Phys. **38**, 447.

J. S. Bell (1987), *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press. Contains reprints of all of Bell's papers on EPR up to 1987.

N. Bohr (1935), "Can Quantum Mechanical Description of Reality be Considered Complete?", Phys. Rev. **48**, 696.

G. L. Bretthorst (1988), *Bayesian Spectrum Analysis and Parameter Estimation*, Springer Lecture Notes in Statistics, Vol. 48.

A. Einstein, B. Podolsky, and N. Rosen (1935), "Can Quantum Mechanical Description of Reality be Considered Complete?", Phys. Rev. **47**, 777.

A. F. Huxley (1957), Prog. Biophys. Chem. **7**, 255.

E. T. Jaynes (1965), ''Gibbs vs Boltzmann Entropies'', Am. J. Phys. **3**, 391-398. Reprinted in E. T. Jaynes, *Papers on Probability, Statistics and Statistical Physics*, R. D. Rosenkrantz, Editor, D. Reidel Publishing Co., Dordrecht-Holland (1983).

E. T. Jaynes (1985), "Generalized Scattering", in *Maximum Entropy and Bayesian Methods in Inverse Problems*, C. R. Smith & W. T. Grandy, Editors, D. Reidel Publishing Co., Dordrecht-Holland; pp. 377-398.

E. T. Jaynes (1986), "Predictive Statistical Mechanics", in Moore & Scully (1986); pp. 33-56.

H. Jeffreys (1931), *Scientific Inference*, Cambridge University Press; later editions 1957, 1973.

H. Jeffreys (1939), *Theory of Probability*, Oxford University Press; numerous later editions.

P. Knight (1987), "Single-atom Masers and the Quantum Nature of Light", Nature, **326**, 329.

A. L. Lehninger (1982), *Biochemistry, The Molecular Basis of Cell Structure and Function*, Worth Publishers, Inc., 444 Park Ave. South, New York; p. 383.

G. T. Moore & M. O. Scully, Editors (1986), *Frontiers of Nonequilbrium Statistical Physics*; Proceedings of the NATO Advanced Study Institute, Santa Fe, June 1984; Plenum Press, New York.

G. Rempe, H. Walther, N. Klein (1987); Phys. Rev. Let **58**, 353

S. Rozental, Editor (1964); *Niels Bohr, His Life and Work as seen by his Friends and Colleagues*, J. Wiley & Sons, Inc., New York.

H. Shimizu (1979), Adv. Biophys. **13**, 195-278.

H. Simon & N. Rescher (1966); "Cause and Counterfactual", Phil. Sci. **33**, 323 - 340.

C. P. Slichter, (1980), *Principles of Magnetic Resonance*, Springer, New York.

A. Zellner (1984), *Basic Issues in Econometrics*, Univ. Chicago Press; pp. 35-74.

FROM RATIONALITY AND CONSISTENCY TO BAYESIAN PROBABILITY

C. RAY SMITH
RESEARCH, DEVELOPMENT AND ENGINEERING CENTER
U.S. ARMY MISSILE COMMAND
REDSTONE ARSENAL, ALABAMA, U.S.A.
AND
GARY ERICKSON
DEPARTMENT OF ELECTRICAL ENGINEERING
SEATTLE UNIVERSITY, SEATTLE, WASHINGTON
AND PUGET SOUND POWER & LIGHT CO., BELLEVUE, WASHINGTON, U.S.A.

ABSTRACT

The presentation by Jaynes of Bayesian probability theory, among other things, served to unify and strengthen the earlier work of Cox and Polya. While the above approach to probability theory is well-known to many proponents of maximum-entropy and Bayesian methods, it deserves to be more widely promulgated and studied. This paper is a tutorial introduction to the Cox–Polya–Jaynes consistency and rationality requirements as the basis of Bayesian probability theory.

1. INTRODUCTION

Bayesian probability theory continues to be applied to many problems of a serious nature [Smith and Grandy, 1985; Justice, 1986; Smith and Erickson, 1987; Erickson and Smith, 1988; Bretthorst, 1987, 1988a, 1988b]. In addition to applying scientific knowledge, there exists a strong propensity in science to dig more deeply, to seek the foundations of that knowledge; so, inquiry into the foundations and rationale of the theory has proceeded apace.

For all practical purposes, the noncollaborative but synergistic efforts of Cox [1946], Polya [1954], and Jaynes [1957] furnish a highly compelling rationale for Bayesian probability theory. Rarely do we wish to say in science that anything is "final;" there is almost always room for refinement, extension and growth. But the work of Cox, Polya, and Jaynes was clearly a satisfactory basis from which to proceed. Most practitioners of maximum-entropy and Bayesian methods have studied — indeed savored — the original works, but too many have not had occasion to delve into them. We therefore present here a quick tutorial tour through these classics, although it must be said that condensation and interpretation by

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 29–44.

their very nature can subtly alter the perspective of an original work. We thus advise the interested reader to consult the original sources for the fullest impression of the authors' intent, tone and accomplishment.

Until very recently, the 1957 report by Jaynes was difficult to obtain. Now that Jaynes' approach to probability theory has become available [Erickson and Smith, 1988, **I**, Ch. 1], it is not reasonable to here reiterate his presentation. Rather, our emphasis will be on background, concepts, definitions and similar elements experience has shown to prove troublesome to many. In the beginning, we tend to avoid even the word probability so we can focus on the question of how far toward an inductive logic rationality and consistency can carry us.

2. PATTERNS OF DEDUCTIVE REASONING

We shall eventually be concerned with potential measures of the plausibility of one proposition given the truth of another. Our discussion will lead to Polya's inductive syllogism, which has its roots in the syllogisms of deductive logic. But we are getting ahead of the story; we must first supply some definitions and essential background.

Deductive reasoning is the inferring of specific conclusions from known principles or premises; the conclusions are unique. Deductive logic formalizes the processes of deductive reasoning by means of symbols and rules. Apparently, Aristotle was the first to observe that deductive reasoning follows definite patterns, called syllogisms. The syllogisms we shall focus on consist of a major premise, a minor premise and a conclusion. A premise is a proposition accepted as true (or false) from the beginning of a development. As used here, "proposition" has a specific meaning.

A proposition is an unambiguous statement which is, or will become, either true or false in the problem under consideration — in other problems, the statement may not be a proposition. It is important to note that in an environment in which our reasoning cannot be deductive, the actual validity of a proposition may not be accessible at the time an inference must be made.

We shall denote propositions by upper case letters like A, B, C, A_1, A_2. It is risky to illustrate propositions without stating the problems in which they are embedded, for, as noted above, one can often think of situations in which an example ceases to be a proposition. The following examples should be viewed with this caveat in mind:

$$A = \text{"On the next toss, a die will show 5 dots,"}$$

$$B = \text{"}\sum_{n=1}^{\infty} \frac{1}{n!} > 3\text{,"}$$

$$C = \text{"It will rain tomorrow."}$$

To a person who had never heard of e^x, some numerical experimentation may be needed to determine that B is false. For C to be a proposition, we must have

adopted a procedure for deciding whether rain has occurred (Is a light mist for a few seconds regarded as rain?).

Certain natural operations for propositions come to mind; these operations and the peculiar property of propositions to be either true or false can be used to establish an algebra of propositions, the subject of the next section of this paper. But at this time, we want to return to our unfinished discussion of syllogisms.

A deductive syllogism has the structure:

$$\begin{array}{c} \text{Major premise} \\ \underline{\text{Minor premise}} \\ \text{Conclusion} \end{array} \tag{1}$$

where the horizontal line plays the role of "therefore." One important syllogism with this structure, called the *modus ponens* (ponere = affirm), is the following:

$$\begin{array}{c} A \text{ implies } B \\ \underline{A \text{ true}} \\ B \text{ true} \end{array} \tag{2}$$

Obviously, if A implies B and A turns out to be true, then B must be true. As a concrete example of *modus ponens*, we take

$$\begin{aligned} A &= \text{“}1 \leq s \leq 4\text{,”} \\ B &= \text{“}1 \leq s \leq 11\text{.”} \end{aligned} \tag{3}$$

Here, s is a parameter in the problem under investigation.

The other deductive syllogism we consider is the *modus tollens* (tollere = deny):

$$\begin{array}{c} A \text{ implies } B \\ \underline{B \text{ false}} \\ A \text{ false} \end{array} \tag{4}$$

That this mode of deduction is valid may not be instantly obvious. Consider: If A *is* true, then B is true; but B is false, so A cannot be true. The propositions in Eq. (3) can be used in the *modus tollens*, as an example.

We are not finished with syllogisms; but to prepare for a more symbolic and quantitative approach, we give next a brief sketch of Boolean algebra.

3. BOOLEAN ALGEBRA

In his investigation of logic, George Boole [1854] developed an algebra of propositions whose importance has increased steadily since the late 1930's. This

Boolean algebra is an algebra of objects which can have only one of two possible values or states, such as numbers like 0 and 1, positions like up and down, and truth values like true and false. The applications of Boolean algebra to switching circuits and digital computers account for much of its current importance; our discussion will center on propositions. Our presentation of Boolean algebra is self-contained, though it is brisk and omits many standard topics.

If a proposition A is true, we say it has truth value $\underset{\sim}{1}$ and write $A = \underset{\sim}{1}$; similarly, if A is false its truth value is $\underset{\sim}{0}$ and $A = \underset{\sim}{0}$. The *negation* of the proposition A means *not* A and is denoted by $\overline{A}$; $\overline{A}$ has the truth value opposite that of A and as a proposition is

$$\overline{A} = \text{“}A \text{ is false.”} \tag{5}$$

For example, if A =“The mayor's house is white,” then $\overline{A}$ =“The mayor's house is not white” or $\overline{A}$ =“It is false that the mayor's house is white.” If B =“The mayor's house is yellow,” even though $B = \underset{\sim}{1}$ requires $A = \underset{\sim}{0}$ and hence $\overline{A} = \underset{\sim}{1}$, care must be exercised in relating B and $\overline{A}$ — for example, both A and B could be false.

Negation is called a *unary* operation, because a single proposition enters the operation. There are many *binary* operations which combine pairs of propositions. Two such operations are of special interest: the logical product and logical sum.

The logical *product* (or conjunction) AB is defined as follows:

$$AB = \text{“Both } A \text{ and } B \text{ are true.”} \tag{6}$$

The logical *sum* (or inclusive disjunction) of any two propositions A and B is denoted by $A + B$ and defined by

$$\begin{aligned} A + B = &\text{“Either } A \text{ is true, or } B \text{ is true, or} \\ &\text{both } A \text{ and } B \text{ are true.”} \end{aligned} \tag{7}$$

Note that AB and $A + B$ are themselves propositions, and are examples of compound propositions. Suppose A_n = “A tossed die shows n dots” and H = “A tossed coin shows a head.” Then, A_nH = “The die shows n dots *and* the coin shows a head,” while $A_n + H$ = “The die shows n dots, *or* the coin shows a head, *or* the die shows n dots *and* the coin shows a head.” Note that paraphrasing or rewording of propositions is permitted provided the meaning is not compromised. The definitions in Eqs. (6) and (7) show clearly that the logical product and logical sum are commutative — cf. Eq. (8) below.

One can employ the operations negation, logical product, and logical sum to construct logical or Boolean functions of propositions. We shall encounter a few Boolean functions below, but we must forgo any discussion of the theory of Boolean functions.

The set of propositions $A, B, C, \ldots, \underset{\sim}{1}, \underset{\sim}{0}$ along with the negation, product and sum operations form a *Boolean algebra* if the following properties (the so–called Huntington axioms) are satisfied [Whitesitt, 1961]:

$$A + B = B + A \quad (a) \qquad AB = BA \quad (b) \tag{8}$$

$$A + \underset{\sim}{0} = A \quad (a) \qquad A \cdot \underset{\sim}{1} = A \quad (b) \tag{9}$$

$$A + \overline{A} = \underset{\sim}{1} \quad (a) \qquad A\overline{A} = \underset{\sim}{0} \quad (b) \tag{10}$$

$$A(B + C) = AB + AC \quad (a) \qquad A + BC = (A + B)(A + C) \quad (b) \tag{11}$$

Note that there is a duality between the operations in the two columns above: If the sum is replaced by the product, the product by the sum, $\underset{\sim}{0}$ by $\underset{\sim}{1}$ and $\underset{\sim}{1}$ by $\underset{\sim}{0}$ in one column, the other column is produced. This principle of duality allows us to translate any Boolean equation into another valid equation. Also, literal notation is understood to apply: symbols for propositions in theorems and other expressions are to be regarded as variables, not fixed propositions — except where the contrary is indicated. The objects $\underset{\sim}{1}$ and $\underset{\sim}{0}$ are considered to be constants in Boolean algebra (they are constant propositions and constant functions).

The Huntington axioms are deliberately parsimonious and yield several important results only as theorems. The theorems are sometimes expected, but others may be surprising and very engaging. We list below several of the more useful theorems [for proofs, see Whitesitt, 1961]:

$$A + (B + C) = (A + B) + C \qquad A(BC) = (AB)C \tag{12}$$

$$A + AB = A \qquad A(A + B) = A \tag{13}$$

$$A + \underset{\sim}{1} = \underset{\sim}{1} \qquad A\,\underset{\sim}{0} = \underset{\sim}{0} \tag{14}$$

$$\overline{\underset{\sim}{0}} = \underset{\sim}{1} \qquad \overline{\underset{\sim}{1}} = \underset{\sim}{0} \tag{15}$$

$$\overline{A + B} = \overline{A}\ \overline{B} \qquad \overline{AB} = \overline{A} + \overline{B} \tag{16}$$

$$A + A = A \qquad AA = A \tag{17}$$

By considering specific verbal propositions, one can verify that propositions indeed satisfy Eqs. (8)–(17). In particular, the commutativity and associativity of the logical product and sum, Eqs. (8) and (12), appear to be completely trivial properties of propositions; yet, these properties play vital roles in Cox's consistency requirements discussed in our Sec. 5.

We now employ the preceding material to develop a more formal notation for deductive syllogisms than used earlier. There are many occasions in which one proposition is conditional on another in a manner that can be expressed by a Boolean equation. We use one such conditional relation to reflect "A implies B" in the syllogisms we consider. The equation

$$AB = A, \tag{18}$$

called the inclusion relation and read "A implies B," embodies some information (some facts) we have regarding A and B. Clearly, for $A = \underset{\sim}{1}$ this equation reduces

to $B = \underset{\sim}{1}$ (thus, A implies B); for $B = \underset{\sim}{0}$ it leads to $A = \underset{\sim}{0}$. These we recognize as the *modus ponens* and *modus tollens* in Eqs. (2) and (4). Using our newly acquired formalism, we can cast the *modus ponens* and *modus tollens* in the forms

$$\begin{array}{ll} AB = A & AB = A \\ \underline{A = \underset{\sim}{1}} & \underline{B = \underset{\sim}{0}} \\ B = \underset{\sim}{1} & A = \underset{\sim}{0} \end{array} \tag{19}$$

(It should be noted that some authors prefer to define the major premise "A implies B" in terms of a binary operation called the material implication.)

4. PATTERNS OF INDUCTIVE REASONING

So far, we have considered syllogisms with the major premise $AB = A$ and the minor premises $A = \underset{\sim}{1}$ and $B = \underset{\sim}{0}$. To complete the study, we need to examine the cases with the minor premises $A = \underset{\sim}{0}$ and $B = \underset{\sim}{1}$. For $A = \underset{\sim}{0}$ and $B = \underset{\sim}{1}$, our information about A and B that is contained in $AB = A$ produces no constraints on B and A, respectively. We can formalize this state of affairs by syllogisms, revealing the ostensible impasse:

$$\begin{array}{lll} AB = A & & AB = A \\ \underline{A = \underset{\sim}{0}} & (a) & \underline{B = \underset{\sim}{1}} \quad (b) \\ B = ? & & A = ? \end{array} \tag{20}$$

This is exactly the situation studied in depth by Polya [1954], so let us see how he resolved the impasse.

Note first of all, any solution of Eq. (20) must be inductive in nature (also, we must define what we mean by solution). Hence, we call the syllogisms in Eq. (20) *inductive syllogisms*, and to shorten our discussion, we concentrate on the second syllogism, (b). Polya engaged in a detailed study to demonstrate that in everyday problems with the structure of Eq. (20) there are identifiable patterns in our reasoning process. Next, we summarize some conclusions Polya reached in his analysis.

In arriving at the condition $AB = A$, we are often using only a portion of the information we have about the problem and about A and B. In the least, we know that B is a consequence of A (it is not suggested that the relationship between A and B must be causal). So, if one consequence of A, namely B, should turn out to be true, what can we infer about the truth value of A? We have inadequate information about A to make any definitive statement of its truth. However, compared to the situation in which the truth value of B is unknown, we have *inductive* evidence that A is true. We shall summarize this by saying: Compared to the case with the truth value of B unknown, learning that B is true *enhances the plausibility* that A is true. (Assuming that A is true, given

that B is true is committing the logical fallacy of assuming the consequent.) The word plausibility arose in earlier discussion. We use plausibility as shorthand for degree of plausibility and to mean something like: credibility, confidence, belief, or appearance of truth. (We are skirting words like probability, likelihood, etc. that have technical meaning.)

Now we are able to complete Eq. (20b):

$$\begin{array}{r} AB = A \\ B = 1\!\!\!\!\sim \\ \hline A \text{ more plausible} \end{array} \tag{21}$$

This result epitomizes the patterns of plausible reasoning revealed by Polya's study [Polya, 1954]. In no way have we been able to convey the depth and spirit Polya brought to bear on his inquiries. By and large, Eq. (21) expresses what we mean by *rationality*, though there remains some fine-tuning (e.g., in respect to continuity). Also, Eq. (21) will be taken as our *solution* of Eq. (20b), as clarified by further discussion. The essence of Eq. (21) is the commitment only to the direction, and not extent, of change of the degree of plausibility. A simple illustration will be given in a moment.

In analogy with deductive logic, inductive logic will use symbols and rules to formalize the processes of inductive reasoning. The symbol E will always represent all of our *prior information* (we use E as if it were a proposition, even though it may be cumbersome to express some information as a proposition). In the present context, E includes everything we know about the problem, in particular whether the condition $AB = A$ is based on a causal connection between A and B. The symbol $u(A|C)$ will be employed to represent the plausibility of A given C (because A and C are variables, we prefer not to say explicitly "the plausibility of A true, given that C is true"). Thus $u(A|E)$ is the plausibility of A given only our prior information E, and $u(A|BE)$ is the plausibility of A given E and the datum B. It appears that we are allowing plausibility to evolve into some numerical object; we need to relate what our plans are.

Before the subject of inductive reasoning can become mathematical in nature, we must identify some quantifiable attributes of plausibilities. If plausibilities are to serve any useful purpose, they must at least have the capacity for describing rationality. This requirement will be met if the plausibilities are capable of (1) varying in a continuous fashion and (2) exhibiting inequality (as well as equality). The simplest means for incorporating these attributes is to associate real numbers with plausibilities. At this time, we are concerned only with identifying a minimal set of attributes; other properties of plausibilities (e.g. the rules they obey) will emerge as a result of the mathematical treatment of the subject.

Using the plausibility as just defined, we can write for Eq. (21)

$$\begin{array}{c} AB = A \\ B = \mathbb{1} \\ \hline u(A|BE) \geq u(A|E) \end{array} \tag{22}$$

In summary, rationality, as so far explicated, requires only that the plausibility not decrease in the face of inductive evidence. Nothing is said about the strengths or magnitudes of plausibilities. The association of larger numbers with greater plausibilities is clearly an inessential convention: it is consistent with our choosing the word "plausibility" (instead of, say, implausibility).

The concept of rationality carries with it a continuity requirement whose meaning will become clearer in the next few sections. For the moment suffice it to say that small changes in plausibilities insinuate only small changes in their numerical representatives.

Most of the conceptual work is behind us. Jaynes [1957] has shown how to translate Polya's rationality and Cox's consistency requirements into desiderata and then to construct a mathematical theory. Before undertaking this task, we give the example promised earlier.

Suppose the propositions in Eq. (3) refer to the radius, s, of some asteroid and suppose crude measurements give us the range of s: $1 \leq s \leq 37$km. We have summarized the prior information, E. The range in proposition A is predicted by some far–out astrophysical theory, and we want very much to test the theory by testing the hypothesis A. While thinking about A we learn B is true. Compared to the situation in which we knew only that s is in the range 1 to 37km, the information $B = \mathbb{1}$ enhances (or at least does not diminish) the plausibility that A is true. This is the central idea Eq. (22) is supposed to convey.

5. THE DESIDERATA

Curiously, Cox [1946, 1961] and Polya [1954] seem never to have encountered each other's work, missing perhaps the stimulation to bring their own works to a more satisfactory and complete form. Jaynes [1957] consolidated the Cox and Polya contributions, supplying additional elements to arrive at the approach to probability theory we are presenting here. The term "consistency" as associated with Cox refers to the requirement that plausibilities be consistent with Boolean algebra. Jaynes developed desiderata which embody rationality and consistency, the latter having a broader significance than that employed by Cox. The desiderata are then applied to create mathematical conditions which the plausibilities must satisfy. This is the program attended to in this section.

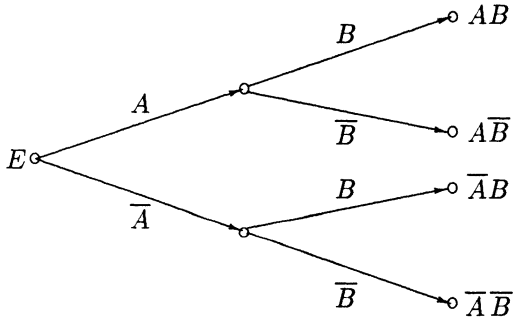
First, a brief comment on the word "desideratum." The primary function of the *Desiderata* is to describe the essential features which the plausibilities must embody. They do not stipulate that the plausibilities are to satisfy specific rules or axioms.

Desideratum 1. *The numerical measures of plausibilities are real numbers.*

Desideratum 2. *Plausibilities must exhibit qualitative agreement with rationality.* As new information supporting the truth of a proposition is supplied, the number which represents the plausibility will increase continuously and monotonically. It is understood that as the plausibility of a proposition increases, the plausibility of its negation will decrease continuously and monotonically. Also, to maintain rationality, the deductive limit must obtain where appropriate. The continuity requirement will be applied to mathematical functions introduced below. Finally, we hope it is unnecessary to remark that all pertinent information is to be used in the course of any analysis of plausibilities.

Desideratum 3. *All rules relating plausibilities must be consistent.* If a result can be derived in more than one way, all legitimate operations on the propositions must lead to the same result. One is free to establish the truth value of a compound proposition by examining the individual propositions in any possible sequence allowed by Boolean algebra (this evaluation need not follow the physical or causal relationships of the propositions.) The final results must be independent of the sequence actually employed.

We investigate next the implications of the desiderata on the plausibility $u(AB|E)$, specifically to determine how $u(AB|E)$ depends upon the plausibilities of A and B. The truth value of AB will be determined by examining first A, then B, as described by the following tree diagram:



The proposition AB can be reached only in the upper branch of the diagram. Thus, the plausibility $u(AB|E)$ depends only on the plausibilities $u(A|E)$ and $u(B|AE)$ — under the circumstances depicted in the tree diagram. We express this dependence as

$$u(AB|E) = F[u(A|E), u(B|AE)], \tag{23}$$

where F represents an unknown function. However, F is not arbitrary: It must maintain the desiderata, and it is the consequences of this that we explore next.

One important result follows easily from the commutativity of the logical product, $AB = BA$. The consistency requirement imposed on Eq. (23) requires the invariance of that expression under interchange of A and B:

$$u(AB|E) = u(BA|E) = F[u(B|E), u(A|BE)]. \tag{24}$$

This does not tell us a great deal about F, however. The associativity of the logical product will be more prolific.

Writing

$$ABC = (AB)C = A(BC) \tag{25}$$

and treating (AB) as one proposition C as the other, then A and (BC) similarly, we obtain from Eq. (23)

$$\begin{aligned} u(ABC|E) &= F[u(AB|E), u(C|ABE)] \\ &= F[u(A|E), u(BC|AE)]. \end{aligned} \tag{26}$$

Applying Eq. (23) to $u(AB|E)$ and $u(BC|E)$ in these expressions leads to

$$F\{F[u(A|E), u(B|AE)], u(C|ABE)\} = F\{u(A|E), F[u(B|AE), u(C|ABE)]\}. \tag{27}$$

Finally, the notation

$$x = u(A|E), \quad y = u(B|AE), \quad z = u(C|ABE) \tag{28}$$

allows us to write Eq. (27) in the form

$$F[F(x, y), z] = F[x, F(y, z)], \tag{29}$$

which is a functional equation known, appropriately, as the associativity equation. As we shall see eventually, this equation determines uniquely the class of functions F which may be associated with plausibilities.

By assuming that F is twice differentiable in both variables, Cox derived from Eq. (29) a differential equation which he then solved. Some fuzzy set advocates have pounced upon this assumption as invalidating Cox's theory, in evident ignorance of the work of Aczél [1966, 1987], who derives the same general solution without assuming differentiability. The earlier book by Aczél provides an extensive bibliography on the associativity equation, starting with Abel [1826] who solved this equation under the condition that it is symmetric in the independent variables x, y, z. The result, as expressed by Aczél, is that all solutions of the associativity equation may be constructed from any continuous, strictly increasing monotonic function $G(x)$ as follows:

$$F(x, y) = G^{-1}[G(x)\, G(y)]. \tag{30}$$

It is easy to verify that this form does indeed satisfy Eq. (29).

Upon reinstating the original variables, Eq. (28), we have from Eqs. (23) and (30)

$$G[u(AB|E)] = G[u(A|E)]\, G[u(B|AE)]. \tag{31}$$

We incur no loss of generality by using the simpler notation (for u itself is arbitrary):

$$v(A|E) = G[u(A|E)]. \tag{32}$$

Then, Eq. (31) along with Eq. (24) give us the product rule for plausibilities:

$$v(AB|E) = v(A|E)\, v(B|AE) \tag{33a}$$

$$= v(B|E)\, v(A|BE). \tag{33b}$$

Our desiderata, specifically with the commutativity and associativity of the logical product, have produced a very impressive result: Irrespective of the scale used, plausibilities must abide by the rules in Eqs. (33a) and (33b). On the question of scales of plausibilities, see Jaynes [1957]. The generality afforded by G is explored by Tribus [1969], pp. 19 and 26–29.

Cox [1946, 1961] adopted Eq. (23) as an axiom, and we adopted it, perhaps too quickly, by taking advantage of the freedom allowed by the desiderata to look at a specific case. Does Eq. (23) describe only the specific case we examined or is it completely general [Eq. (24) is treated as equivalent to Eq. (23)]? The book by Tribus [1969], pp. 14–18, discusses all functional relations that the problem allows. The conclusion is that Eq. (23), and no other functional relation, describes the general case.

Though the scale of plausibility is quite arbitrary (and will remain so, much as the relationship of temperature scales with thermodynamic relations), Eq. (33) already fixes numerical values of plausibilities (more correctly, $v(A|E) = G[u(A|E)]$, where $u(A|E)$ is the plausibility) in the deductive limits we discuss next.

We consider first the extreme case in which our prior information E dictates that A is true ($A = \mathbb{1}$), while B remains arbitrary (provided it does not contradict E) — cf. the *modus ponens* in Eq. (19). Using $AB = \mathbb{1} \cdot B = B$ and $v(A|BE) = v(\mathbb{1}\,|BE) = v(\mathbb{1}\,|E)$ [A is already true by E, so $v(\mathbb{1}\,|BE)$ is conditional only on E] in Eq. (33a), we see that

$$v(B|E) = v(B|E)\, v(\mathbb{1}\,|E). \tag{34}$$

Because B is arbitrary, there are cases in which $v(B|E) \neq 0$; in such cases, the solution of Eq. (34) is

$$v(\mathbb{1}\,|E) = 1. \tag{35}$$

Thus, the certain proposition has a plausibility equal to unity — we continue to refer to $v(A|E)$ as the plausibility of A to avoid more terminology; cf. Eqs. (44)—(49) and the discussion following Eq. (49).

The other limiting case we consider obtains in a problem in which E informs us that B is false ($B = \mathbb{0}$) and A is compatible with E and B but is otherwise arbitrary. This limit corresponds to *modus tollens* in Eq. (19). Using $AB = A \cdot \mathbb{0} = \mathbb{0}$ and $v(B|AE) = v(\mathbb{0}\,|AE) = v(\mathbb{0}\,|E)$ in Eq. (23) gives

$$v(\mathbb{0}\,|E) = v(A|E)\, v(\mathbb{0}\,|E). \tag{36}$$

For $v(A|E)$ arbitrary, this equation has two solutions: $v(\mathbb{0}\,|E) = 0$ and $v(\mathbb{0}\,|E) = \infty$. Because A is arbitrary, it too can have truth value $\mathbb{0}$; this precludes $v(\mathbb{0}\,|E) =$

$-\infty$ as a solution. So we have a choice: $0 \leq v \leq 1$ or $1 \leq v \leq \infty$. Strictly as a convention we choose

$$v(\underset{\sim}{0}\,|E) = 0, \tag{37}$$

because of its accord with the word plausibility and because the other choice involves an infinite quantity which cannot be realistically implemented in hardware or software. The two solutions can be interchanged by replacing v with v^{-1} [Jaynes, 1957].

Up to this point, we have used only the logical product in investigating the consequences of the desiderata on the plausibilities. Naturally, one would turn next to the negation and the logical sum to explore their relationship with the plausibilities. In fact, we need consider only the negation, because the logical product and negation, forming an exhaustive set of operations (meaning all others can be represented by this set), already subsume the logical sum [cf. Eq. (16)]:

$$A + B = \overline{\overline{A}\,\overline{B}}. \tag{38}$$

There are other exhaustive sets of logical operations, but it is not appropriate to consider them here [Whitesitt, 1961].

Our information about the plausibility of $\overline{A}$ is identical to that for A; note too that A and $\overline{A}$ are exhaustive and mutually exclusive ($A + \overline{A} = \underset{\sim}{1}$, $A\,\overline{A} = \underset{\sim}{0}$). Thus, by the requirement of rationality, $v(A|E)$ and $v(\overline{A}\,|E)$ must be related by some function $T(x)$ as follows:

$$v(\overline{A}\,|E) = T[v(A|E)]. \tag{39}$$

The rationality requires further that $T(x)$ be a monotonic decreasing function of x (x represents a general argument). Moreover, because $A = \overline{\overline{A}}$, we see that

$$T^{-1}(x) = T(x), \tag{40}$$

so $T(x)$ is self-reciprocal. However, this condition is not stringent enough to completely characterize $T(x)$, for it must also be compatible with the product rule, Eq. (33). By considering $v(AB|E)$ for the particular propositions satisfying the conditional relation $\overline{A}\;\overline{B} = \underset{\sim}{0}$, one can show that $T(x)$ satisfies the functional equation [see Cox, 1946 and Jaynes, 1957]:

$$xT[T(y)/x] = yT[T(x)/y]. \tag{41}$$

Cox [1946] solved this equation by deriving from it a second-order differential equation, the solution of which is given by

$$[T(x)]^n + x^n = 1, \tag{42}$$

where Eq. (40) has been taken into account and where n is arbitrary except $n \neq 0$. Again, Aczél [1963] derives the same general solution without assuming differentiability. Finally, for our convention $v(\underset{\sim}{0}\,|E) = 0$, Eq. (37), only $n > 0$ is allowed. If we let $x = v(A|E)$ in Eq. (42) and take Eq. (39) into account, we obtain

$$[v(A|E)]^n + [v(\overline{A}\,|E)]^n = 1, \tag{43}$$

which provides the desired relation between $v(A|E)$ and $v(\overline{A}\,|E)$.

6. THE RULES OF BAYESIAN PROBABILITY THEORY

We can eliminate the appearance of the constant n in Eq. (43) by introducing a new function $p(A|E)$, where

$$p(A|E) = [v(A|E)]^n. \tag{44}$$

But, referring to Eq. (32), we see also that

$$p(A|E) = \{G[u(a|E)]\}^n. \tag{45}$$

Hence, the arbitrariness in G renders superfluous the dependence of $p(A|E)$ on n; that is, the dependence on n in Eq. (45) can be absorbed by G. Raising Eqs. (33), (35) and (37) to the power n and using Eq. (43), we obtain

$$p(AB|E) = p(A|E)\,p(B|AE) \tag{46a}$$

$$= p(B|E)\,p(A|BE) \tag{46b}$$

$$p(A|E) + p(\overline{A}\,|E) = 1 \tag{47}$$

$$p(\mathfrak{1}\,|E) = 1\ ,\ p(\mathfrak{0}\,|E) = 0. \tag{48}$$

We have noted already in Eq. (38) that the logical sum can be defined in terms of the logical product and negation operations. Using Eqs. (38), (46) and (47), one can show in a few lines

$$p(A + B|E) = p(A|E) + p(B|E) - p(AB|E). \tag{49}$$

The rewards of our quest are before us: Eqs. (46)–(49) represent the quantitative rules we sought. These rules are unique: *Any rules which represent degrees of plausibility by real numbers and conflict with them will necessarily violate rationality or consistency.* Of course, it makes no difference whether one calls that real number a likelihood, a significance level, a degree of membership in a set, or anything else.

In Eqs. (46), (47) and (49) we have three rules for plausibilities, along with the limiting values in Eq. (48), which coincide with those of probability theory. However, our derivation of these rules was based on rationality and consistency requirements, with no reference to sets, counting, frequencies or mass phenomena. Bearing this in mind, we shall refer henceforth to $p(A|E)$ as the *probability* of A, given the prior information E. In any line of reasoning in accord with the desiderata — and we know of no valid inductive reasoning at variance with the desiderata — one is entitled to the broadest possible interpretations and applications of the theory allowed by these foundations. This is not a trivial point, for the actual numerical values one uses for the probabilities are often reflective of the broad interpretation in which probabilities encode any information we are clever enough to use — a few additional comments on numbers are provided in the next

section. The meaning of the title of this paper should be clear, once we note that the theory we have just developed is called *Bayesian probability theory.* Of course, a full understanding of Bayesian probability theory will require an in–depth study of its further development and applications — for a start, consult the relevant books and papers in the references at the end of this paper, especially those of Jaynes.

We would be remiss were we to omit one final formal result. The commutativity of logical products as embodied by Eqs. (46a) and (46b) leads immediately to

$$p(A|BE) = \frac{p(A|E)\,p(B|AE)}{p(B|E)}. \tag{50}$$

This important expression is known as *Bayes' Theorem* (or Bayes' Rule). In light of the preceding discussion, it is not a stunning result. Nonetheless, it is astonishingly useful in applications. It affords the means for using datum B (which may be a compound proposition) to update the prior probability $p(A|E)$ to the posterior probability $p(A|BE)$ for A, incorporating both the datum B and our prior information, E.

7. CONCLUDING REMARKS

This paper is a brief tutorial on the foundations of Bayesian probability theory. The space allocated to this tutorial does not allow us to explore several key issues we suspect will occur to many readers. So, we want to close with a few remarks which may point the reader in a direction for resolving the issues.

First, the probability $p(A|E)$ we have arrived at is an arbitrary (positive, strictly increasing monotonic) function of the plausibility $u(A|E)$, itself a quantity that is anthropomorphic and intuitive. Moreover, the necessity and role of the prior information E in relation to $p(A|E)$ [$p(A|E)$ is conditional on E, and $p(A)$ has no meaning] deserve more discussion. Both of these points are discussed at length in Jaynes [1957], Secs. 2 and 3 (see especially pp. 8-11).

The next question that arises is: Where do the numerical values of Bayesian probabilities come from? The most fruitful methods are the principle of insufficient reason, the principle of maximum entropy and the symmetry principles based on transformation groups. These methods are developed from first principles in Jaynes [1957], Secs. 2 through 6 — see also Jaynes [1968] and [1983]. The subject of actual Bayesian probabilities brings to mind the *Jeffreys prior* and other pioneering work of Sir Harold Jeffreys [1939]. One can carry out estimations of parameters by direct application of Bayes' theorem, Eq. (50) — important and impressive results using this approach are achieved by Bretthorst [1987, 1988a, 1988b].

ACKNOWLEDGMENTS

We are grateful to Dr. John Skilling for the invitation to present this paper at the Eighth International Maximum Entropy Workshop, St. John's College, Cam-

bridge, England. The hospitality of John, Jennifer and Martin Skilling will always be dear to us.

The critical reading of and suggestions for improving this paper by Professor Edwin T. Jaynes add to our considerable indebtedness to him. We thank Dr. Larry Bretthorst, Mr. Mark Smith, Mrs. Carolyn Pardue, Dr. Harriet Shaklee and Dr. Rama Inguva for reading the manuscript. Thanks are also due Professor J. Aczél for corresponding with the first author and providing suggestions concerning the functional equations.

REFERENCES

Abel, N.H. (1826), *J. reine und angew. Math.* (Crelle's Jour.) **1**, pp. 11–15.

Aczél, J. (1963), 'Remarks on Probable Inference,' *Ann. Univ. Sci. Budapest. R. Eötvös Sect. Math.* **6**, pp. 3– 11.

Aczél, J. (1966), *Lectures on Functional Equations and Their Applications*, Academic Press, New York.

Aczél, J. (1987), *A Short Course on Functional Equations*, D. Reidel, Dordrecht.

Boole, George (1854), *An Investigation of the Laws of Thought*, Macmillan, London. Reprinted by Dover Publications, New York (1958).

Bretthorst, G.L. (1987), 'Bayesian Spectrum Analysis and Parameter Estimation,' Ph.D. Thesis, Washington University, St. Louis, Missouri. Available from University Microfilms, Ann Arbor, Michigan.

Bretthorst, G. L. (1988a), 'Excerpts from Bayesian Spectrum Analysis and Parameter Estimation,' in Erickson and Smith [1988, **I**], pp. 75–145.

Bretthorst, G.L. (1988b), 'Bayesian Spectrum Analysis and Parameter Estimation,' *Lecture Notes in Statistics* **48**, Springer–Verlag, Berlin.

Cox, R.T. (1946), 'Probability, Frequency and Reasonable Expectation,' *American Journal of Physics* **14**, pp. 1–13.

Cox, R.T. (1961), *The Algebra of Probable Inference*, The Johns Hopkins Press, Baltimore, Maryland.

Erickson, Gary J., and C. Ray Smith, eds. (1988), *Maximum-Entropy and Bayesian Methods in Science and Engineering.* **I.** *Foundations* and **II.** *Applications*, Kluwer Academic Publishers, Dordrecht.

Jaynes, E.T. (1957), 'How Does the Brain Do Plausible Reasoning?' Stanford University Microwave Laboratory Report 421. Reprinted in Erickson and Smith [1988, **I**], pp. 1–23.

Jaynes, E.T. (1968), 'Prior Probabilities,' *IEEE Trans. Syst. Sci. Cybern.* **SSC–4**, pp. 227–241. Reprinted in Jaynes [1983].

Jaynes, E.T. (1983), *Papers on Probability, Statistics and Statistical Physics*, R. D. Rosenkrantz, ed., D. Reidel, Dordrecht.

Jeffreys, H. (1939), *Theory of Probability*, Oxford. Second edition 1948, third edition 1961.

Justice, J.H., ed. (1986), *Maximum Entropy and Bayesian Methods in Applied Statistics*, Cambridge University Press, Cambridge.

Polya, George (1954), *Patterns of Plausible Inference* (Vol **II** of *Mathematics and Plausible Reasoning*), Princeton University Press, Princeton, New Jersey.

Smith, C. Ray, and G.J. Erickson, eds. (1987), *Maximum-Entropy and Bayesian Spectral Analysis and Estimation Problems*, D. Reidel, Dordrecht.

Smith, C. Ray, and W.T. Grandy, Jr., eds.(1985), *Maximum-Entropy and Bayesian Methods in Inverse Problems*, D. Reidel, Dordrecht.

Tribus, Myron (1969), *Rational Descriptions, Decisions and Designs*, Pergamon Press, New York.

Whitesitt, J.E. (1961), *Boolean Algebra*, Addision-Wesley Publishing Co., Reading, Massachusetts.

CLASSIC MAXIMUM ENTROPY

John Skilling
Dept. of Applied Mathematics
and Theoretical Physics
Silver Street
Cambridge CB3 9EW, U.K.

Abstract

This paper presents a fully Bayesian derivation of maximum entropy image reconstruction. The argument repeatedly goes from the particular to the general, in that if there are general theories then they must apply to special cases. Two such special cases, formalised as the "Cox axioms", lead to the well-known fact that Bayesian probability theory is the only consistent language of inference. Further cases, formalised as the axioms of maximum entropy, show that the prior probability distribution for any positive, additive distribution must be monotonic in the entropy. Finally, a quantified special case shows that this monotonic function must be the exponential, leaving only a single dimensional scaling factor to be determined a posteriori. Many types of distribution, including probability distributions themselves, are positive and additive, so the entropy exponential is very general.

The following paper (Gull 1989) applies these ideas to image reconstruction, showing how a sophisticated treatment can incorporate prior expectation of spatial correlations.

1. Introduction

There is a simple mode of reasoning - compelling or infuriating according to one's point of view - which allows us to construct general theories. It is this. If there is a general theory at all, it must apply to particular cases. In particular, if we already know the answer for a simple case, this constrains the general theory by falsifying all those which give wrong answers. If enough such cases can be found to constrain the general theory completely, then there will be no freedom left, and the theory will have been fully defined.

In experimental science, general theories usually have some "loose ends", such as relativistic corrections to Newtonian dynamics, or cosmological terms in general relativity, which are allowed by the experimental errors on the constraining observations. However, in an argument about logic, there can be no such loose ends.

We use this mode of reasoning three times here, leading successively to Bayesian probability theory, to maximum entropy (MaxEnt), and finally to a quantified prior for images. Firstly, if there is a general language for inference, it must be that of ordinary probability theory, with our inferences being quantified as probabilities: the proof is due to Cox (1946). However, probability theory describes how we must modulate our inferences in the light of evidence, but it does not tell us how to assign the prior probabilities which we need in order to start the scheme. Secondly, if there is a general way of assigning positive additive distributions (such as probability distributions), then it must be MaxEnt: one source for the original form of this argument is Shore and Johnson (1980), though we generalise away from unit normalisation.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 45–52.

However, this argument does not address the reliability of a MaxEnt distribution, in the sense of quantifying how much worse (i.e. less probable) a different distribution might be. Thirdly, if there is a general quantification of distributions, their probabilities must be exponential in their entropy (scaled by some factor which can not be fixed a priori). This work builds upon that of Frieden (1972), Gull and Daniell (1979) and Jaynes (1986), supersedes all the author's previous writings on the subject, and completes what we call "Classic" MaxEnt.

Of course, at any stage it remains possible that there is no general theory, in which case the argument would break down. Different problems would need different theories. Although it is a sociological fact that different problems are indeed currently analysed in a multitude of different ways, the author knows of no example in which a correct application of classical Bayesian methods would give a demonstrably incorrect result. In the absence of such contrary evidence, we shall avoid the Babel of Tongues by assuming that there are general theories.

Far from being restricted and thereby impoverished, the Classic MaxEnt formulae in fact allow wide freedom, and the rigour of the underlying mathematics can be turned to advantage. The following paper (Gull 1989) introduces a sophisticated new use of Classic MaxEnt in the realm of image reconstruction, and we surmise that more such developments will arise in the future.

2. Bayesian probability theory: The Cox axioms.

One of the principal aims of science is to enable us to infer the plausible outcomes of different situations, and thereby help us to predict the future, and to understand the past. Logical reasoning, aided by mathematics, is the principal intellectual tool we bring to bear upon this central problem of inference.

Whatever the content of our discussions, be it Raman spectroscopy or Roman history, we wish to be able to express our preferences for the various possibilities i,j,k,... before us. A minimal requirement is that we be able to rank our preferences consistently (i.e. transitively)

$$\text{(Prefer i to j) AND (Prefer j to k)} \Rightarrow \text{(Prefer i to k)} . \qquad [1]$$

Any transitive ranking can be mapped onto real numbers, by assigning numerical codes P(i), P(j),... such that

$$P(i) > P(j) \iff \text{(Prefer i to j)} . \qquad [2]$$

Now, if there is a common general language, it must apply in simple cases. Cox (1946) formulated two such simple cases as axioms, which we restate briefly. It is difficult to argue against either.

Axiom A:
If we first specify our preference for i being true, and then specify our preference for j being true (given i), then we have implicitly defined our preference for i and j together.

Axiom B:
If we specify our preference for i being true, then we have implicitly specified our preference for its negation ~i.

As a consequence of these remakably mild requirements, Cox showed that there is a mapping of the original codes P into other codes pr that obey the usual rules of probability theory

$$pr(i,j|h) = pr(i|h)\ pr(j|i,h) \qquad [3]$$

$$pr(i|h) + pr(\sim i|h) = 1\ . \qquad [4]$$

Many authors have repeated this proof in greater or lesser detail, as in the preceding paper (Smith and Erickson 1989).

Therefore, <u>if</u> there is a common language, then it can only be this one, and in accordance with historical precedent set by Bernoulli and Laplace (Jaynes 1978) we call the codes pr thus defined "probabilities". Logically, of course, there may be no common language. There may be a lurking "Axiom C", just as convincing as Axioms A and B, which contradicts them. Although much effort has been expended on such arguments (Klir 1987), no such contradictory axiom has been demonstrated to our satisfaction, and accordingly we submit to the Bayesian rules.

Bayes' Theorem itself, which is a simple corollary of these rules, then tells us how to <u>modulate</u> probabilities in accordance with extra evidence. It does not tell us how to <u>assign</u> probabilities in the first place. It turns out that such prior assignments should be accomplished by MaxEnt.

3. The axioms of maximum entropy.

The probability distribution pr(x) of a variable x is an example of a positive, additive distribution. It is positive by construction. It is additive in the sense that the overall probability in a domain D equals the sum of the probabilities in any decomposition into sub-domains, and we write it as $\int_D pr(x)dx$. It also happens to be normalised, $\int_{all\ x} pr(x)dx = 1$.

Another example of a positive, additive distribution is the intensity or power f(x,y) of incoherent light as a function of position (x,y) in an optical image. This is positive, and additive because the integral $\iint_D f(x,y)dxdy$ represents the physically meaningful power in D. (By contrast, the amplitude of incoherent light, though positive, is not additive.) For brevity, we shall call a positive, additive distribution a "PAD".

It turns out to be simpler to investigate the general problem of assigning a PAD than the specific problem of assigning a probability distribution, which carries the technical infelicity of normalisation. Accordingly, we investigate the assignment of a PAD f(x), given some definitive but incomplete constraints on it: such constraints have been called "testable information" by Jaynes (1978). Now <u>if</u> there is a general rule for assigning a single PAD, <u>then</u> it must give sensible results in simple cases. The four "entropy axioms" - so-called because they lead to entropic formulae - relate to such cases. Shore and Johnson (1980) and Tikochinsky, Tishby and Levine (1984) give related derivations pertaining to the special case of probability distributions. Proofs of the consequences of the axioms as formulated below appear in Skilling (1988),

though our phraseology improves upon that paper.

Axiom I: "Subset Independence"
Separate treatment of individual separate distributions should give the same assignment as joint treatment of their union.
More formally, if constraint C_1 applies to $f(x)$ in domain $x \in D_1$ and C_2 applies to a separate domain $x \in D_2$, then the assignment procedure should give

$$f[D_1|C_1] \cup f[D_2|C_2] = f[D_1 \cup D_2 | C_1 \cup C_2] \ , \qquad [5]$$

where $f[D|C]$ means the PAD assigned in domain D on the basis of constraints C.

Consequence: The PAD f should be assigned by maximising over f some integral of the form

$$S(f,m) = \int dx\ m(x)\ \Theta(f(x),x) \ . \qquad [6]$$

Here Θ is a function, as yet unknown, and m is the Lebesgue measure associated with x which must be given before an integral can be defined. The effect of this basic axiom is to eliminate all cross-terms between different domains.

Axiom II: "Coordinate invariance"
The PAD should transform as a density under coordinate transformations.

Consequence: The PAD f should be assigned by maximising over f some integral of invariants

$$S(f,m) = \int dx\ m(x)\ \phi(f(x)/m(x)) \ , \qquad [7]$$

where ϕ is a function, as yet unknown. The crucial axiom is the next.

Axiom III: "System independence"
If a proportion q of a population has a certain property, then the proportion of any sub-population having that property should properly be assigned as q.

For example, if 1/3 of kangaroos have blue eyes (Gull and Skilling 1984), then the proportion of left-handed kangaroos having blue eyes should also be assigned the value 1/3.

Consequence: The only integral of invariants whose maximum always selects this assignment, regardless of any other subdivisions which may be present, is

$$S(f,m) = -\int dx\ f(x)\ \log(f(x)/cm(x)) \ , \qquad [8]$$

where c is a constant.

Axiom IV: "Scaling"
In the absence of additional information, the PAD should be assigned equal to the given measure (instead of being merely proportional). This is a practical convenience rather than a deep requirement.

Consequence: The PAD f should be assigned by maximising over f

$$S(f,m) = \int dx\ (\ f(x) - m(x) - f(x)\ \log(f(x)/m(x))\ ) \ . \qquad [9]$$

The additive constant $\int m dx$ in this expression ensures that the global maximum of S, at $f(x)=m(x)$, is zero, which is both convenient and required for other purposes (Skilling 1988).

Because of its entropic form, we call S as defined in [9] the entropy of the positive, additive distribution f. It reduces to the usual cross-entropy formula $-\int dx\ f \log(f/m)$ if f and m happen to be normalised, but is actually more general. (Holding that the general concept should carry the generic name, we deliberately avoid [9] a qualified or personalised name.)

We see that MaxEnt is the only method which gives sensible results in simple cases, so if there is a general assignment method, it must be MaxEnt. (Logically, there may be a lurking, contradictory "Axiom V", but we have not found one, and accordingly we submit to this "principle of maximum entropy".) Two major applications follow from this analysis. Firstly, MaxEnt is seen to be the proper method for assigning probability distributions pr(x), given testable information. Secondly, in practical data analysis, if it is agreed that prior knowledge of a PAD satisfies axioms I-IV, and if testable information is given on it, then any single PAD to be assigned on this basis must be that given by MaxEnt.

However, the arguments above do not address the reliability of the MaxEnt assignment: would a slightly different PAD be very much inferior?. Furthermore, experimental data are usually noisy, so that they do not constitute testable information about a PAD f. Instead, they define the likelihood or conditional probability $pr(data|f)$ as a function of f. In order to use this in a proper Bayesian analysis, we need the quantified prior probability pr(f) - or strictly $pr(f|m)$ because we have needed to set a measure m.

4. Quantification.

The reliability of an estimate is usually described in terms of ranges and domains, leading us to investigate probability integrals over domains V of possible PADs f(x), digitised for convenience into r cells as $(f_1,f_2,\ldots,f_r)$.

$$pr(f\in V|m) = \int_V d^r f\ M(f)\ pr(f|m)\ , \qquad [10]$$

where M(f) is the measure on the space of PADs. By definition, the single PAD we most prefer is the most probable, and we identify this with the PAD assigned by MaxEnt. Hence $pr(f|m)$ must be of the form

$$pr(f|m) = \text{monotonicfunction}(S(f,m))\ , \qquad [11]$$

but we do not yet know which function. Now S has the units (dimensions) of the total f, so this monotonic function must incorporate a dimensional constant, α say, not an absolute constant, so that

$$pr(f\in V|m) = \int_V d^r f\ M(f)\ \Phi(\alpha S(f,m))\ /\ Z_S(\alpha,m)\ , \qquad [12]$$

where Φ is a monotonic function of dimensionless argument and

$$Z_S(\alpha,m) = \int_\infty d^r f\ M(f)\ \Phi(\alpha S(f,m)) \qquad [13]$$

is the partition function which ensures that $pr(f|m)$ is properly normalised.

In order to find Φ, we repeat our earlier mode of reasoning, finding a simple case for which the result is known, and arguing that any general theory must apply to this specific example. Let the traditional team of monkeys throw balls (each of quantum size q) at r cells (i=1,2,...,r), at random with Poisson expectations μ_i. This arrangement satisfies the entropic axioms (I-IV), and the probability of occupation numbers n_i is known (from symmetry and straightforward counting of possible outcomes) to be

$$pr(n|\mu) = \prod_i \mu_i^{n_i} e^{-\mu_i} / n_i! \ . \qquad [14]$$

Define $f_i=n_iq$ and $m_i=\mu_iq$ to remain finite as the quantum size q is allowed to approach zero. Then the image-space of f becomes constructed from microcells of volume q^r, each associated with one lattice-point of integers $(n_1,n_2,...,n_r)$. Hence we have, as q tends to 0,

$$pr(f\in V|m) = \sum_{\text{lattice points in V}} pr(n|\mu)$$

$$= \int_V (d^rf\ /q^r) \prod_i \mu_i^{n_i} e^{-\mu_i} / n_i! \ . \qquad [15]$$

Because we are taking n large, we may use Stirling's formula

$$n_i! = (2\pi n_i)^{1/2} n_i^{n_i} e^{-n_i} \qquad [16]$$

to obtain (accurately to within O(1/n))

$$pr(f\in V|m) = \int_V \frac{d^rf}{\prod (2\pi q f_i)^{1/2}} \exp \frac{\Sigma(f_i-m_i-f_i \log(f_i/m_i))}{q} \qquad [17]$$

Here we recognise the entropy on r cells,

$$\Sigma(f_i-m_i-f_i \log(f_i/m_i)) = S(f,m) \ , \qquad [18]$$

so that

$$pr(f\in V|m) = \int_V \frac{d^rf}{\prod f_i^{1/2}} \frac{\exp(S(f,m)/q)}{(2\pi q)^{r/2}} \ . \qquad [19]$$

Comparing this with the previous formula [12], we must identify

$$q = 1/\alpha \ , \quad \Phi(\alpha S(f,m)) = \exp(\alpha S(f,m)) \qquad [20]$$

and

$$Z_S(\alpha,m) = (2\pi/\alpha)^{r/2} \ , \quad M(f) = \prod f_i^{-1/2} \ . \qquad [21]$$

save possibly for multiplicative constants in Φ, Z_S, M which can be defined to be unity. Note how the often-ignored "square-root" factors in Stirling's

formula have enabled us to derive the measure M, which allows us to make the passage between pointwise probability comparisons and full probability integrals over domains.

A natural interpretation of the measure is as the invariant volume $(\det g)^{1/2}$ of a metric g defined on the space. Thus the natural metric for the space of PADs is

$$g_{ij} = \begin{cases} 1/f_i & \text{if } i=j \\ 0 & \text{otherwise} , \end{cases} \qquad [22]$$

which happens to equal (minus) the entropy curvature $\nabla\nabla S \equiv \partial^2 S / \partial f \partial f$. This quantity, also known as the Fisher information matrix, has also been given this geometrical interpretation by Levine (1986) and by Rodriguez (1989) in these Proceedings.

Although this analysis has used large numbers of small quanta q, so that α is large, this limit also ensures that each n_i will almost certainly be close to its expectation μ_i. Indeed, the expected values of αS remain O(1), so that the identification

$$\Phi(u) = \exp(u) \qquad [23]$$

holds for finite arguments u. Finally, if there is a general form of Φ, it must be valid for the small quantum case, so Φ must be exponential.

To summarise, if there is a general prior for positive, additive distributions f, it must be

$$\text{pr}(f|m) = \exp(\alpha S(f,m)) \, / Z_S(\alpha) \qquad [24]$$

and furthermore

$$\text{pr}(f \in V|m) = \int_V \frac{d^r f}{\Pi \, f_i^{1/2}} \, \frac{\exp(\alpha S(f,m))}{Z_S(\alpha)} , \qquad [25]$$

where

$$Z_S(\alpha) = \int_\infty \frac{d^r f}{\Pi \, f_i^{1/2}} \, \exp(\alpha S(f,m)) . \qquad [26]$$

This quantified prior contains just one un-determined parameter α which can not be fixed a priori because it is dimensional. (Logically, there may be a lurking, contradictory thought experiment, but we have not found one, and accordingly we commend this mode of quantification.)

5. Conclusions

The Classic MaxEnt prior ([24] and [9]) for positive, additive distributions is the only one which gives the correct results in simple cases, so if there is a general prior at all, it can only be this one. It is fully quantified except for the single dimensional number α which can not be assigned a priori. As a bonus, the formal derivation has given us the metric [22] which we need in order to define integrals over ranges of distributions.

ACKNOWLEDGEMENTS

This work was partly supported by Maximum Entropy Data Consultants Ltd.

REFERENCES

Cox, R.P. (1946). Probability, Frequency and Reasonable Expectation. Am. Jour. Phys. 17, 1-13.

Frieden, B.R. (1972). Restoring with maximum likelihood and maximum entropy. J. Opt. Soc. Am., 62, 511-518.

Gull, S.F. (1989). Developments in maximum entropy data analysis. In these Proceedings.

Gull, S.F. & Daniell, G.J. (1979). The maximum entropy method. In Image Formation from Coherence Functions in Astronomy, ed. C. van Schooneveld, pp. 219-225, Reidel.

Gull, S.F. & Skilling, J. (1984). The maximum entropy method. In Indirect Imaging, ed. J.A. Roberts. Cambridge University Press.

Jaynes, E.T. (1978). Where do we stand on maximum entropy? Reprinted in E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics, ed. R. Rosenkrantz, 1983 Dordrecht: Reidel.

Jaynes, E.T. (1986). Monkeys, kangaroos and N. In Maximum Entropy and Bayesian Methods in Applied Statistics, ed. J.H. Justice, pp. 26-58, Cambridge Univ. Press.

Klir, G.J. (1987). Where do we stand on measures of uncertainty, ambiguity, fuzziness and the like? Fuzzy sets and systems, 24, 141-160.

Levine, R.D. (1986). Geometry in classical statistical thermodynamics, J. Chem. Phys., 84, 910-916.

Rodriguez, C. (1989). The metrics induced by the Kullback number. In these Proceedings.

Shore, J.E. & Johnson, R.W. (1980). Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy. IEEE Trans.Info.Theory, IT-26, 26-39 and IT-29, 942-943.

Skilling, J. (1988). The axioms of maximum entropy. In Maximum Entropy and Bayesian Methods in Science and Engineering, Vol. 1., ed. G.J. Erickson & C.R. Smith, pp. 173-188. Kluwer.

Smith, C.R. & Erickson, G.J. (1989). From rationality and consistency to Bayesian probability. In these Proceedings.

Tikochinsky, Y., Tishby, N.Z. & Levine, R.D. (1984). Consistent inference of probabilities for reproducible experiments. Phys. Rev. Lett., 52, 1357-1360.

Developments in Maximum Entropy Data Analysis

Stephen F. Gull
Cavendish Laboratory
Madingley Road
Cambridge CB3 0HE, U.K.

Abstract

The Bayesian derivation of "Classic" MaxEnt image processing (Skilling 1989a) shows that $\exp(\alpha S(f,m))$, where $S(f,m)$ is the entropy of image f relative to model m, is the only consistent prior probability distribution for positive, additive images. In this paper the derivation of "Classic" MaxEnt is completed, showing that it leads to a natural choice for the regularising parameter α, that supersedes the traditional practice of setting $\chi^2=N$. The new condition is that the dimensionless measure of structure $-2\alpha S$ should be equal to the number of good singular values contained in the data. The performance of this new condition is discussed with reference to image deconvolution, but leads to a reconstruction that is visually disappointing. A deeper hypothesis space is proposed that overcomes these difficulties, by allowing for spatial correlations across the image.

1. Introduction

The Maximum Entropy method (MaxEnt) has now become the standard method for data analysis in many fields. It has been used most spectacularly in radio-astronomical interferometry, where it deals routinely with images of up to a million pixels, with high dynamic range. A review of the method, together with many examples taken from fields such as optical deblurring and NMR spectroscopy is given by Gull & Skilling (1984). Despite the success of the method in practical applications, the underlying rationale of MaxEnt has caused widespread controversy. This paper, together with the one preceding it (Skilling 1989a), presents a Bayesian justification for the use of MaxEnt.

The desire for a fully Bayesian interpretation of MaxEnt is not new: the advantage of such a probabilistic formulation being that it would then allow us to quantify the reliability of MaxEnt images. The "team of monkeys" argument as applied to image reconstruction (e.g. Gull & Daniell 1979, following Frieden 1972, see also Jaynes 1986b) was an attempt to derive a prior probability distribution on the space of images. But these earlier attempts had a fundamental drawback: why should we consider that all images are made randomly by monkeys? Clearly they are not. However, the arguments presented here by Skilling are of a completely different character. By asking merely that any supposed general procedure should also work in every specific, simple case, he shows that if there is a consistent prior on the space of images it must be of the form $\exp(\alpha S(f,m))$, where $S(f,m)$ is the entropy of image f relative to model m. This prior is, of course, consistent with that derived by the "monkey" argument, because it is conceivable (though unlikely) that some images could actually be made that way.

The purpose of the present paper is to complete the derivation of "Classic"

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 53–71.

MaxEnt by addressing the choice of the regularising parameter α that appears in the prior. With noisy data, traditional practice has been to select a value of α that makes the χ^2 misfit statistic equal to the number of observations, but this is ad hoc and does not allow for the reduction in effective number of degrees of freedom caused by fitting accurate data. Section 2 gives a Bayesian determination of α, finding that the amount of structure in the image, quantified as $-2\alpha S$, must equal the number of "good" (accurate) singular vectors contained in the data. The value of χ^2 is not relevant to the choice of α, but instead allows an estimate of the overall noise level if it is unknown.

The application of this method is discussed (Section 3) by reference to a specific deconvolution example. Disconcertingly, the "Classic" reconstruction is visually disappointing, with an unfortunate level of "ringing". This can only be due to a poor choice of initial model m. Indeed, the initial, flat model is very far from the final reconstruction. In order to allow the "good" singular data vectors to be fitted, α must be small, so that there is little entropic smoothing, and the consequence is under-smoothing of the "bad" noisy data.

The next step must be a better model, incorporating some expectation of correlated spatial statistics in a deeper hypothesis space (Section 4). We introduce a set of "hidden variables" $\tilde{m}(x)$ which are then blurred to make the model m(x) used in "Classic". The prior for these hidden variables must also be of entropic form $\exp(\beta S(\tilde{m},\text{flat}))$. The new multiplier β and the width of the hidden blur are also determined by Bayesian methods.

The results from this deeper hypothesis space are excellent, and provide a coherent rationale for some of the manipulations of the model m that have been found useful in current practice.

2. The choice of α in Classic MaxEnt

In the preceding paper (Skilling 1989a) it was shown that the only consistent prior for positive, additive images is of the form:

$$\text{pr}(f|m,\alpha) = \exp \alpha S(f,m) \;/\; Z_S(\alpha,m) \;, \qquad [2.1]$$

where S in the entropy of image f relative to model m and Z_S is the normalising partition fuction. Explicit forms for S and Z were derived for the case of an image discretised to r pixels:

$$S(f,m) = \sum_j (f_j - m_j - f_j \log(f_j/m_j)) \qquad [2.2]$$

$$Z_S(\alpha,m) = \int d^r f \; \Pi \, f^{-1/2} \; \exp \alpha S \;. \qquad [2.3]$$

The only remaining parameter in this "Classic" hypothesis space is the constant α. We do not believe that we can determine α a priori by general arguments. Not only is α dimensional, so that it depends on the scaling of the problem, but its best-fitting value varies quite strongly with the type and quality of the data available. It can only be determined a posteriori.

We therefore turn to the other side of the problem, the likelihood, which we write as:

$$\text{pr}(D|f) = \exp(-L(f)) \;/\; Z_L, \qquad [2.4]$$

where $Z_L = \int d^N D \exp(-L)$, [2.5]

N being the number of data. The log-likelihood L(f) defined by this expression contains all the details of the experimental setup and accuracies of measurement. For the common case of independent, Gaussian errors, this reduces to $L = \chi^2/2$, but other types of error such as Poisson noise are also important. Quite frequently, the overall level of noise is not well-known, so we will eventually generalise to

$$pr(D|f,\sigma) = \exp(-L(f)/\sigma^2) \; / \; Z_L(\sigma) \; , \qquad [2.6]$$

but for now we assume that the errors are known in advance, so that $\sigma = 1$.

We now write down the joint p.d.f. of data and image:

$$pr(f,D|\,\alpha,m) = Z_L^{-1} \, Z_S^{-1} \, \exp(\alpha S - L) \; . \qquad [2.7]$$

Bayes' Theorem tells us that this is also proportional to the posterior probability distribution for f: $pr(f|D,\alpha,m)$. The maximum of this distribution as a function of f is then our "best" reconstruction, and occurs at the maximum of

$$Q = \alpha S - L. \qquad [2.8]$$

This brings us back once again to the choice of α, which can now be viewed as a regularising parameter. When seen this way, α controls the competition between S and L: if α is large, the data cannot move the reconstruction far from the model - the entropy term dominates. If α is low there is little smoothing and the reconstruction will show wild oscillations as the noise in the data is interpreted as true signal. We have to control α carefully, but there is usually a large range of sensible values.

Our practice hitherto (Gull & Daniell 1978, Gull & Skilling 1984) has been to set α so that the misfit statistic χ^2 is equal to the number of data points N. Although this has a respectable pedigree in the statistical literature (the discrepancy method (Tikhonov & Arsenin 1977)), it is ad hoc, and can be criticised on several grounds.

1) The only "derivation" of the χ^2=N condition that has been produced is a frequentist argument. If the image was known in advance and the data were then repeatedly measured, χ^2=N would result on average. However, the data are only measured once and the image is not known a priori, but is instead estimated from the one dataset we have.

2) There is no allowance for the fact that good data cause structure in the reconstruction f. These "good" degrees of freedom are, in effect, parameters that are being fitted from the data and because of this they no longer contribute to the variance. In general terms, χ^2=N leads to "under-fitting" of data (Titterington 1985). This is particularly apparent for imaging problems where there is little or no blurring. The χ^2=N criterion leads to a uniform, one standard deviation bias towards the model. This bias is very unfortunate: it is the job of a regulariser such as entropy to cope with noise and missing information, not to bias the data that we do have.

3) For many problems (such as radioastronomical imaging, where I started) the data are nearly all noise, so that $\chi^2 \approx$ N for any reasonable α. The

statistic χ^2 is in any case expected to vary by $\pm\sqrt{N}$ from one data realisation to another, and this can easily swamp the difference between χ^2 at $\alpha=\infty$ and the χ^2 appropriate to a sensible reconstruction.

For these reasons we now believe that there is no acceptable criterion for selecting α that looks only at the value of a misfit statistic such as χ^2. However, within our Bayesian framework there is a natural way of choosing α. We simply treat it as another parameter in our hypothesis space, with its own prior distribution. The joint p.d.f. is now

$$\text{pr}(f,D,\alpha|m) = \text{pr}(\alpha)\ \text{pr}(f,D|\alpha,m) \quad . \qquad [2.9]$$

To complete the assignment of the joint p.d.f. we select an uninformative prior, uniform in $\log(\alpha)$: $\text{pr}(\log\alpha)$ = constant over some "sensible" range $[\alpha_{min},\alpha_{max}]$. We shall return to the definition of "sensible" later.

Using Bayes' Theorem, this joint distribution is also proportional to the posterior distribution $\text{pr}(f,\alpha|D,m)$ and we proceed to estimate the best value of α by marginalisation over the reconstruction f:

$$\text{pr}(\alpha|D,m) = \int d^r f\ \Pi\, f^{-1/2}\ \text{pr}(f,\alpha|D,m) \ .$$

$$\propto Z_Q\ Z_S^{-1}\ Z_L^{-1} \ , \qquad [2.10]$$

where $$Z_Q = \int d^r f\ \Pi\, f^{-1/2}\ \exp(\alpha S - L) \ . \qquad [2.11]$$

It is essential to perform this integral carefully, rather than estimating α by maximising the integrand with respect to f and α simultaneously, because the distribution in f-α space is significantly skew. In fact, the maximum of $\text{pr}(f,\alpha|D,m)$ is usually at $\alpha=\alpha_{max}$=large; $f \approx m$, which is certainly not what we want.

We now evaluate the integrals involved. The integrand for Z_S has a maximum at f = m and, using Gaussian approximations, we find that for all α a reasonable approximation to log Z_S is:

$$\log Z_S = r/2\ \log(\alpha/2\pi). \qquad [2.12]$$

In performing this integral, the terms from the volume element cancel with those from the curvature $\nabla\nabla S$. This is a happy consequence of the fact that the entropy curvature is also the natural metric tensor of the f space.

The Z_Q integral is done similarly, expanding about the maximum of $Q(f,m,\alpha)$ at $\hat{f}$. We can aid our understanding by introducing at this point the eigenvalues $\{\lambda_i\}$ of the symmetric matrix

$$A = \text{diag}(f^{1/2}) \ . \ \nabla\nabla L \ . \ \text{diag}(f^{1/2}) \ , \qquad [2.13]$$

which is the curvature of L viewed in a the entropy metric. The eigenvalues λ and eigenvectors in f space define the natural coordinates for our problem, and the $\lambda^{1/2}$ are the appropriate "singular values". A large value of λ implies a "good" or measured direction, whereas a low or zero λ corresponds to a poorly measured quantity.

Evaluating the integrals in the Gaussian approximation, we find

$$\log \mathrm{pr}(\alpha|D,m) = \text{constant} + r/2 \log(\alpha) - 1/2 \log\det(\alpha I + B) + Q(\hat{f},m,\alpha)$$

$$= \text{constant} + 1/2 \sum_j \log(\alpha/(\alpha+\lambda_j)) + \alpha S(\hat{f},m) - L(\hat{f}). \quad [2.14]$$

For large datasets this has a sharp maximum at a particular value of α. Differentiating with respect to $\log\alpha$, and noting that the $\hat{f}$ derivatives cancel, we find the condition:

$$-2\,\alpha\, S(\hat{f},m) = \sum_j \lambda_j/(\alpha+\lambda_j) \; . \quad [2.15]$$

This fixes our estimate of $\alpha = \hat{\alpha}$ quite closely, provided we have many data, so that we can return to the determination of the reconstruction $\hat{f}$. Strictly, having already integrated out f to determine $\mathrm{pr}(\alpha)$, the formalism does not allow us to return with a single value $\hat{\alpha}$. However, we are allowed to find the distribution of any integral $R \equiv \int d\alpha\, f(x)\, r(x)$ by integrating the joint p.d.f. successively over f and then α. Because $\mathrm{pr}(\alpha)$ is so sharply peaked, the effect on R is just as if α were set equal to $\hat{\alpha}$. We may as well simplify the notation by setting $\alpha = \hat{\alpha}$ in the derivation of f itself:

$$\mathrm{pr}(f|D,m) = \int d\alpha\, \mathrm{pr}(f,\alpha|D,m)$$

$$= \int d\alpha\, \mathrm{pr}(\alpha|D,m)\, \mathrm{pr}(f|\alpha,D,m)$$

$$\simeq \mathrm{pr}(f|\hat{\alpha},D,m)$$

$$= Z_Q^{-1} \exp(\hat{\alpha} S(\hat{f},m) - L(\hat{f})) \; . \quad [2.16]$$

The fluctuations (uncertainty) of f about $\hat{f}$ can also be investigated, at least in principle, by using the known curvature:

$$\langle \delta f \delta f^t \rangle = [\nabla\nabla Q]^{-1} \; . \quad [2.17]$$

We can understand our Bayesian formula for the best value $\hat{\alpha}$ as follows.

1) The statistic $\lambda/(\alpha+\lambda)$ is a measure of the quality of the data along any given singular vector. If $\lambda >> \alpha$ the data are good and $\lambda/(\alpha+\lambda)$ adds one to the statistic. If, on the other hand, $\lambda << \alpha$, then the regularising entropy dominates the observations and the contribution is approximately zero. We can therefore say that $\Sigma\lambda/(\alpha+\lambda)$ specifies the number of good, independent data measurements, or the number of degrees of freedom ndf(S) associated with the entropy. We associate the degrees of freedom with the entropy rather than the likelihood because these are the directions (dimensions) that contribute to the entropy.

2) The quantity $-2\alpha S$ is a dimensionless measure of the amount of structure in the image relative to the model, or the distance that the likelihood has been able to pull the reconstruction away from the starting model.

The formula thus has a very plausible interpretation: the dimensionless measure of the amount of structure demanded by the data is equal to the number of good, independent measurements. We also note that, as we indicated earlier, the value of the misfit statistic L is irrelevant to the choice of α. However, it too has a rôle to play. To see this we now generalise to the case of unknown overall noise level

$$pr(D|f,\sigma) = \exp -L(f)/\sigma^2 \; / \; Z_L(\sigma), \qquad [2.18]$$

and this time keeping all terms involving σ find:

$$\log pr(\alpha,\sigma|D) = \text{constant} - N \log(\sigma) + 1/2 \; \Sigma \; \alpha'/(\alpha'+\lambda_j) + \alpha S - L/\sigma^2, \qquad [2.19]$$

where $\alpha' = \alpha\sigma^2$. There is now an additional Bayesian choice for σ and its estimate $\hat{\sigma}$,

$$2 \; L(\hat{f})/\sigma^2 = N - \Sigma \; \lambda/(\alpha+\lambda) \equiv ndf(L) \qquad [2.20]$$

The interpretation of this condition is also very plausible: the expected χ^2 (=2L) is equal to the number of degrees of freedom controlled by the entropy, that is, the poorly measured "bad" directions of f space. This is less than the number of data, thereby answering our first objection to χ^2=N, and showing that the χ^2 (or L) is really suited to estimation of the noise level, not α. Notice also how there is a clean division of degrees of freedom between S and L, so that

$$N = ndf(S) + ndf(L) \; . \qquad [2.21]$$

The choice of regularising parameters has been much debated in the statistical literature (Titterington 1985 gives a review). Our arguments in this section have reproduced (albeit for an entropic variation) one of these prescriptions, known elsewhere as Generalised Maximum Likelihood (Davies & Anderssen 1986).

3. Performance of the Bayesian α

To illustrate both the power and the shortcomings of the Bayesian choice for α, we turn now to a practical example, a picture of "Susie". Figure 1 shows Susie, digitised on a 128x128 pixel grid, with grey-level values between 40 and 255. This picture was blurred with a 6-pixel radius Gaussian point-spread function (PSF) and noise of unit variance added. This is a traditional example for MaxEnt processing (e.g. Daniell & Gull 1980, Gull & Skilling 1984), and we show a χ^2=N reconstruction. Our previously-published "Susies" have used a disc PSF, appropriate to an out-of-focus camera, and for which the MaxEnt results at this signal-to-noise are more impressive visually. A Gaussian PSF gives less improvement in resolution because the eigenvalues of $\nabla\nabla L$ fall off very fast.

We now reach the first practical difficulty associated with our Bayesian answer. The log-determinant and the ndf(S) statistic require a knowledge of the eigenvalue spectrum of $f^{1/2}\nabla\nabla L f^{1/2}$. For the present case, this is a 16384x16384 matrix, a size which is well in excess of the limits for conventional computational methods of calculating eigenvalues. However, Skilling (1989b) has recently developed a method based on the application of the matrix to random vectors, together with the use of MaxEnt, that allows an estimate of the eigenvalue spectrum to be obtained. In particular, the accuracy of estimation of scalars such as ndf(S) is excellent using this technique. It seems, therefore, that practical computation of the Bayesian solution is in general possible.

For the moment, the problem of the eigenvalues is avoided in a different way: we change the definition of S. All of the Bayesian analysis of the last

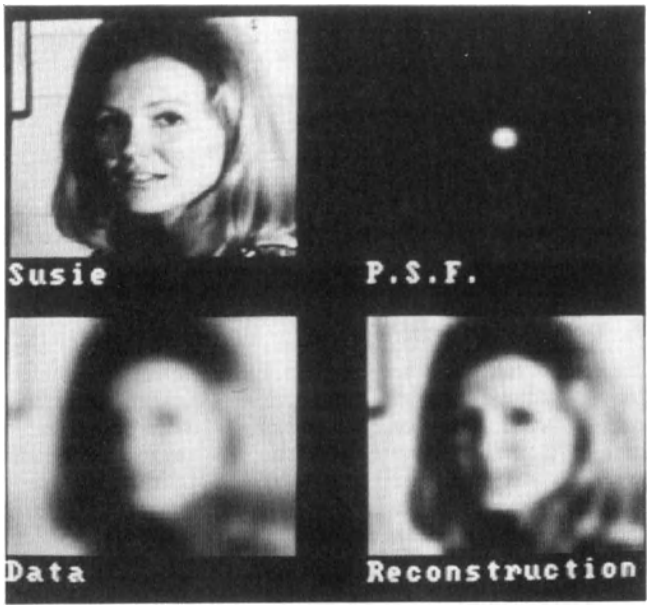


Figure 1. 128x128 image of Susie, blurred with a 6-pixel Gaussian PSF. MaxEnt reconstruction using $\chi^2 = N$.

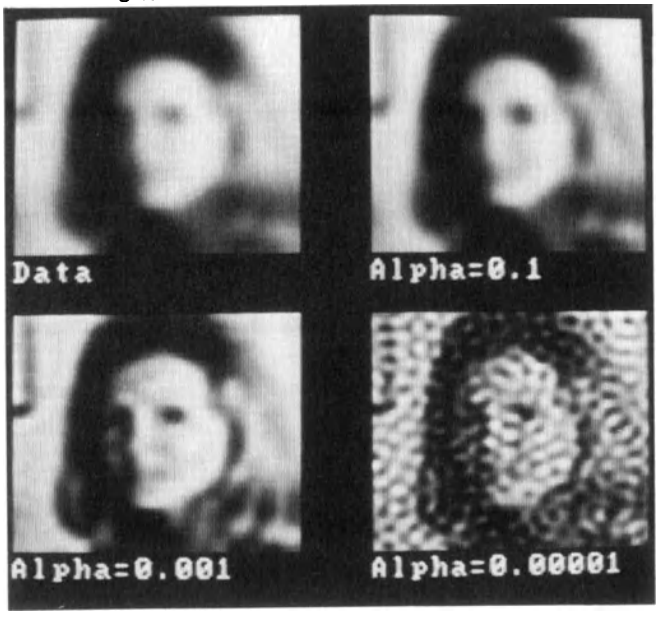


Figure 2. Susie images showing the behaviour of reconstruction quality as α is varied.

section applies equally to any regularising function, so we select a simple one that allows us to to diagonalise $\nabla\nabla L$ and $\nabla\nabla S$ simultaneously. This is the case for a spatially-invariant, circulant PSF and for the quadratic

$$S = -1/2 \sum_j (f_j - m_j)^2 \quad , \qquad [3.1]$$

which is a linearised version of the correct form, and a reasonable approximation for a low-contrast image such as Susie. The computations can now be performed easily in eigenvector (Fourier Transform) coordinates. The change in the definition of S makes no difference to the formulae, except that the metric is now flat, the $f^{1/2}$ terms disappear and f might possibly go negative. The change makes no difference whatever to our conclusions about the performance of the Bayesian solution.

Figure 2 shows the reconstruction from blurred Susie for a selection of α values. When α is high the reconstruction looks like the original blurred data, and when α is too low unsightly ripples appear due to the amplification of noise. Note, however that this behaviour covers a wide range of α ($\sim 10^4$) and that there is a large region where the reconstruction is generally satisfactory.

For our example the Bayesian solution suggests that there are ~790 good degrees of freedom out of the total 16384. As might be expected, this is somewhat greater than the $16384/36\pi$=145 independent PSFs contained in the image, the excess being a rough measure of the degree of deconvolution obtained. Its estimate of the noise level was correct to within the expected error and, indeed, we have always found that the noise level prediction performance of the Bayesian solution is excellent. Figure 3 shows a plot of the posterior probability of α, both as its logarithm, and also linearly, to emphasise the discrimination in the determination of $\hat{\alpha}$, which is better than 1 db for this dataset. The posterior p.d.f. is normalisable as α approaches zero (towards the left of Figure 3a,b,c) if the noise level is known, but a global view shows that it levels off once α exceeds the highest eigenvalue (towards the right of Figure 3c), resulting in a technically improper distribution. We therefore return to the definition of a "sensible" cutoff for α_{max} referred to earlier. The scale of Figure 3c is rather large: in order to make a 50 per cent contribution to the probability integral, the α_{max} cutoff has to exceed $\exp(\exp(1.4\times10^7))$. Such numbers are typical of the "singularities" encountered in this type of Bayesian analysis and we are content to take α_{max} less than this bizarre value.

The reconstruction $\hat{f}(\hat{\alpha})$ is shown as Figure 4. It it visually disappointing, and is clearly in the range of the "over-fitted" solutions for which α is too low. It is very easy to understand why this is so. The initial model used for these reconstructions was everywhere uniform, at approximately the mean of the data. This model is very far from the final reconstruction, because there is plenty of real structure in the picture produced by the 790 good measurements in the data. α must be reduced sufficiently to accommodate this structure, or a large penalty in L results. An unfortunate consequence is that α now becomes too low to reject noise properly along the "bad" directions. In general terms, the Bayesian solution will tend to allow fluctuations of the same order of magnitude as the deviation of the reconstruction from the initial model.

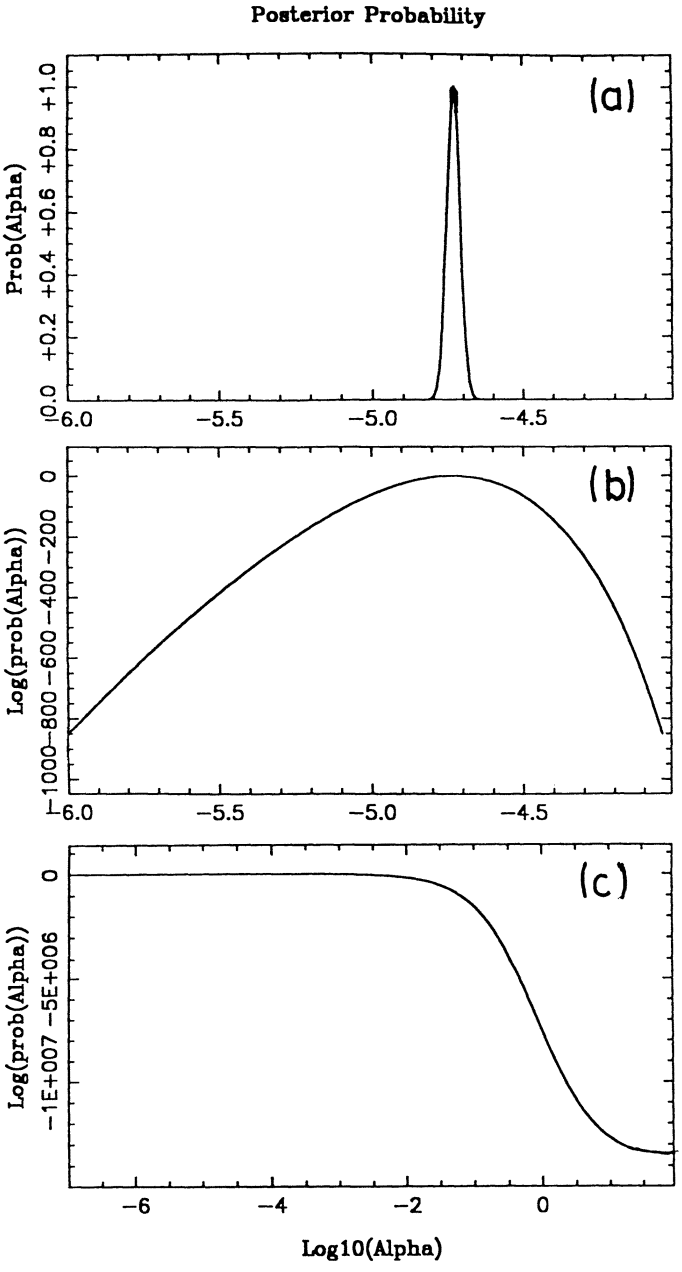


Figure 3. Posterior distribution of the smoothing parameter α for the Susie image, plotted (a) logarithmically, (b) linearly, (c) logarithmically over a large range of α.

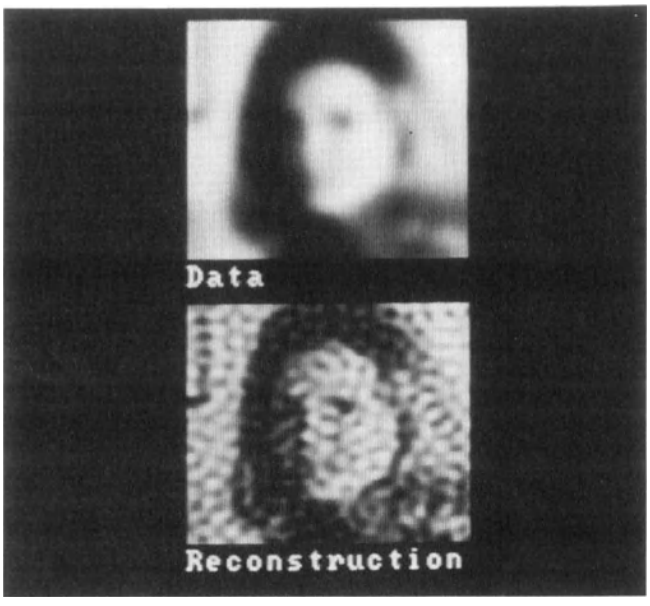


Figure 4. "Classic MaxEnt" reconstruction of Susie.

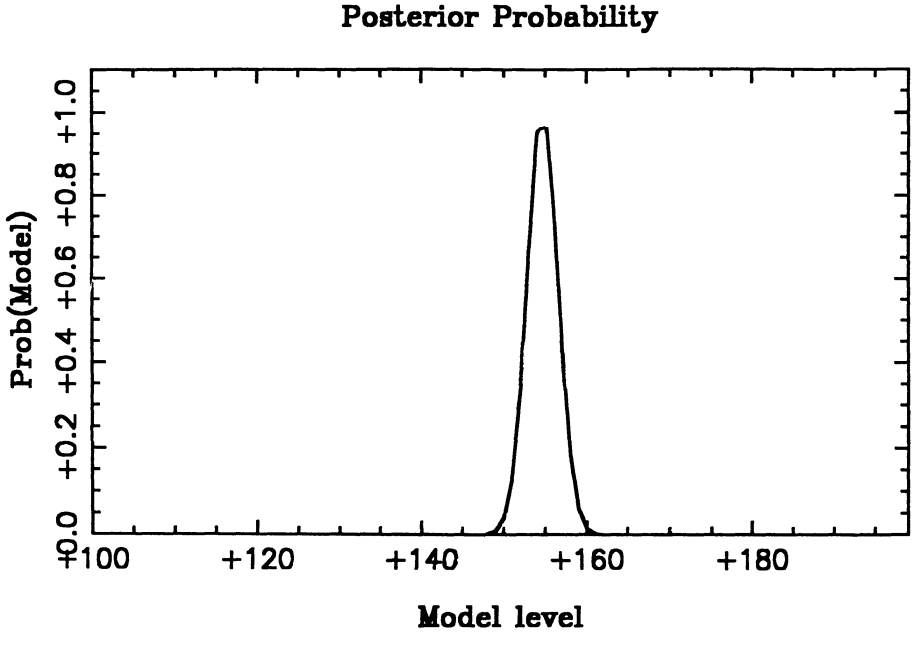


Figure 5. Posterior probability distribution of the initial model level m_0 for the Susie image. The maximum occurs at the mean of the data.

4. New MaxEnt

We have seen that the Bayesian choice of α will often lead to a reconstruction that is over-fitted. Despite this, we feel that this "Classic" choice is the correct answer to the problem that we have so far formulated. In fact, it was the purity of its derivation, combined with problems of its performance that led us to propose the name "Classic" for it. We have derived a joint p.d.f $pr(D,f,\alpha,\sigma|m)$ which is still conditional on the knowledge of an initial model m. This m was first introduced as a "measure" on the x-space of pixels, but it is a point in f-space and acts as a "model" there. The only freedom that we have left in our hypothesis space is to consider variations in this model, which we recall was a flat, uniform picture set to the average of the data m_0. The fact that the model was flat expresses our lack of prior information about the structure of the picture, but where did the brightness level m_0 come from?

The answer is again: Bayes' Theorem. We expand the hypothesis space to $pr(D,f,\alpha,\sigma,m_0|flat)$ and select an uninformative prior for $pr(m_0|flat)$. The posterior distribution for m_0 (Figure 5) is again sharply peaked and in the Gaussian approximation has a maximum at exactly the mean of the data. Reconstructions using values of m_0 different from this Bayesian optimum exacerbate the over-fitting problem, as one would expect. However, this exercise of varying the model is very instructive, because it emphasises the cause of the problem; the picture is very non-uniform. There are large areas of the picture where the lighting is generally light or dark, with interesting details superimposed. There are correlations from pixel to pixel present in the image that we have so far ignored. Indeed, our earlier MaxEnt Axiom I forbids us to put pixel-pixel correlations directly into our prior $pr(f|m,\alpha)$. We wish to circumvent this axiom, but we must be subtle.

Suppose we imagine a silly case where the left half of our picture is Susie, but the right half is a distant galaxy. Axiom I is designed to protect us from letting the reconstruction of Susie influence our astrophysics, or vice-versa. But there is nothing stopping us from having a different m_0 level for each half. In fact, in view of the grossly different luminance levels involved, it would be extremely desirable to have different levels of m_R and m_L. When seen this way, there is nothing to prevent us considering the right and left halves of the original Susie picture separately, because the average luminance levels are different. A new hypothesis space involving $pr(m_R,m_L|flat,L/R)$ will again fix suitable levels for m_R and m_L a posteriori. If there is a strong right/left brightness variation across the picture, then this two-value model will be closer to the reconstruction and $\hat{\alpha}$ will increase, reducing the ripples. But in that case why not use 4 subdivisions (top/bottom, left/right), or 8, or more?

If we continue to subdivide, we can get a better model, closer to the reconstruction, so we expect that $\hat{\alpha}$ will increase. However, we are introducing extra parameters, so that we would expect there to be a penalty for this, and that it would be likely to have some effect on the choice of $\hat{\alpha}$. A further consideration is that, if at all possible, we should like to avoid the sharp boundaries that such a crude division of the model would involve.

We are now in a position to formulate a new, flexible hypothesis space that is suitable for pictures such as Susie. We suppose that the model m for use in "Classic" MaxEnt is itself generated from a blurred image of hidden variables $\tilde{m}$:

$$m = \tilde{m} * b = B\,\tilde{m}\,, \qquad [4.1]$$

where b is our "model-blur" PSF, which can also be written as a circulant matrix B. For the case of Susie we might like to think of $\tilde{m}$ as the source of background lighting. If this model-blur is broad, then our model in "Classic" is smooth, and there are effectively very few parameters in it. If b is narrow, there are many parameters. The shape and width of the model-blur are to be determined by Bayesian methods as well. We do not expect the shape of this blur to matter greatly and we arbitrarily restrict it to be a Gaussian. The crucial parameter is the width and we expect that the most useful width will be about equal to the size of the correlation-length that is actually present in the picture. Our Bayesian analysis of the larger, richer hypothesis space will then tell us how useful is the freedom provided by the hidden variables. The final probability levels will quantify for us the level of improvement relative to "Classic", which is contained in our new space as limiting cases.

To complete the analysis we must assign a prior for the "pre-model" $\tilde{m}$. We treat $\tilde{m}$ as an image and again use the entropic prior:

$$\text{pr}(\tilde{m}|\beta,\text{flat}) = Z_T^{-1} \exp(\beta T), \qquad [4.2]$$

where $T = S(\tilde{m},\text{flat})$ and we have introduced β as a new Lagrange multiplier for the $\tilde{m}$-space entropy T. We again restrict ourselves to the mathematically tractable (but still interesting) case of quadratic S and T, circulant blurs and spatially uniform noise level, for which the $\nabla\nabla L$, $\nabla\nabla S$ and $\nabla\nabla T$ matrices are all simultaneously diagonal in Fourier transform space. The Bayesian calculation of $\hat{\alpha}$ and $\hat{\beta}$ now yields:

$$-2\,\alpha\, S(\hat{f},\hat{m}) = \text{ndf}(S) = \sum_i \beta\lambda_i/(\alpha\beta+\beta\lambda_i+\alpha\lambda_i b_i^2)\ , \qquad [4.3]$$

$$-2\,\beta\, T(\hat{\tilde{m}},\text{flat}) = \text{ndf}(T) = \sum_i \alpha\lambda_i b_i^2/(\alpha\beta+\beta\lambda_i+\alpha\lambda_i b_i^2) \qquad [4.4]$$

where b_i^2 are the eigenvalues of B^tB and

$$\log \text{pr}(\alpha,\beta,b|D) = \text{constant} + 1/2 \sum_i \alpha\beta/(\alpha\beta+\beta\lambda_i+\alpha\lambda_i b_i^2) + \beta T + \alpha S - L\ . \qquad [4.5]$$

The noise level σ can also be estimated as before:

$$\chi^2 = 2\,L(\hat{f})/\hat{\sigma}^2 = N - \text{ndf}(S) - \text{ndf}(T)\ . \qquad [4.6]$$

Notice how there is once again a neat division of the degrees of freedom between S, T and L.

We have tested the performance of New MaxEnt on the Susie picture. Classic MaxEnt is contained in New MaxEnt in several ways:

1) as $\beta \to \infty$, because $\tilde{m}$ cannot move from the initial m_0.
2) As $b \to \infty$, because the model becomes flat.
3) (rather surprisingly) As $b \to 0$. This last case illustrates a general peculiarity of

$$\log \text{pr}(\alpha,\beta,b|D) = \text{constant} + 1/2 \log(\det) + \alpha S + \beta T - L\ , \qquad [4.7]$$

an object which would be known elsewhere in physics as a Gibbs' surface. Our new

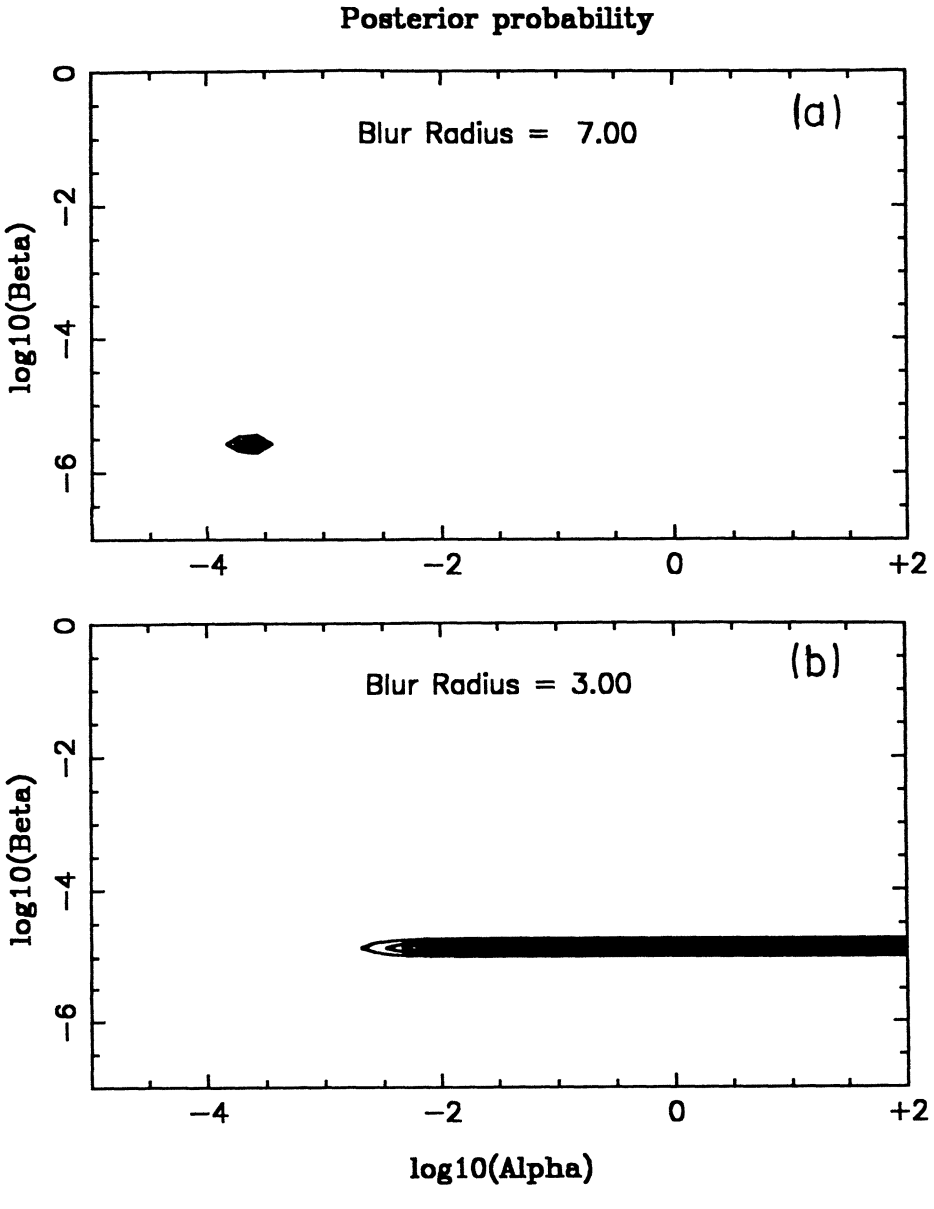


Figure 6. Posterior p.d.f. of the Lagrange multipliers α and β for the New MaxEnt reconstruction of Susie, having b=3 and b=7 pixels. The contour intervals are linear.

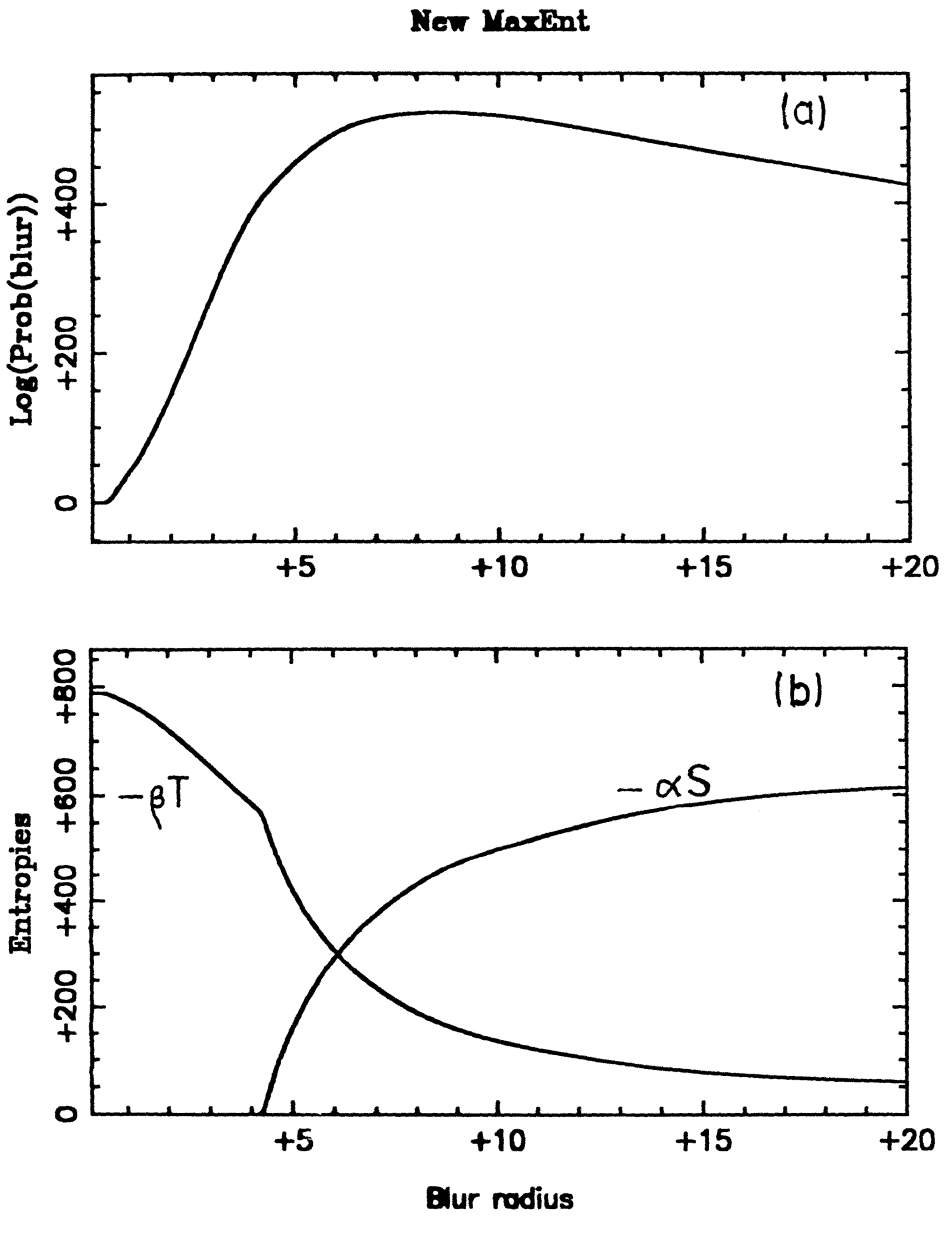


Figure 7. (a) Posterior distribution of the model-blur width for New MaxEnt Susie images. (b) Image-space entropy S and model-space entropy T. Note that S is zero for model-blurs narrower than 4.27 pixels.

hypothesis space has sufficient structure to contain phase transitions and one such occurs for the Susie image as the the width of b is reduced below 4.27 pixels. Below this value of the model-blur, the model is sufficiently detailed to cope with all the structure in the image demanded by the data, and S(f,m) no longer adds anything that is useful. The New MaxEnt $\hat{\alpha}$ increases to infinity at this point; S switches off and the reconstruction is the model $m = \tilde{m} * b$. This is illustrated in Figure 6, which shows the posterior distribution of α and β for b=3 and b=7 pixels.

Figure 7 shows the posterior distribution of the width of b, which rises to a maximum at ~8.5 pixels. This diagram also answers the question of how useful our new hypothesis space is. It is useful to the extent of being more probable than Classic MaxEnt by exp(520). The extrinsic variables S and T are also plotted, showing a change of slope at the phase transition. There is no specific heat associated with this phase change! Inspection of the reconstruction and effective model $m = \tilde{m} * b$ for the optimum width of b (Figure 8) confirms that the New MaxEnt has indeed achieved its promise.

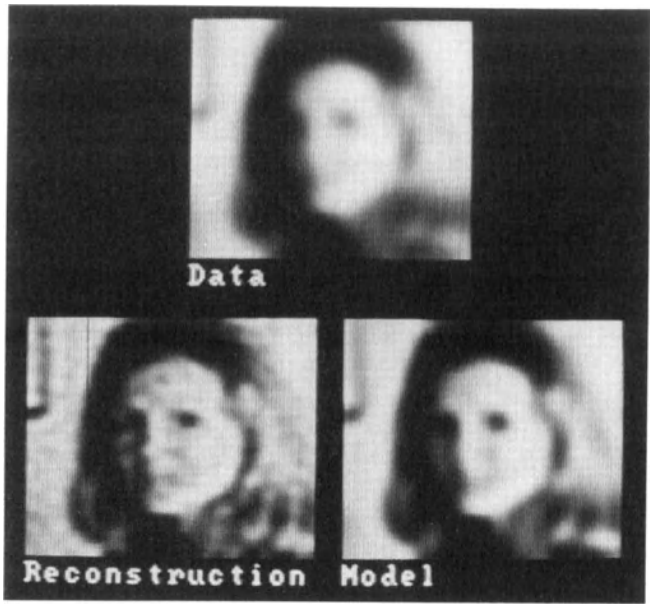


Figure 8. New MaxEnt reconstruction of Susie. Compare with Figure 4.

Of course, our New MaxEnt can be used to encourage smoothness in any image, whether or not it is actually blurred. Indeed, our failure to offer a solution the problem to analysing noisy, but unblurred pictures has been a continual source of frustration over the years. We test the noise-smoothing properties of the method with a picture of Susie which is in focus, but which has had 25 units of noise added. For this type of problem, the Classic MaxEnt reconstruction is almost identical to the data. The best value of the model blur is now ~3 pixels, and there is an increase in probability of exp(10000) over Classic for this case. The picture produced (Figure 9a) is also very good, and shows all the

Figure 9. (a) Comparison of Classic and New MaxEnt reconstructions of a noisy Susie picture. (b) Detail of Figure 9 (a), showing the improvement due to noise suppression.

structure that can be reliably produced from this noisy dataset. A detail from this (Figure 9b) confirms that the pixel-to-pixel noise has been greatly reduced, without degrading the information content of the picture in any way.

5. Discussion

Our New MaxEnt approach is related to other methods of introducing spatial smoothing that have been found useful in practice. Within the context of maximum entropy image processing, there are now many examples of "reconstruction-dependent" models $m(\hat{f})$. A particularly successful application to tomographic mapping of stellar accretion discs is presented by Marsh and Horne (1989), following Horne (1985). To improve the quality of the images, they used a model that was a blurred form of their current reconstruction. We have also found such techniques useful: Charter & Gull (1988) give an example of studies of drug absorption rate into the bloodstream, in which a blurred version of the reconstruction is again used as the model.

Such tricks have previously lacked any rigorous justification, because the development of the MaxEnt story treats m as a point in f-space that is given a priori. It was thus difficult to see how we could legally let it depend on $\hat{f}$. However, in New MaxEnt, the effective model m looks very much like a blurred version of $\hat{f}$, although it is actually a blurred version of the hidden variables $\tilde{m}$. We can now justify the above tricks in terms of New MaxEnt. Thus in the drug absorption problem, f represents the rate of absorption into the bloodstream, $\tilde{m}$ is the rate at which the tablets break down in the stomach, and b represents the time delay as the drug passes through the liver. Charter (private communication) also gives another, intriguing example, in which he simply pretends that the data are more blurred than is actually true, adding an additional "pre-blur" to the real PSF. Often the results are improved by this device, encouraging smoothness and eliminating noise. We can now see that this trick too is covered in New MaxEnt as the degenerate case $\alpha \dashrightarrow \infty$ that occurs in the case of Susie for small model-blurs. The New MaxEnt hypothesis space provides a natural justification for these variants, and automatically includes any consequential effect upon the value of α due to the additional parameters in the model.

It is also useful to examine our new procedure in the context of spatial statistics, where the currently favoured techniques are things such as Markov random fields (Kinderman and Snell 1980, Geman & Geman 1984) and smoothness-enforcing regularisers (Titterington 1985). We can compare New Maxent with these techniques by marginalising out $\tilde{m}$ to get an effective prior for $pr(f|\alpha,\beta,b,flat)$. We have not so far done this, because it would obscure the real structure of our hypothesis space, which is still faithful to the spirit of Axiom I. When we do it, we find

$$pr(f|\alpha,\beta,b,m_0) \propto \exp -1/2\ \delta f^t R^{-1} \delta f\ , \qquad [5.1]$$

where $\delta f_i = f_i - m_0$ and R is a circulant matrix that has eigenvalues $1/\alpha + b^2/\beta$.

By varying the shape of the model-blur b we can clearly mimic any given spectral behaviour of spatial smoothing. Markov random fields correspond to particular functional forms of b. New MaxEnt contains these techniques as special cases. However, we prefer the rationale of our new hypothesis space, because we feel it is more closely related to our prior knowledge of the imaging problem.

6. Conclusions

The Bayesian choice of the regularising parameter α completes the derivation of Classic MaxEnt and represents a major advance over our previous practice of setting $\chi^2=N$. The resulting formula $-2\alpha S = ndf(S)$ is theoretically appealing, and expresses the fact that the amount of structure produced in the reconstruction is equal to the number of good, independent measurements present in the dataset.

For some problems we have found the Classic value of α to be satisfactory, but there are general grounds for supposing that it leads to over-fitting, because α has to be reduced to allow for the structure produced by good data. This leads to under-smoothing of bad data, as we have illustrated with our picture of Susie.

The New MaxEnt hypothesis space which incorporates spatial correlations is sufficiently powerful to correct these problems and is considerably more probable than Classic, showing that the inclusion of spatial information is useful.

New MaxEnt also provides a consistent rationale for a wide class of model manipulations that are found to be useful in practical applications. Athough we have, for reasons of computational expediency, illustrated the New MaxEnt only in the quadratic (Wiener filter) approximation, the results are already excellent. We do not expect our conclusions to change when the correct entropic forms are used, indeed the results can only impove.

Finally, we ask the question: "Is our hypothesis space good enough?" Of course, the answer depends on what we are trying to achieve. Certainly our new procedure is good enough to overcome the over-fitting problems of Classic MaxEnt and produce a good reconstruction of Susie. However, looking at the images produced for different values of the model-blur width, our eyes tell us that the reconstruction for b=5 pixels is visually slightly better than that for the Bayesian optimum b=8.5 pixels, although the probability of b=5 is lower by exp(50). This is a warning that we may eventually find another, deeper hypothesis space even more useful for the imaging problem (as envisaged by Jaynes 1986a). We speculate that the improvement we get by going to b=5 tells us something about human vision. We pay attention to the fine details present in Susie's face and relatively ignore the background. The computer, with its spatially-invariant model PSF sees the smooth surfaces in the background and weights them equally, thereby arriving at a slightly larger correlation length than our eyes would like.

Acknowledgments

I am grateful to all the past and present members of the Cambridge MaxEnt Group for discussions about the MaxEnt stopping criterion during the last twelve years. This work was partly supported by Maximum Entropy Data Consultants Ltd.

References

Charter, M.K. & Gull, S.F. (1988) Maximum entropy and its application to the calculation of drug absorption rates. J. Pharmacokinetics and Biopharmaceutics, 15, 645-655.

Daniell, G.J. & Gull, S.F. (1980). Maximun entropy algorithm applied to image enhancement, IEE Proc., 127E, 170-172.

Davies, A.R. & Anderssen, R.S. (1986). Optimisation in the regularisation of ill-posed problems. J. Austral. Math. Soc. Ser. B., 28, 114-133.

Frieden, B.R. (1972). Restoring with maximum likelihood and maximum entropy. J. Opt. Soc. Am., 62, 511-518.

Geman, S. & Geman, D. (1984) Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, IEEE Trans PAMI-6, 721-741

Gull, S.F. & Daniell, G.J. (1978). Image reconstruction from incomplete and noisy data. Nature, 272, 686-690.

Gull, S.F. & Daniell, G.J. (1979). The maximum entropy method. In Image Formation from Coherence Functions in Astronomy, ed. C. van Schooneveld, pp. 219-225, Reidel.

Gull, S.F. & Skilling, J. (1984). Maximum entropy method in image processing. IEE Proc.,131(F), 646-659.

Horne, K. (1985) Images of accretion discs I: The eclipse mapping method, Mon. Not. R. astr. Soc., 213, 129-141.

Jaynes, E.T. (1978). Where do we stand on maximum entropy? Reprinted in E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics, ed. R. Rosenkrantz, pp 211-314. Dordrecht 1983: Reidel.

Jaynes, E.T. (1986a). Bayesian methods: general background. In Maximum Entropy and Bayesian Methods in Applied Statistics. ed. J.H. Justice., pp 1-25. Cambridge University Press.

Jaynes, E.T. (1986b). Monkeys, kangaroos and N. In Maximum Entropy and Bayesian Methods in Applied Statistics, ed. J.H. Justice, pp. 26-58, Cambridge Univ. Press.

Kinderman, R. & Snell, J.L. (1980) Markov random fields and their applications. Amer. Math. Soc. Providence, RI.

Marsh, T.R. & Horne, K. (1989) Maximum entropy tomography of accretion discs from their emission lines. In these Proceedings.

Skilling, J. (1989a). Classic Maximum Entropy. In these Proceedings.

Skilling, J. (1989b). The eigenvalues of mega-dimensional matrices. In these Proceedings.

Tikhonov, A.N. & Arsenin, V.Y. (1977). Solutions of ill-posed problems. Wiley, New York.

Titterington, D.M. (1985) General structure of regularisation procedures in image reconstruction. Astron. Astrophys. 144, 381-387.

THE THREE PHASES OF STATISTICAL MECHANICS

W.T. Grandy, Jr.
Department of Physics and Astronomy
University of Wyoming
Laramie, Wyoming 82071 USA

ABSTRACT. The foundations of statistical mechanics are reviewed, based on the principle of maximum entropy, and this principle is shown to underlie the fundamental mechanisms of both equilibrium and nonequilibrium phenomena. Representative applications are provided—to quantum statistical systems in the first case, and to classical hydrodynamics in the second. Extensions of these ideas inspired by modern notions of chaos are mentioned, as well as ongoing work directed toward models of fully-developed turbulence.

Because a great deal of the discussion at this workshop is related as much to Bayes' theorem as it is to maximum entropy, let us begin by restating that theorem. If $\mathrm{P}(A|C)$ denotes the probability of a proposition A given hypothesis C (itself a proposition), then Bayes' theorem tells us that receipt of additional information B leads to a reassessment of that probability in the form

$$\mathrm{P}(A|BC) = \mathrm{P}(A|C)\frac{\mathrm{P}(B|AC)}{\mathrm{P}(B|C)}\,. \tag{1}$$

The proposition C is often thought of as prior information, so that $\mathrm{P}(A|C)$ is a prior probability of A based only on that information—called simply the *prior*. Then the left-hand side of Eq.(1) is called a posterior probability, and the fraction on the right-hand side is the ratio of the direct probability of the data to their prior probability. This theorem, of course, merely reflects the symmetry present in the standard rules for manipulating probabilities.

Now, when studying the behavior of some 10^{20} molecules in a given volume we are forced to the use of probability theory primarily because of an inability to formulate that many initial conditions, let alone follow the individual trajectories of the particles. The ensuing formalism—which we call *statistical mechanics*—provides us with some surprises in this respect, however, the first of which is that we never get past the prior in Eq.(1). That is, we rarely obtain more data to be employed in updating our estimates, so that it is necessary to make predictions

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 73–91.

about the system based only on our initial data. In addition, that initial data set usually consists of only a few pieces of *macroscopic* information, unlike the apparently-copious amount in a time series, say. Nevertheless, even with such burdensome constraints we are able to predict all other properties of such systems which we are desirous of measuring, and to develop all of thermodynamics. Surely this is something which John Wheeler would call 'Magic Without Magic' ! It is the exposition of this magic which is the concern of this lecture.

The many-body system in thermal equilibrium provides the quintessential example of the above remarks, for this state is *defined* through the observation that measured macroscopic quantities such as temperature remain unchanged under repeated measurement. The initial data provide our only information beyond a general knowledge of the problem, and so prior probabilities are the only ones we are led to consider. But, given sparse macroscopic data concerning constants of the motion, how is one to construct these priors? As is very well known by now, and first recognized by Gibbs, the optimum procedure in the present case is provided by the principle of maximum entropy. Hence, let us first review the relevant results of this prescription as rediscovered and reformulated by Jaynes many years later.

PRINCIPLE OF MAXIMUM ENTROPY

Suppose data to be available in the form of values of some function $f(x)$ at discrete values of the variable x (for convenience), such that these numbers can be interpreted as expectation values of $f(x)$ over the n possible mutually-exclusive and exhaustive alternatives $\{x_i\}$. That is, we make the identification

$$\langle f(x)\rangle \equiv \sum_i^n \mathrm{P}_i f(x_i)\,, \tag{2a}$$

such that

$$\sum_i^n \mathrm{P}_i = 1\,, \qquad \mathrm{P}_i \equiv \mathrm{P}(x_i) > 0\,. \tag{2b}$$

At first glance, the information provided in Eq.(2a) does not appear adequate to determine the probabilities in general. But if this is all we have—an extraordinarily underdetermined problem—some means for assigning these probabilities must be found. As first demonstrated by Shannon (1948), the optimum measure of uncertainty as to the appropriate distribution in this situation is the entropy of the probability distribution,

$$S(\mathrm{P}_1,\ldots,\mathrm{P}_n) \equiv -K\sum_i^n \mathrm{P}_i \ln \mathrm{P}_i, \qquad K > 0\,. \tag{3}$$

And about nine years later Jaynes (1957) stated the principle of maximum entropy (PME) as the optimal means for determining the set $\{P_i\}$ subject to the constraints (2).

The implied variational problem is most readily solved by means of Lagrange's method of undetermined multipliers, the result being

$$P_i = \frac{1}{Z(\lambda)} e^{-\lambda f(x_i)},$$
$$Z(\lambda) \equiv \sum_i e^{-\lambda f(x_i)}. \tag{4}$$

The partition function $Z(\lambda)$ is defined by substitution into the constraint equation (2b), and the Lagrange multiplier by substitution into (2a):

$$F \equiv \langle f(x) \rangle = -\frac{\partial}{\partial \lambda} \ln Z(\lambda). \tag{5}$$

The point is that λ is adjusted so as to reproduce the known datum F, which is *all* we can logically ask of any procedure of this kind. The expectation value of any other function $g(x)$ is then given by

$$\langle g(x) \rangle = \sum_i P_i g(x_i). \tag{6}$$

Some generalization is immediate and necessary. When data are specified about $m < n$ functions $f_r(x)$, we have constraints

$$F_r = \langle f_r(x) \rangle = \sum_i P_i f_r(x_i), \quad r = 1, \ldots, m < n. \tag{7}$$

A certain economy in notation is achieved by defining a 'scalar product'

$$\lambda \cdot f(x_i) \equiv \lambda_1 f_1(x_i) + \cdots + \lambda_m f_m(x_i). \tag{8}$$

Then the probability distribution maximizing the entropy subject to the constraints (7) is

$$P_i = \frac{1}{Z} e^{-\lambda \cdot f(x_i)},$$
$$Z(\lambda_1 \cdots \lambda_m) = \sum_i e^{-\lambda \cdot f(x_i)}, \tag{9}$$

with the Lagrange multipliers determined by the set of coupled differential equations

$$F_r = -\frac{\partial}{\partial \lambda_r} \ln Z(\lambda_1 \cdots \lambda_m), \quad r = 1, \ldots, m. \tag{10}$$

The maximum entropy itself is found from substituting Eqs.(9) into (3):

$$S_{\mathrm{I}} = K \ln Z + K\lambda \cdot F\,, \tag{11}$$

where the subscript I indicates that this is the information-theoretical entropy, providing a convenient distinction from other such quantities later. Note the explicit parameter dependence of S_{I}:

$$\frac{\partial S_{\mathrm{I}}}{\partial \lambda_r} = 0\,, \qquad \lambda_r = \frac{1}{K}\frac{\partial S_{\mathrm{I}}}{\partial F_r}\,, \tag{12}$$

which also constitutes a Legendre transformation between descriptions in terms of $\{F_r\}$ and $\{\lambda_r\}$.

Often the functions f_r will also depend on a common external parameter α, so that $f_r = f_r(x;\alpha)$. (They can also depend on an external parameter peculiar to each function). Then

$$Z = Z(\lambda_1 \cdots \lambda_m;\alpha)\,, \quad S_{\mathrm{I}} = S_{\mathrm{I}}(F_1 \cdots F_m;\alpha)\,. \tag{13}$$

If we define

$$\langle df_r\rangle \equiv \left\langle \frac{\partial f_r}{\partial \alpha}\right\rangle d\alpha\,, \tag{14}$$

then a short calculation yields for the total differential

$$dS_{\mathrm{I}} = K\lambda \cdot dQ\,, \tag{15}$$

where

$$dQ_r \equiv d\langle f_r\rangle - \langle df_r\rangle \tag{16}$$

is an inexact differential.

Prior to applying these results to something concrete, it is somewhat informative to mention one other way of looking at this problem of making predictions from very sparse data. What we really have here is a rather severe inverse problem, in that we are expected to estimate the causes of certain phenomena based on knowledge of very few effects. Abstractly, consider the general operator equation $F = Kf$, in which K is a known kernel specific to the particular problem under consideration. If f is known, then this mathematical expression constitutes the so-called direct problem and is solved by straightforward (though possibly quite difficult) calculation. The *inverse problem* consists of determining f if it is F that is known. Standard matrix, integral transform, and integral equations provide simple examples of such problems when all quantities are known fully and precisely. Often, however, F is only known incompletely and, rather than being deductive, the problem becomes one of inference based on incomplete information. One already senses a relation to the earlier discussion.

A common example of this scenario occurs when n trials of some process are carried out in which each trial has m possible outcomes, so that there are m^n conceivable outcomes to the total experiment. If the ith result occurs n_i times, it is useful to define frequencies

$$f_i \equiv \frac{n_i}{n}, \qquad 1 \leq i \leq n. \tag{17}$$

Suppose that we are given data in terms of M numbers F_j,

$$F_j = \sum_{1=1}^{m} K_{ij} f_i, \quad 1 \leq j \leq M < m, \tag{18}$$

where the K_{ij} are known, and asked to determine the true frequencies which might have produced these data. Although at first glance a seemingly outrageous request, we do have substantial prior information concerning such a problem. That is, we do know the number of ways a particular set of occupation numbers $\{n_i\}$ can be realized, for it is just the multinomial coefficient:

$$W \equiv \frac{n!}{(nf_1)! \cdots (nf_m)!}, \tag{19}$$

also called a *multiplicity factor*. Common sense then tells us that by maximizing W subject to the given data we determine that set $\{n_i\}$ that can be realized in the greatest number of ways. It is an equivalent procedure to maximize any monotonic function of W, and if n is very large the result will be that set that can be realized in the overwhelmingly greatest number of ways. Use of Stirling's formula then leads to the problem of maximizing

$$n^{-1} \log W = -\sum_i f_i \log f_i \tag{20}$$

subject to the constraints (18), and thus we have simply reformulated the prescription of maximizing the entropy. Clearly, the solution is

$$f_i = \frac{1}{Z} \exp\left\{\sum_j \lambda_j K_{ji}\right\}. \tag{21}$$

Classical statistical mechanics provides an example *par excellence* of this scenario, although it is equally useful in areas such as image processing, say.

1. Equilibrium Phenomena

Perhaps the simplest application of the PME is to the many-body system in thermal equilibrium, which we determine to be the correct state of the system

by finding that repeated measurement yields the same value for the total system energy, say: $\langle E \rangle$. There is only one set of alternatives, then, the system energy levels $E_i(V)$, and we consider just the one external parameter V, the system volume. Thus, under the constraints

$$\langle E \rangle = \sum_i \mathrm{P}_i E_i \,, \qquad \sum_i \mathrm{P}_i = 1 \,, \tag{22}$$

we find from above that

$$\mathrm{P}_i = \frac{1}{Z} e^{-\beta E_i} \,, \quad Z(\beta) = \sum_i e^{-\beta E_i} \,, \tag{23}$$

and the Lagrange multiplier β is determined from

$$\langle E \rangle = -\frac{\partial}{\partial \beta} \ln Z(\beta) \,. \tag{24}$$

The maximum entropy is then

$$S_{\mathrm{I}} = \kappa \ln Z + \kappa \beta \langle E \rangle \,, \tag{25}$$

where in this application we denote the constant K appearing in the definition of entropy by κ, for reasons which will become clear presently.

From the general expression (15) we see that

$$dS_{\mathrm{I}} = \kappa \beta dQ \,, \tag{26a}$$

with

$$dQ = d\langle E \rangle - \langle dE \rangle \,. \tag{26b}$$

But from the original discussion of $\langle dE \rangle$ it is clear immediately that this quantity is an element of mechanical work,

$$\begin{aligned} dW \equiv \langle dE \rangle &= \sum_i \mathrm{P}_i \left(\frac{\partial E_i}{\partial V} \right) dV \\ &= -\mathcal{P} \, dV \,, \end{aligned} \tag{27}$$

because this is just the definition of the physical pressure. Hence, Eq.(26b) is simply an expression of the first law of thermodynamics—$dE = dQ + dW$—owing to the physical meanings of the quantities involved. That is, dQ must be the element of heat introduced in classical thermodynamics, which is an inexact differential.

One now sees that the Lagrange multiplier β is determined immediately as the integrating factor for dQ, which is the way the Kelvin temperature scale is *defined.* Thus, β^{-1} must be proportional to the absolute temperature T. The units are determined by choosing the constant K to be Boltzmann's constant, κ, yielding the Kelvin temperature scale: $\beta = (\kappa T)^{-1}$. With these observations we have now identified S_{I} with the physical entropy of a system in thermal equilibrium, in which $dS = dQ/T$. These are just the equations of Gibbs for the canonical ensemble, thereby allowing us to write $S = S_{\mathrm{I}}$ and omit the expectation-value symbols in the present context. In addition, Eq.(25) can now be written in a more familiar form,

$$\begin{aligned} E - TS &= -\kappa T \ln Z \\ &\equiv F(T,V)\,, \end{aligned} \tag{28}$$

where F is called the *Helmholtz free energy.* The pressure is now written explicitly as

$$P = \beta^{-1} \frac{\partial}{\partial V} \ln Z\,. \tag{29}$$

and from Eq.(12) we obtain the well-known expression

$$\frac{1}{T} = \left(\frac{\partial S}{\partial E}\right)_V\,. \tag{30}$$

This last relation provides a Legendre transformation illustrating that, although in practice we usually measure the temperature in defining the equilibrium state, it is an equivalent procedure to measure the energy, as we have found convenient here.

There are now a number of other notions which logically should be discussed in some detail at this point, such as fluctuations, and stability conditions for the equilibrium state. But, owing to a lack of space here, we shall have to refer elsewhere for those details (e.g., Grandy, 1987). Two points, however, merit some comment now, the first having to do with the 'magic' of our prescription, mentioned earlier. It appears that we have put almost nothing into the PME and come out with all of classical thermodynamics. Some have felt that this is beyond belief, for we do not seem to have inserted the dynamics of the actual physical system into the development. This is only a mild subtlety, however, for in Eq.(22) we have presumed that it is possible to enumerate the spectrum of global energy levels of the system, and this is usually a rather nontrivial calculation. Thus, while the *structure* of the theory is a simple result of the rules of inference, its application to physical systems requires some decidedly serious knowledge of basic physical theory.

The second point to be made concerns the interpretation of statistical mechanics as an inverse problem, as in Eqs.(17)-(21). Let us recall Boltzmann's

expression for the physical entropy,

$$S = \kappa \log W\,, \tag{31}$$

where W is proportional to the number of microscopic states available to the system. That is, W is a multiplicity factor. From Eq.(28) we have $S = (E-F)/T$, so that now the probability for a macroscopic state of the system to be realized is

$$e^{-\beta F} = W\, e^{-\beta E}\,. \tag{32}$$

Because W is a rapidly increasing function of energy and both β and E are positive, one sees that the probability is sharply-peaked about the equilibrium energy.

A significant generalization occurs when the Gibbs algorithm is extended to a manifestly quantum-mechanical description in terms of the statistical operator, or density matrix $\hat{\rho}$. In thermal equilibrium one now considers linear Hermitian operators $\hat{F}_i$ which are constants of the motion in the quantum-mechanical sense, and which possibly are noncommuting. Expectation values are written

$$\langle \hat{F} \rangle = \mathrm{Tr}(\hat{\rho}\hat{F})\,, \qquad \mathrm{Tr}\,\hat{\rho} = 1\,, \tag{33}$$

incorporating both aspects of probability to be found in quantum statistical mechanics: that arising in connection with incomplete information, and that intrinsic to quantum mechanics itself. The entropy is now defined as

$$S = -\kappa\, \mathrm{Tr}\,\hat{\rho} \ln \hat{\rho}\,, \quad \kappa > 0\,. \tag{34}$$

Maximization of S subject to the constraints (33) then yields the statistical operator

$$\hat{\rho} = \frac{1}{Z} e^{-\lambda_1 \hat{F}_1 - \cdots - \lambda_m \hat{F}_m}\,, \tag{35a}$$

with partition function

$$Z(\lambda_1 \cdots \lambda_m) = \mathrm{Tr}\, e^{-\lambda_1 \hat{F}_1 - \cdots - \lambda_m \hat{F}_m}\,. \tag{35b}$$

The Lagrange multipliers are once again found from a set of coupled differential equations:

$$\langle \hat{F}_k \rangle = -\frac{\partial}{\partial \lambda_k} \ln Z\,, \quad k = 1, \ldots, m\,. \tag{36}$$

One finds for the maximum entropy

$$S_{\mathrm{I}} = \kappa \ln Z + \lambda \cdot \langle \hat{F} \rangle\,, \tag{37}$$

in the dot-product notation introduced earlier, and

$$\lambda_k = \frac{1}{\kappa}\frac{\partial S_{\mathrm{I}}}{\partial \langle \hat{F}_k \rangle} . \tag{38}$$

Except for the dual interpretation of expectation values, these equations are identical to those obtained above. The canonical ensemble is regained if only the total energy (or temperature) is specified, which corresponds to an expectation value of the N-body, time-independent Hamiltonian $\hat{\mathrm{H}}_N$. The statistical operator is then

$$\hat{\rho} = e^{\beta(\hat{F} - \hat{\mathrm{H}}_N)} , \tag{39}$$

where $\hat{F}$ is the unit operator times $F = -\kappa T \ln Z$. Because $\hat{F}$ commutes with $\hat{\mathrm{H}}_N$ and is conserved, the trace of the last expression yields immediately the partition function

$$Z(\beta) = \mathrm{Tr}\, e^{-\beta \hat{\mathrm{H}}_N} . \tag{40}$$

Further examples are plentiful, and we mention just two. If, in addition to the total energy, the system is rotating uniformly with angular velocity $\boldsymbol{\omega}$ and we measure a component of angular momentum $\hat{J}_i$, we can consider the system to be in thermal equilibrium in its rest frame (Gibbs, 1902; p.39). The resulting description is called the *rotational ensemble*, with statistical operator

$$\hat{\rho}_r = \frac{1}{Z_r} e^{-\beta(\hat{\mathrm{H}}_N - \omega_i \hat{J}_i)} . \tag{41}$$

Should both the total energy and total particle number N be provided—$\langle \hat{\mathrm{H}} \rangle$ and $\langle \hat{N} \rangle$—then we obtain the equations of the grand canonical ensemble, which are expressed in terms of the grand partition function:

$$Z_G = \sum_{N=0}^{\infty} e^{\beta \mu N}\ \mathrm{Tr}\, e^{-\beta \hat{\mathrm{H}}_N} , \tag{42}$$

where the chemical potential μ provides the additional Lagrange multiplier corresponding to conservation of particle number. The average number of particles, Helmholtz free energy, and total energy per particle, respectively, are given by

$$N = \frac{1}{\beta}\frac{\partial}{\partial \mu} \ln Z_G , \tag{43a}$$

$$F = G - \beta^{-1} \ln Z_G , \tag{43b}$$

$$\frac{E}{N} = \mu - \frac{1}{N}\frac{\partial}{\partial \beta}\left(V^{-1} \ln Z_G\right) , \tag{43c}$$

and $G = \mu N$ is called the Gibbs function.

As an aside, we note that there *is* a sense in which the spirit of Bayes' theorem emerges here, and which is a major strength of the PME. No probabilistic theory can guarantee its predictions, of course, and it may happen that the latter do not agree with observation. In that event the theory is telling us that there are constraints operating of which we were unaware, and hence alerts us to the possible existence of new physics. This call for re-assessment is clearly Bayesian in spirit.

2. Nonequilibrium Systems

For systems clearly not in thermal equilibrium, and for equilibrium systems in which the observed quantities are not constants of the motion, a much more general algorithm is needed than that developed in the preceding section. With reference to Eq.(1), one might think that we are ready finally to move past the prior and employ the full content of Bayes' theorem, for if the need for 'updating' were ever evident one would think it would be for time-dependent problems of this kind. We shall see, however, that this is not the case, and that we are still concerned primarily with prior probabilities.

A detailed understanding of arbitrary irreversible problems necessarily passes through three distinct stages of calculation:

(*i*) Construction of the initial 'ensemble', or statistical operator $\hat{\rho}(t_0)$, describing the initial state of the system of interest;

(*ii*) Solution of he microscopic dynamical problem so as to obtain the time-evolved operator $\hat{\rho}(t)$;

(*iii*) Prediction of the final macroscopic physical quantities of interest using $\hat{\rho}(t)$.

Stage (*iii*) does not present any difficulties of principle, for one merely calculates expectation values of the operators of interest via the prescription $\langle \hat{F} \rangle = \mathrm{Tr}(\hat{\rho}\hat{F})$—a procedure justified within the theory itself. Stage (*ii*) is technically the most difficult, but also the one which has received the most attention over the past three decades. In one way or another, and usually to some degree of approximation, one must solve the equation of motion

$$i\hbar \frac{\partial \hat{\rho}(t)}{\partial t} = [\hat{\mathrm{H}}, \hat{\rho}(t)]\,, \tag{44}$$

or the equivalent for open systems. The fundamental aspect of the calculational stage is to solve this equation of motion subject to the initial conditions describing the physical situation, and this brings us back to stage (*i*).

We presume the initial data defining the nonequilibrium state of a system can be put into the form of expectation values of a number of Heisenberg operators $\hat{F}_i(\mathbf{x}, t)$, for which the variables $\mathbf{x}$ and t vary over some information-gathering space-time interval $R_i(\mathbf{x}, t)$. The general time development of the Heisenberg

operators is described by the unitary transformation

$$\hat{F}_i(t) = \hat{U}^\dagger(t,t_0)\hat{F}(t_0)\hat{U}(t,t_0)\,, \tag{45}$$

where for the moment we suppress the spatial variable. The time-development operators are solutions of the equation of motion

$$i\hbar\frac{d\hat{U}(t,t_0)}{dt} = \hat{\mathrm{H}}(t)\hat{U}(t,t_0)\,, \tag{46}$$

subject to the initial condition $\hat{U}(t_0,t_0) = \hat{1}$. Should $\hat{\mathrm{H}}$ not be explicitly dependent on the time, Eq.(46) has the solution

$$\hat{U}_0(t,t_0) = e^{i(t-t_0)/\hbar}\,; \tag{47}$$

otherwise, it is very difficult to find an expression for $\hat{U}$ in closed form.

Although the statistical operator $\hat{\rho}(t_0)$ remains stationary in the Heisenberg picture, this is conventionally taken to coincide with the Schrödinger picture at $t = t_0$. In the latter $\hat{\rho}$ evolves in time according to the prescription

$$\hat{\rho}(t) = \hat{U}(t,t_0)\hat{\rho}(t_0)\hat{U}^\dagger(t,t_0)\,, \tag{48}$$

which is equivalent to the equation of motion

$$i\hbar\frac{d\hat{\rho}(t)}{dt} = [\hat{\mathrm{H}}(t),\hat{\rho}(t)]\,. \tag{49}$$

If now we are given several pieces of data $\langle\hat{F}_k(\mathbf{x},t)\rangle$ over space-time regions $R_k(\mathbf{x},t)$, then we can once again construct the initial statistical operator encompassing only this information by maximizing the entropy subject to these constraints. The result is

$$\hat{\rho} = \frac{1}{Z}\exp\left[\sum_k\int_{R_k}\lambda_k(\mathbf{x},t)\hat{F}_k(\mathbf{x},t)\,d^3x\,dt\right]\,, \tag{50a}$$

where now

$$Z[\{\lambda_k\}] = \mathrm{Tr}\exp\left[\sum_k\int_{R_k}\lambda_k(\mathbf{x},t)\hat{F}_k(\mathbf{x},t)\,d^3x\,dt\right] \tag{50b}$$

is called the *partition functional.* The Lagrange-multiplier functions $\lambda_k(\mathbf{x},t)$ are identified from the initial data by means of the coupled set of functional differential equations

$$\langle\hat{F}_k(\mathbf{x},t)\rangle \equiv \mathrm{Tr}[\hat{\rho}\hat{F}_k(\mathbf{x},t)] = \frac{\delta}{\delta\lambda_k(\mathbf{x},t)}\ln Z\,, \quad (\mathbf{x},t)\in R_k\,, \tag{51}$$

and the predicted expectation of any other Heisenberg operator $\hat{J}$ at $(\mathbf{x},t)$ is

$$\langle \hat{J}(\mathbf{x},t)\rangle = \mathrm{Tr}[\hat{\rho}\hat{J}(\mathbf{x},t)] = \mathrm{Tr}[\hat{\rho}(t)\hat{J}(\mathbf{x})]\,. \tag{52}$$

It is important to emphasize that the $\hat{\rho}$ constructed here is nothing more than the initial statistical operator describing only what is known about the initial state of the system. Aside from a possible clearly specified driving mode, this is generally all one can hope to know regarding any experimentally reproducible situation. Although a number of authors over the years have attempted to construct a $\hat{\rho}(t)$ intended to describe the nonequilibrium system for all time, that now appears to be an entirely unrealistic goal. Thus, we are still working on the prior in Bayes' theorem, at least in the absence of specified dynamical or thermal driving. One should also note here that the above expressions contain as a special case the well-known theory of dynamical response (e.g., Grandy, 1988).

LINEAR APPROXIMATION

The preceding expressions are rather difficult to employ in practice, for they are exact and completely nonlinear. Often, however, a linear approximation is adequate for describing the various phenomena. Suppose that initially the system is described by the equilibrium ensemble

$$\begin{aligned} \hat{\rho}_0 &= \frac{1}{Z_0(\beta)} e^{-\beta \hat{\mathrm{H}}}\,, \qquad \hat{\mathrm{H}} \neq \hat{\mathrm{H}}(t) \\ \langle \hat{F}\rangle_0 &= \mathrm{Tr}(\hat{\rho}_0 \hat{F})\,. \end{aligned} \tag{53}$$

Again we suppress the spatial variable temporarily and consider new data $\langle \hat{F}(t)\rangle$ obtained over a time interval $-\tau \leq t \leq 0$, so that the new description is

$$\begin{aligned} \hat{\rho} &= \frac{1}{Z} \exp\left[-\beta \hat{\mathrm{H}} + \int_{-\tau}^{0} \lambda(t)\hat{F}(t)\,dt\right], \\ Z[\beta,\lambda(t)] &= \mathrm{Tr}\exp\left[-\beta \hat{\mathrm{H}} + \int_{-\tau}^{0} \lambda(t)\hat{F}(t)\,dt\right]. \end{aligned} \tag{54}$$

The explicit time dependence here is given by

$$\hat{F}(t) = \hat{U}_0^{\dagger}(t)\hat{F}\hat{U}_0(t)\,, \quad \hat{U}_0(t) = e^{-i\hat{\mathrm{H}}t/\hbar}\,. \tag{55}$$

Define the *Kubo transform* as

$$\overline{\hat{B}} \equiv \int_0^1 e^{-x\hat{A}}\,\hat{B}e^{x\hat{A}}\,dx\,, \tag{56}$$

and the *covariance function*

$$K_{CB} \equiv \langle \overline{\hat{B}}\hat{C}\rangle_0 - \langle \hat{B}\rangle_0\langle \hat{C}\rangle_0 = K_{BC}\,. \tag{57}$$

Then, through leading order in λ, the expectation value of some other Heisenberg operator at some time t is

$$\langle \hat{C}(t)\rangle - \langle \hat{C}\rangle_0 \simeq \int_{-\tau}^{0} K_{CF}(t,t')\lambda(t')\,dt'\,, \tag{58}$$

and

$$K_{CF}(t,t') \equiv \langle \overline{\hat{F}(t')}\hat{C}(t)\rangle_0 - \langle \hat{F}\rangle_0\langle \hat{C}\rangle_0\,. \tag{59}$$

If $\hat{\mathrm{H}} \neq \hat{\mathrm{H}}(t)$ we have a reciprocity relation exhibiting time translational invariance:

$$K_{CF}(t,t') = K_{CF}(t-t') = K_{FC}(t'-t)\,. \tag{60}$$

Finally, a complete generalization yields for the linear approximation to expectation values in a region $(\mathbf{x},t)$

$$\langle \hat{C}(\mathbf{x},t)\rangle - \langle \hat{C}(\mathbf{x})\rangle_0 = \int_R K_{CF}(\mathbf{x},t;\mathbf{x}',t')\lambda(\mathbf{x}',t')\,d^3x'\,dt'\,, \tag{61a}$$

with covariance function

$$\begin{aligned} K_{CF}(\mathbf{x},t;\mathbf{x}',t') &= \langle \overline{\hat{F}(\mathbf{x}',t')}\hat{C}(\mathbf{x},t)\rangle_0 - \langle \hat{F}(\mathbf{x}')\rangle_0\langle \hat{C}(\mathbf{x})\rangle_0 \\ &= \frac{\delta^2}{\delta\lambda_C(\mathbf{x},t)\,\delta\lambda_F(\mathbf{x}',t')}\ln Z\,. \end{aligned} \tag{61b}$$

As an example of this formalism we consider some of the equations of linear hydrodynamics.

LINEAR HYDRODYNAMICS

When a many-body system is perturbed from thermal equilibrium the resulting situation is one of considerable chaos, compounded by the fact that in a fluid containing a very large number of particles there is a corresponding large number of degrees of freedom. If the system is then allowed to relax, most of these degrees of freedom return rather quickly to their equilibrium values in ways determined by the microscopic characteristics of the system. But this relaxation can be described on the macroscopic level by only a few long-lived modes which decay relatively slowly, and these modes are related to the locally-conserved densities

in the medium. That is, local excesses of these quantities can disappear neither locally nor quickly, but must relax by spreading out over the entire system.

In a simple fluid the locally-conserved quantities are the number, energy, and momentum densities, resulting in five long-lived hydrodynamic modes. We denote these densities generically by the symbol $\hat{e}(\mathbf{x},t)$, along with associated current densities $\hat{\mathbf{J}}(\mathbf{x},t)$, and recall that local conservation equations take the form

$$\partial_t \hat{e}(\mathbf{x},t) + \nabla \cdot \hat{\mathbf{J}}(\mathbf{x},t) = 0\,, \tag{62}$$

which are microscopic operator equations. When appropriate driving mechanisms are introduced (e.g, Grandy, 1988), one derives from this last expression a *macroscopic* conservation law:

$$\partial_t \langle \hat{e}(\mathbf{x},t) \rangle_t + \nabla \cdot \langle \hat{\mathbf{J}}(\mathbf{x},t) \rangle_t = \dot{\sigma}(\mathbf{x},t)\,, \tag{63}$$

where $\dot{\sigma}$ is the rate at which the source drives the density.

Of primary interest at the moment is the momentum density, for which the current is the stress tensor $\hat{T}_{ij}(\mathbf{x},t)$, and the Lagrange-multiplier function associated with the momentum density is identified as $\beta m \mathbf{v}(\mathbf{x},t)$, where $\mathbf{v}$ is referred to as the velocity field. Let us consider an incompressible fluid, $\nabla \cdot \mathbf{v} = 0$. Then some further calculation converts the macroscopic equation (63) into what are usually called the Navier-Stokes equations:

$$mn_0\big(\partial_t v_i + [(\mathbf{v}\cdot\nabla)\mathbf{v}]_i\big) = -\partial_j P \delta_{ij} + \partial_k p_{ik} + mn_0 F_i\,, \tag{64}$$

which are nonlinear in the fluid velocity. (Although the formal approximation we have made is linear in the departure from equilibrium, the nonlinearity here arises from the convective contribution to the total time derivative.) Indeed, we see that the macroscopic equations are just equations of motion for the Lagrange-multiplier functions. The notation is as follows: n_0 is the equilibrium number density of the system of particles with mass m, P is the pressure, we have represented the possible driving force by F_i, and p_{ik} is the shear tensor. Explicitly,

$$p_{ik} \equiv \eta\left(\frac{\partial v_k}{\partial x^i} + \frac{\partial v_i}{\partial x^k} - \tfrac{2}{3}\delta_{k\ell}\frac{\partial v_\ell}{\partial x^\ell}\right) + \varsigma \frac{\partial v_\ell}{\partial x^\ell}\delta_{ki}\,. \tag{65}$$

The transport coefficients in this expression are derived within the linearized theory in terms of space-integrated covariance functions. Specifically, the shear viscosity in the steady state is

$$\eta \simeq \frac{\beta m}{V} \lim_{\epsilon\to 0^+} -\int_0^\infty e^{-\epsilon t} \big\langle \overline{\hat{T}^{ij}(-t)}\hat{T}^{ij}\big\rangle_0 \, dt\,, \qquad i \neq j\,, \tag{66a}$$

whereas the bulk viscosity is expressed in terms of the trace $\hat{T} \equiv \sum_i \hat{T}_i^i$:

$$\varsigma \simeq \frac{1}{9}\frac{\beta m}{V} \lim_{\epsilon \to 0^+} \int_0^\infty e^{-\epsilon t} \langle \overline{\hat{T}(-t)}\hat{T} \rangle_0 \, dt \,. \tag{66b}$$

At this point one would think that the standard many-body theory has completed its task, resulting in macroscopic deterministic equations such as those of Eq.(64). Probability theory has fulfilled its role in describing these systems and can retire with honor. Recent investigations into the dynamics of classical systems give one reason to pause, however, and it is quite possible that there is one more phase to go through.

3. Macroscopic Processes

Ever since the work of Poincaré it has been known that the number of integrable dynamical systems is severely limited, and that nonlinear equations of motion possess solutions exhibiting highly irregular behavior for given parameter values. Only relatively recently, however, have the enormous advances in computational ability made it possible to study this behavior in any detail. Examples of deterministic dynamical systems in which irregular or 'chaotic' motions can occur are now commonplace. Hénon and Heiles (1964) presented an early and important physical model in connection with the distribution of stellar velocities within the galaxy. One is compelled to ask what bearing, if any, these developments might have on the statistical-mechanical description of many-body systems.

If the microscopic equations are nonlinear one must then allow for the possibility that the microscopic trajectories may exhibit irregular behavior. But a major role of the microscopic equations of motion is to provide us with an enumeration of the various alternatives over which the probability index ranges, and this has always been a technically difficult matter irrespective of whether or not those equations are linear. Introspection suggests that nothing really changes in this respect if the particle equations are highly nonlinear, for the procedures remain the same. That is, because it always considers the *full* equations of motion, the PME is rather transparent to the actual structure of those equations. If those equations are nonlinear it is possible that new phenomena *could* appear in our macroscopic predictions, but that will not affect the way in which we *make* those predictions. As long as the spectrum can be presented in principle, however difficult in practice, then it matters little how irregular the microscopic motion may be—that, after all, is just the point of statistical mechanics!

Entirely different conclusions emerge, however, with respect to *macroscopic* motions of a system if the governing equations are nonlinear, for we usually wish to—and often can—follow the trajectories in this case. We continue to use the example of conventional hydrodynamics for an incompressible fluid, so that

the equations of motion are just the Navier-Stokes equations (64). As already pointed out, these are a linear approximation in the Lagrange-multiplier-function $\mathbf{v}(\mathbf{x},t)$, and higher-order equations can be obtained in a straightforward way from the perturbation expansion of the nonequilibrium expectation values. The nonlinearity in Eq.(64) arises solely from the convection term in the derivative on the left-hand side. It is useful to rewrite these equations in terms of dimensionless variables by introducing a characteristic speed u of the fluid, and a characteristic length ℓ. Then, in vector notation, and without the external-force term, Eq.(64) becomes

$$\partial_t \mathbf{v} + R(\mathbf{v}\cdot\nabla)\mathbf{v} = -\nabla p + \nabla^2 \mathbf{v}\,, \tag{67}$$

where $R \equiv u\ell/\nu$ is the *Reynolds number*, and ν is called the kinematic viscosity (the ratio of viscosity coefficient to the density). Clearly, the effect of nonlinearity is controlled completely by R, and when $R = 0$ these are known as the Stokes equations.

In some systems the experimentalist observes a series of spectacular instabilities as R increases from zero past some critical value, and eventually complete turbulence in the fluid flow emerges for sufficiently large Reynolds numbers. (There is some controversy on this point, and we shall return to it below.) Many of these stages in the progression to fully-developed turbulence for various systems have been captured photographically, and can be observed in the beautiful collection of Van Dyke (1982). One believes that this progression is described theoretically by Eq.(67) as R varies, but these equations are very difficult either to solve or to analyze in general. Nevertheless, in some applications it is possible to approximate them without destroying the essential nonlinearity.

The now-classic example of this latter procedure begins with the attempt by Saltzman (1962) to model the Rayleigh-Bénard instability in two dimensions by Fourier expansion in Eq.(67) and truncation into a set of ordinary differential equations. Shortly thereafter these equations were adopted by Lorenz (1963) as a model for the unpredictable behavior of the weather, and were studied extensively by him—with remarkable results. These reduced equations for convection of the fluid are

$$\begin{aligned} \frac{dx}{dt} &= \sigma(y-x)\,,\\ \frac{dy}{dt} &= rx - y - xz\,,\\ \frac{dz}{dt} &= xy - bz\,, \end{aligned} \tag{68}$$

where $x(t)$ is proportional to the amplitude of convective motion, $y(t)$ and $z(t)$ are proportional to two temperature modes, σ is called the *Prandtl number* (the ratio of kinematic viscosity to the thermal diffusivity), r is the *Rayleigh number*

in units of its critical value (the convective analog of the Reynolds number), and b is a constant related to the wavenumber of the fundamental mode.

This last set of equations is completely deterministic, so that we can study (with the aid of the computer) the trajectories generated from various initial conditions by fixing σ and b at the values adopted by Lorenz, say, and varying r. For $r < 1$ all trajectories are attracted to a stable solution at the origin of the variables in Eq.(68): $x = y = z = 0$. If r exceeds unity by much the model is no longer physically realistic, but nevertheless still worth studying. There are two stable solutions for $1 < r < 13.9$, to which all stable trajectories are attracted, and in the region $13.9 < r < 24.1$ a complicated transition begins to take place. For $r > 24.1$ all trajectories are attracted toward a subspace in which they wander 'chaotically' forever. That is, the motion is highly irregular and essentially unpredictable. This subspace is called a *strange attractor.*

To use the word 'chaos' here is to risk conveying an impression of motion which is not deterministic. In reality, the motion is no more chaotic than that of particles in an equilibrium gas—they all obey well-defined equations of motion. But 'chaos' now assumes a more technical meaning—namely, the result of an extraordinary sensitivity to initial conditions. In the chaotic regime it is virtually impossible to specify initial conditions precisely enough to be sure of the ensuing trajectory, and it is in this sense we employ the above phrase 'essentially unpredictable'.

The importance of these results in the present context lies with the possibility of being able to describe turbulence in some detail as a solution of the Navier-Stokes equations, say. As noted above, it has been thought for many years that smooth laminar flow will become unstable and cascade into turbulence eventually when R exceeds some critical value. This is a *macroscopic* phenomenon, and so would seem to be outside the purview of statistical mechanics. That is, the role of the latter should cease with the derivation of the macroscopic equations of motion and provision for calculation of the relevant parameters.

But completely-developed turbulence is more than just 'chaotic' motion, and the phenomena uncovered by study of the Lorenz equations only provide us with a beginning. There is, for example, some current controversy as to whether a final state of fully-developed turbulence is always attainable, or whether so-called coherent structures persist indefinitely in some systems (e.g., Lesieur, 1987). Nevertheless, the onset of chaos may well signal the approach to a turbulent state, which is intrinsically nonequilibrium and collective in nature. As the parameters of the macroscopic equations continue to change, and full turbulence develops, one realizes that the number of *macroscopic* degrees of freedom has increased enormously (owing to nonlinearity). There are now a great many possible trajectories available to the system, but it is very difficult to know which is taken owing to the extreme sensitivity to initial conditions. Although the system state may well be described by only a few macroscopic variables—or 'supermacroscopic'

variables—just what that state may be is difficult to determine exactly. It is as if one did not really know the precise initial conditions.

Everything begins to sound familiar at this point, as if statistical mechanics were emerging anew, but on a higher level. In the problems of hydrodynamics it appears that the elementary volumes associated with the velocity field $\mathbf{v}(\mathbf{x}, t)$ play the role of the basic units, or 'particles', with laminar flow being analogous to the equilibrium state. [Years ago Hopf (1952) attempted to construct a statistical theory of turbulence based on much the same point of view, but he did not have available perspectives which were only to emerge from the more recent computer-assisted understanding of chaos.] Some systems can then pass through a number of 'second-order phase transitions', corresponding to the hydrodynamic instabilities, and for a given range of parameter values the various states are both stable and reproducible. A striking example of this kind of sequence is provided by Couette flow between rotating concentric cylinders (e.g., DiPrima and Swinney, 1985).

In order to verify such notions as these, however, there are a number of questions which must be addressed and resolve to a degree that has not yet been achieved—questions suggested in part by our experience with the microscopic theory. For example, one must identify the experimentally reproducible phenomena on the macroscopic level and construct a definite catalog. What are the macroscopic quantities we can measure or observe both on and off the strange attractor? One known class of quantities consists of power spectra of the velocity field, but it is not clear that this class is sufficient to characterize the phenomena adequately.

Precisely how the notion of 'insensitivity to initial conditions' arises in a specific real problem of this kind is not entirely clear. But the suggestion is strong that, as R increases and the nonlinearities become increasingly more important, the observed instabilities signal a breakdown in the severely rigid uniformity of laminar flow, or in 'coherent' structures. The onset of turbulence is characterized by 'insensitivities' which are analogous to the ignorance of microscopic initial conditions leading to the statistical description of many-body systems discussed earlier. A higher-level statistical description will require construction of a probability distribution over possible macroscopic trajectories, which in turn requires a very clear understanding of what kind of information can be obtained and how it can be utilized for that purpose. As these points are clarified it is quite possible there will emerge a 'canonical' form of probability distribution every bit as effective as that of Gibbs in describing ordinary thermodynamics. We are far from reaching that point, however, and the observations made here merely serve to outline a program in need of a great deal of development.

REFERENCES

DiPrima, R.C., and H.L.Swinney: 1985, 'Instabilities and Transition in Flow Between Concentric Rotating Cylinders', in H.L. Swinney and J.P. Gollub (eds.), *Hydrodynamic Instabilities and the Transition to Turbulence*, Springer-Verlag, Berlin.

Grandy, W.T., Jr.: 1987, *Foundations of Statistical Mechanics, Volume I: Equilibrium Theory*, Reidel, Dordrecht.

Grandy, W.T. Jr.: 1988, *Foundations of Statistical Mechanics, Volume II: Nonequilibrium Phenomena*, Reidel, Dordrecht.

Hénon, M., and C.Heiles: 1964, 'The Applicability of the Third Integral of the Motion: Some Numerical Experiments', *Astron. J.* **69**, 73.

Hopf, E.: 1952, 'Statistical Hydromechanics and Functional Calculus', *J. Rat. Mech. Anal.* **1**, 87.

Jaynes, E.T.: 1957, 'Information Theory and Statistical Mechanics', *Phys. Rev.* **106**, 620.

Lesieur, M.: 1987, *Turbulence in Fluids*, Nijhoff, Dordrecht.

Lorenz, E.N.: 1963, 'Deterministic Nonperiodic Flows', *J. Atmos. Sci.* **20**, 130.

Saltzman, B.: 1962, 'Finite Amplitude Free Convection as an Initial Value Problem–I', *J. Atmos. Sci.* **19**, 329.

Shannon, C.E.: 1948, 'A Mathematical Theory of Communication', *Bell System Tech. J.* **27**, 379, 623.

Van Dyke, M.: 1982, *An Album of Fluid Motion*, The Parabolic Press, Stanford, CA.

BELL'S THEOREM. INFERENCE AND QUANTUM TRANSACTIONS

A.J.M. Garrett

Department of Physics & Astronomy
University of Glasgow
GLASGOW G12 8QQ
Scotland
U.K.

ABSTRACT. Bell's theorem is expounded as an analysis in Bayesian inference. Assuming the result of a spin measurement on a particle is governed by a causal variable internal (hidden, "local") to the particle, one learns about it by making a spin measurement: thence about the internal variable of a second particle correlated with the first; and from there predicts the probabilistic result of spin measurements on the second particle. Such predictions are violated by experiment: locality/causality fails. The statistical nature of the observations rules out acausal signalling, superluminal or otherwise. Quantum mechanics is irrelevant to this reasoning, although its correct predictions of experiment imply it is a nonlocal/acausal theory. Cramer's new transactional interpretation of the quantum formalism, which incorporates this feature, is advocated as an invaluable way of envisaging quantum processes. The usual paradoxes melt before it, and one, the "delayed choice" experiment, is interpreted in detail.

1. BAYESIAN INTERPRETATION OF BELL'S THEOREM

In this section it is shown that no theory, postulating that the results of spin measurements on a particle are causally governed by variables internal to the particle, can reproduce the findings of measurements on particle pairs. Quantum theory is irrelevant to the argument: to test whether nature is nonlocal/acausal, it is the class of local/causal theories which must be compared with experiment. Nevertheless quantum theory correctly predicts the outcome of these experiments, and is discussed in the light of this later on.

Let us begin by postulating the existence within a particle of internal (i.e. local, "hidden") variables. There may be any number of these, denoted collectively by λ. We might hope to learn about these by measuring the spin of a particle in a particular direction, and reasoning back to λ using Bayes' theorem. From there, we hope to make (probabilistic) predictions about future spin measurements on that particle in any direction.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 93–105.

Unfortunately the disturbance caused by measuring the spin would in general alter the value of λ in an unknown way; so with a single particle we cannot predict the future, only improve our knowledge of the past. This problem is circumvented by using two particles which are correlated in some way, an ingenious and well-known idea due originally to Einstein, Podolsky and Rosen (1935); application to spin correlations is due to Bohm (1951). By measuring the spin of the first particle in a selected direction, we infer something about its internal variable; through the correlation, we then learn about the internal variable of the second particle; and from there, we make probabilistic predictions about spin measurements on the second particle. Locality has it that the internal variable of the second particle is unchanged by the measurement on the first, and causality that the first measurement is uninfluenced by the second. Below, this analysis is made quantitative, using the laws of probability as consistent laws of inference (Cox 1946). This is the famous analysis of Bell (1964), rephrased in Bayesian language.

For two photons or two spin-½ particles, correlated by having zero net angular momentum, results lie outside the predictions of causal internal variable theories (Bell 1964). Because we are only able to predict on a statistical basis, many pairs of particles are examined at each direction setting of the apparati. Nevertheless, locality/causality fails in this situation, and we accordingly conclude that nature is nonlocal/acausal.

For simplicity we work with those particles observed to have only two spin states, which we call $\pm$½. It has been claimed that nonlocality need not be implied should particles have three exclusive categories: +½, -½, and undetectable. (The idea is due originally to Pearle, 1970.) Quite apart from the ad hoc nature of this assumption, which has no correspondent in quantum mechanics, the result still stands if the analysis is applied only to those particles in the first two categories. One is free to seek nonlocality/acausality wherever one wants. The notation $S^{\pm}|\underline{v}$ denotes "spin measurement in direction $\underline{v}$ (a unit vector) is $\pm$½"; this will accord with standard probability notation. S is not a variable but a measurement. I denotes the information that the total angular momentum of a pair is zero, or (in quantum parlance) that the pair is in a singlet state, and that both members of the pair are detected. Subscripts $_1$ and $_2$ denote the two particles of a pair, ordered according to the times of measurement in the laboratory frame.

Let us now calculate the probabilities of spin measurements on each particle, $p(S_1^{\pm}|\underline{v}_1,I)$ and $p(S_2^{\pm}|\underline{v}_2,S_1^{\pm},\underline{v}_1,I)$. First, $p(S_1^{\pm}|\underline{v}_1,I)$ is a marginal distribution over the internal variables of the particles:

$$p(S_1^{\pm}|\underline{v}_1,I) = \iint d\lambda_1 d\lambda_2 \; P(\lambda_1,\lambda_2,S_1^{\pm}|\underline{v}_1,I) \qquad (1)$$

$$= \iint d\lambda_1 d\lambda_2 \; P(\lambda_1,\lambda_2|I)\; p(S_1^{\pm}|\underline{v}_1,\lambda_1,\lambda_2,I) \qquad (2)$$

where a capital P denotes a probability density. Nothing other than the sum and product rules are involved here, and there is no implication that (1) is an ensemble average over the internal variables, for this is a procedure of inference. Because of the correlation I, $P(\lambda_1,\lambda_2|I)$ is not separable, though indifference demands it be symmetrical. (Bell's analysis combined λ_1 and λ_2 into a single variable.) Next, we connect to the physics by making the local/causal assumption that the probability of measured spin values depends only on a particle's internal variable and the specified direction of measurement:

$$p(S_1^{\pm}|\underline{v}_1,\lambda_1,\text{ anything else}) = F_{\pm}(\underline{v}_1,\lambda_1) \tag{3}$$

where $F_{\pm}$ are definite functions, complementary for the same value of argument. Our reasoning remains valid no matter how fine, or fractal, the structural dependence of $F_{\pm}$ on $\underline{v}$. We shall in fact infer from the observed exact (anti)correlation $S_2=-S_1$ when the directions are identical ($\underline{v}_2=\underline{v}_1$) that $F_{\pm} = 0$ or 1 everywhere in $(\underline{v},\lambda)$-space.

Experiment shows that

$$p(S_1^{+}|\underline{v}_1,I) = p(S_1^{-}|\underline{v}_1,I) = \tfrac{1}{2}. \tag{4}$$

Define for convenience the "expectation"

$$A(\underline{v},\lambda) = (+1)F_{+}(\underline{v},\lambda) + (-1)F_{-}(\underline{v},\lambda), \tag{5}$$

which has the property

$$|A(\underline{v},\lambda)| \leqslant 1. \tag{6}$$

From (2)-(5) we have, on subtracting $p(S_1^{-}|\underline{v}_1,I)$ from $p(S_1^{+}|\underline{v}_1,I)$,

$$0 = \int\!\!\int d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ A(\underline{v}_1,\lambda_1). \tag{7}$$

Adding merely confirms normalisation:

$$1 = \int\!\!\int d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I). \tag{8}$$

Now examine the probabilities of measured spin values on the second particle of the pair, conditioned on the result of the first measurement. From the laws of probability,

$$p(S_2^{\pm}|\underline{v}_2,S_1^{\pm},\underline{v}_1,I) = p(S_1^{\pm},S_2^{\pm}|\underline{v}_2,\underline{v}_1,I) \,/\, p(S_1^{\pm}|\underline{v}_2,\underline{v}_1,I) \tag{9}$$

$$= \frac{1}{p(S_1^{\pm}|\underline{v}_1,I)} \int\!\!\int d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ p(S_1^{\pm},S_2^{\pm}|\underline{v}_2,\underline{v}_1,\lambda_2,\lambda_1,I) \tag{10}$$

where, in passing from (9) to (10), we have again made the local/causal assumption that $p(S_1^{\pm}|\underline{v}_2,\underline{v}_1,I)$ is independent of $\underline{v}_2$: this follows by

marginalizing (3) over λ. Locality/causality demands further that

$$p(S_1^{\pm}, S_2^{\pm} | \underline{v}_2, \underline{v}_1, \lambda_2, \lambda_1, I) = F_{\pm}(\underline{v}_2, \lambda_2)\, F_{\pm}(\underline{v}_1, \lambda_1), \qquad (11)$$

for only then are the marginals for $S_1^{\pm}$ and $S_2^{\pm}$ dependent on just $(\underline{v}_1, \lambda_1)$ and $(\underline{v}_2, \lambda_2)$ respectively.

We now demonstrate explicitly that expression (10) is identical to the result derived from the proposed strategy of updating our knowledge of λ_2 using the result of S_1. This is in fact a consequence of the consistency conditions that probability not depend on how the conditioning data are partitioned, or on whatever has been marginalized out, from which Cox derived the two laws of probability. We seek

$$p(S_2^{\pm} | \underline{v}_2, S_1^{\pm}, \underline{v}_1, I) = \int d\lambda_2\, P(\lambda_2, S_2^{\pm} | \underline{v}_2, S_1^{\pm}, \underline{v}_1, I) \qquad (12)$$

$$= \int d\lambda_2\, p(S_2^{\pm} | \underline{v}_2, \lambda_2, S_1^{\pm}, \underline{v}_1, I)\, P(\lambda_2 | \underline{v}_2, S_1^{\pm}, \underline{v}_1, I) \qquad (13)$$

using only the laws of probability; and on demanding locality/causality

$$= \int d\lambda_2\, F_{\pm}(\underline{v}_2, \lambda_2)\, P(\lambda_2 | S_1^{\pm}, \underline{v}_1, I). \qquad (14)$$

Next, we work out $P(\lambda_2 | S_1^{\pm}, \underline{v}_1, I)$ as a marginal of the joint probability of λ_1 and λ_2:

$$P(\lambda_2 | S_1^{\pm}, \underline{v}_1, I) = \int d\lambda_1\, P(\lambda_1, \lambda_2 | S_1^{\pm}, \underline{v}_1, I) \qquad (15)$$

and retrodict the joint probability, incorporating the result of S_1 using Bayes' theorem:

$$P(\lambda_1, \lambda_2 | S_1^{\pm}, \underline{v}_1, I) = K\, P(\lambda_1, \lambda_2 | \underline{v}_1, I)\, p(S_1^{\pm} | \underline{v}_1, \lambda_1, \lambda_2, I), \qquad (16)$$

which on demanding locality/causality becomes

$$= K\, P(\lambda_1, \lambda_2 | I)\, F_{\pm}(\underline{v}_1, \lambda_1). \qquad (17)$$

Normalisation demands that

$$K^{-1} = \iint d\lambda_1 d\lambda_2\, P(\lambda_1, \lambda_2 | I)\, F_{\pm}(\underline{v}_1, \lambda_1) \qquad (18)$$

$$= p(S_1^{\pm} | \underline{v}_1, I). \qquad (19)$$

On substituting (19) into (17), the result into (15), and that into (14) we have

$$p(S_2^{\pm} | \underline{v}_2, S_1^{\pm}, \underline{v}_1, I) = \frac{1}{p(S_1^{\pm} | \underline{v}_1, I)} \iint d\lambda_1 d\lambda_2\, P(\lambda_1, \lambda_2 | I)\, F_{\pm}(\underline{v}_2, \lambda_2)\, F_{\pm}(\underline{v}_1, \lambda_1), \qquad (20)$$

which is just (10) with condition (11) already incorporated. This completes our equivalence proof.

The experimental result is

$$p(S_2^+|\underline{v}_2,S_1^+,\underline{v}_1,I) = p(S_2^-|\underline{v}_2,S_1^-,\underline{v}_1,I) = \tfrac{1}{2}(1-\underline{v}_1.\underline{v}_2), \quad (21a)$$

$$p(S_2^-|\underline{v}_2,S_1^+,\underline{v}_1,I) = p(S_2^+|\underline{v}_2,S_1^-,\underline{v}_1,I) = \tfrac{1}{2}(1+\underline{v}_1.\underline{v}_2) \quad (21b)$$

(Clauser and Shimony, 1982 (a review); Aspect et al (1982)). On substituting these into (20) and re-arranging, we have the four relations

$$\tfrac{1}{2}(1-\underline{v}_1.\underline{v}_2)\ p(S_1^+|\underline{v}_1,I) = \iint d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ F_+(\underline{v}_2,\lambda_2)\ F_+(\underline{v}_1,\lambda_1), \quad (22)$$

$$\tfrac{1}{2}(1-\underline{v}_1.\underline{v}_2)\ p(S_1^-|\underline{v}_1,I) = \iint d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ F_-(\underline{v}_2,\lambda_2)\ F_-(\underline{v}_1,\lambda_1), \quad (23)$$

$$\tfrac{1}{2}(1+\underline{v}_1.\underline{v}_2)\ p(S_1^+|\underline{v}_1,I) = \iint d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ F_-(\underline{v}_2,\lambda_2)\ F_+(\underline{v}_1,\lambda_1), \quad (24)$$

$$\tfrac{1}{2}(1+\underline{v}_1.\underline{v}_2)\ p(S_1^-|\underline{v}_1,I) = \iint d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ F_+(\underline{v}_2,\lambda_2)\ F_-(\underline{v}_1,\lambda_1). \quad (25)$$

Any of these implies determinism in the internal variable. For example, putting $\underline{v}_2 = \underline{v}_1 = \underline{v}$, (22) reduces to

$$\iint d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ F_+(\underline{v},\lambda_2)\ F_+(\underline{v},\lambda_1) = 0 \quad (26)$$

and since $P(\lambda_1,\lambda_2|I)$ is a measure of justified belief and is not physical, this relation must hold regardless of its form. Since also the physical quantity F_+ is non-negative, it follows that

$$F_+(\underline{v},\ \lambda_2=\lambda_2(\lambda_1))\ F_+(\underline{v},\lambda_1) = 0\ \forall\ \underline{v},\lambda_1, \quad (27)$$

where λ_2 is functionally related to λ_1 because of the correlation I. This in turn implies that one of $F_+(\underline{v},\ \lambda_2(\lambda_1))$ and $F_+(\underline{v},\lambda_1)$ is always zero. A similar argument applies to $F_+(\underline{v},\lambda_2)$ and $F_+(\underline{v},\ \lambda_1(\lambda_2))$; and to F_- rather than F_+. The result $F_+.\ F_- = 0$ or 1 at every direction and value of the internal variable, corresponding to determinism, now follows.

If the probability of spin measurements (3) depends on further physical quantities - such as time, in a dynamical hidden variable theory - $F_\pm$ would be a marginal distribution over an imperfectly known quantity, and would not be deterministic. Such theories are, therefore, untenable.

Since (21a) implies (21b) and vice-versa, two of equations (22)-(25) are redundant. Moreover, the sum of all four reduces to the normalisation condition (8) on $P(\lambda_1,\lambda_2|I)$. There is therefore only one independent relation, which we choose as (22) + (23) - (24) - (25):

$$-\ \underline{v}_1.\underline{v}_2 = \iint d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ A(\underline{v}_2,\lambda_2)\ A(\underline{v}_1,\lambda_1). \quad (28)$$

This is the expectation value of the product of the spins. Equations (8), (7) and (28) give the zeroth, first and second moments of

$P(\lambda_1,\lambda_2|I)$ with respect to A. In passing from (22)-(25) to (28) it was not necessary to use result (4), the measurement of S_1. Next, based on an inequality satisfied by the RHS of (28) but violated by the LHS, one deduces the master result that no (non-negative, normalised) solution exists for $P(\lambda_1,\lambda_2|I)$; this is a slight generalisation of Bell's original calculation (see Clauser and Shimony, 1978). It cannot therefore be a probability, and ergo not the probability of anything; existence of its arguments, the internal variables λ_1 and λ_2, is simply incompatible with the facts. To demonstrate this, first label the RHS as $E(\underline{v}_1,\underline{v}_2)$:

$$E(\underline{u},\underline{v}) \equiv \int\int d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ A(\underline{u},\lambda_1)\ A(\underline{v},\lambda_2). \tag{29}$$

Then

$$E(\underline{u},\underline{w}) - E(\underline{v},\underline{w}) = \int\int d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ [A(\underline{w},\lambda_2)A(\underline{u},\lambda_1)-A(\underline{w},\lambda_2)A(\underline{v},\lambda_1)]. \tag{30}$$

Now add and subtract a new term:

$$E(\underline{u},\underline{w})-E(\underline{v},\underline{w}) = \int\int d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ \{A(\underline{w},\lambda_2)A(\underline{u},\lambda_1)\}[1+A(\underline{v},\lambda_1)A(\underline{s},\lambda_2)]+$$
$$+\int\int d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ \{-A(\underline{w},\lambda_2)A(\underline{v},\lambda_1)\}[1+A(\underline{u},\lambda_1)A(\underline{s},\lambda_2)]. \tag{31}$$

Since $|A|\leqslant 1$, the square brackets are non-negative, and the magnitudes of the curly brackets are $\leqslant 1$. Thus

$$|E(\underline{u},\underline{w}) - E(\underline{v},\underline{w})| \leqslant \int\int d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ [1+A(\underline{v},\lambda_1)A(\underline{s},\lambda_2)]$$
$$+ \int\int d\lambda_1 d\lambda_2\ P(\lambda_1,\lambda_2|I)\ [1+A(\underline{u},\lambda_1)A(\underline{s},\lambda_2)], \tag{32}$$

or

$$|E(\underline{u},\underline{w}) - E(\underline{v},\underline{w})| \leqslant 2 + E(\underline{v},\underline{s}) + E(\underline{u},\underline{s}). \tag{33}$$

Symmetry of $P(\lambda_1,\lambda_2|I)$ has not in fact been employed. Next, a similar inequality is derived with $E \Rightarrow -E$ throughout. The more stringent of the two inequalities is always

$$|E(\underline{u},\underline{w}) - E(\underline{v},\underline{w})| + |E(\underline{u},\underline{s}) + E(\underline{v},\underline{s})| \leqslant 2. \tag{34}$$

Since equality is attained at $A(\underline{v},\lambda) = 1\ \forall\ \underline{v},\lambda$, this inequality is sharp over A. Further inequalities can be generated using the method of Braunstein and Caves (1988).

It is easy to find directions $\underline{s}$, $\underline{u}$, $\underline{v}$, $\underline{w}$, such that (34) is violated for $E(\underline{u},\underline{v}) = -\underline{u}.\underline{v}$; for example if $\underline{s}$ is parallel to $\underline{u}$ and perpendicular to $\underline{w}$, with $\underline{v}$ in the same plane at acute angles θ to $\underline{u}$ and $\pi/2-\theta$ to $\underline{w}$. Condition (34) then demands that $\sin\theta + \cos\theta < 1$, which is clearly false because $\sin^2\theta + \cos^2\theta = 1$.

A generalized proof goes through even if different functions $F(\underline{v},\lambda)$ are defined for each detector, to allow for differences between them. We simply attach subscripts $_1$ and $_2$ to the functions $A(\cdot,\lambda_1)$ and $A(\cdot,\lambda_2)$ in the foregoing.

The entire analysis is a train of inference, constrained only by the assumption of locality/causality (3); physics of the particles is irrelevant. Therefore the idea can be applied (and illustrated) in other areas. Suppose one isolates the individual members of a series of couples, and asks each person a yes/no question off a prescribed list. The analysis indicates the range of questions needed, and how to tell from the collected replies, whether individuals were in clandestine communication with their partners during the interrogation, as distinct from every couple sticking to its own pre-arranged story. The choice of question corresponds to the direction $\underline{v}$; memory to the internal variable, correlated through pre-arrangement in each couple; and communication corresponds to nonlocality. In this simplified problem we do not consider acausality.

Translating from particle experiments into this parlance, partners always give opposite replies if the same question is put to each. This is possible in isolation provided the couple pre-agreed answers to every question on the list; but then the replies taken over many couples and over many questions could not (in fact) have the completely random character observed. This insight into the analysis has been highlighted in a particularly clear model problem: Mermin (1985).

Another actualisation is: that it is impossible to program two independent computers C_1, C_2 so as to prompt successive users to input a direction $\underline{v}$, and to respond either "+" or"-", such that over many pairs of users

$$p(C_1^{\pm}|\underline{v}_1) = p(C_2^{\pm}|\underline{v}_2) = \tfrac{1}{2}, \tag{35}$$

$$p(C_2^{\pm}|\underline{v}_2, C_1^{\pm}, \underline{v}_1) = \begin{cases} \tfrac{1}{2}(1 - \underline{v}_1.\underline{v}_2) & ++,\ -- \\ \tfrac{1}{2}(1 + \underline{v}_1.\underline{v}_2) & +-,\ -+. \end{cases} \tag{36}$$

Here it is the program which corresponds to the internal variable.

Further insight into Bell's theorem is gained by looking at special cases. Suppose, for example, that the internal variable is a direction $\underline{\lambda}$, oriented in opposite directions for the members of a singlet pair, and our state of knowledge corresponds to uniform probability over solid angle. This is the usual "first try", corresponding to

$$P(\lambda_1,\lambda_2|I)\, d\lambda_1 d\lambda_2 = \frac{1}{4\pi}\, \delta^{(2)}(\underline{\lambda}_1+\underline{\lambda}_2)\, d^2\underline{\lambda}_1 d^2\underline{\lambda}_2, \tag{37}$$

where $\delta^{(2)}$ is the delta function over the surface of the unit sphere. $A(\underline{v},\lambda)$ must be a function, denoted g, of $\underline{v}.\underline{\lambda}$, and equation (26) becomes

$$\int d^2\underline{\lambda}_1 \; g(\underline{v}_1.\underline{\lambda}_1)\; g(-\underline{\lambda}_1.\underline{v}_2) = -\; \underline{v}_1.\underline{v}_2. \qquad (38)$$

Expansion of g in Legendre polynomials P_n leads to the unique solution $g(\underline{v}.\underline{\lambda}) = \sqrt{3}P_1(\underline{v}.\underline{\lambda}) = \sqrt{3}\underline{v}.\underline{\lambda}$. There is no zeroth Legendre polynomial, so this result also satisfies the "first moment" equation (7). But g exceeds unity in part of $(\underline{v},\lambda)$-space, contrary to (6).

Nonlocality/acausality are not observed in everyday life because we do not customarily observe individual particles travelling through a vacuum. Nevertheless the conclusion is firm: they are present. The novel concept of a nonlocal universe has led some physicists into mysticism, to their detriment, for physics is about prediction and its improvement. Shortly we shall exhibit a more fruitful alternative.

2. SIGNALLING?

Can nonlocality be exploited for long distance communication, or acausality for picking up signals from the future? By comparing the outputs of the two detectors after testing many particle pairs, we find that the character of the randomness in spin measurements on a particle is identical before and after its partner undergoes measurement. No matter how we choose the detector directions, nature arranges it so that any sufficiently long sample of output from either detector looks like any other. Therefore we cannot pick up any information from a set of particles about what is happening to their distant partners; signalling using this mechanism is impossible. Mathematically, from Bayes' theorem,

$$p(S_1 \text{ measured}|S_2^{\pm},I) = p(S_1 \text{ measured}|I)\, \frac{p(S_2^{\pm}|S_1 \text{ measured},I)}{p(S_2^{\pm}|I)}, \qquad (39)$$

and since, observationally,

$$p(S_2^{\pm}|S_1 \text{ was measured},I) = p(S_2^{\pm}|I), \qquad (40)$$

the posterior probability that S_1 was measured equals the prior; no information has been gained.

Bell's theorem proves that measurement of particles has the definite effect of altering the results that would otherwise have eventuated for their distant partners, increasing the correlation beyond what mere pre-arrangement could achieve. However, because this doesn't alter the degree of randomness at the second detector, we can only confirm it by conparing the results from <u>both</u> detectors, over many particle pairs. Signalling, by contrast, is a stronger form of nonlocality, different because of the random character of the observations. It is testable from the output of the second detector alone, and it is disconfirmed.

Some tests have been done using photons. If the influence travels from

the first detector at the speed of light or slower, it cannot catch up with the second particle and "prime" it. Does it therefore travel superluminally? (The two photons do not in fact propagate in exactly opposite directions, but the problem remains.) Also, because of relativistic space-time transformations, certain observers moving fast enough see the measurements take place in reverse order. Which particle tells which? Since no Lorentz frame is preferred, the resolution should be symmetrical with respect to the particles, seemingly implying acausality. The answers must lie with the theory describing the particles: quantum mechanics.

Finally, if hidden variables (necessarily nonlocal/acausal) are ever uncovered, tangible signals could be received before they had been sent, and at speeds faster than light. Doubtless this is why we have never seen hidden variables; it also hints that we never shall, and that we are stuck with quantum randomness. Entirely identical systems do behave differently.

3. QUANTUM MECHANICS

Although we compared the predictions of local/causal theories with experiment, we know that quantum mechanics correctly predicts the results (see, for example, Clauser and Shimony, 1978). It is therefore a nonlocal/acausal theory. This feature was not built in explicitly; indeed it emerges unexpectedly. To ask where inside quantum mechanics it comes from is simply not fruitful; better to accept it and go on from there. We recall Einstein's bold re-orientation of the constancy of light-speed as a starting point rather than something to be explained; and what it led to. In quantum mechanics too, reorientation achieves a dramatic breakthrough; but first let us tidy up some loose ends.

It is remarkable that, nature being nonlocal/acausal, we can make prediction at all. There is a further, distinct, locality problem: when working with two fermions (bosons), how can we get away without using a monstrous Slater (anti)determinant for all the other identical particles in the universe? The explanation is that the relative phases of the others are unknown, and marginalizing ("averaging") over them recovers the usual results.

Second, what is meant by stating that quantum theory accords with experiment? Quantum mechanics predicts, for example, that

$$p(S_1^{\pm}|\Psi(1,2)) = \tfrac{1}{2} \tag{41}$$

where Ψ represents the wavefunction. Experiment begins by noting that there are two alternatives S^+ and S^-. This is the "coin-tossing" problem, solved by Jaynes (1968): given information Σ that we know nothing in advance distinguishing S^+ from S^-, we assign for the probability of n_+ measurements of S^+ and $n_-=N-n_+$ measurements of S^-

$$p(n_+|N,\Sigma) = \int_0^1 dq\ F(q)\ \frac{N!}{n_+!n_-!}\ q^{n_+}\ (1-q)^{n_-} \qquad (42)$$

where the "prior" $F(q) \propto q^{-1}(1-q)^{-1}$ and the constant of proportionality is determined by demanding that (42) be normalised over n_+, $0 \leqslant n_+ \leqslant N$. Then $p(\pm\text{next}|n_+,n_-,\Sigma)$ is easily calculated via Bayes' theorem. Experimental results converge to ½, in agreement with the theoretical prediction (41).

This procedure is fully Bayesian, and avoids the "frequentist" view that repeated results constitute an ensemble of the distribution (41). Instead, $p(S^{\pm}|\text{theory})$ and $p(S^{\pm}|\text{experiment})$ are compared.

Finally, measurement is not a well-defined act in (quantum) reality; it merely reflects an interaction of some form. Needles on dials are themselves quantum objects, evaporating when examined closely into a blur of elementary particles. This Gordian knot is severed by <u>assuming</u> there is a well-defined answer, to be determined as best we can by looking at the pointer closely, but not <u>too</u> closely. Quantum mechanics predicts the result statistically. Since this procedure works, measurement remains a legitimate concept.

4. THE TRANSACTIONAL INTERPRETATION OF QUANTUM MECHANICS

The predictive formalism of quantum mechanics works as perfectly in the Bell experiments as everywhere else. But quantum theory still gives cause for unease. What is needed is a fresh <u>interpretation</u> which takes our new understanding into account. A nonlocal/acausal interpretation of the Schrodinger representation has recently been proposed by J. G. Cramer (1986), the seeds of which go back as far as Tetrode (1922). What this looks like in the Heisenberg representation, closest to our classical way of thinking, remains to be seen; but a similar concept, two-point boundary conditions in time corresponding to the past and future, has been proposed for the Feynman path integral representation (Roberts 1978). Cramer's interpretation is by far the most helpful way of thinking about quantum phenomena yet found, and it hints at future revisions of the physics. Certainly it consigns the Copenhagen interpretation - "don't try to think, the formalism's the thing" - to history.

Cramer's idea is this: the wavefunction Ψ satisfies the Schrodinger equation

$$i\hbar\ \frac{\partial\Psi}{\partial t} = H\Psi \qquad (43)$$

(H is the Hamiltonian) and propagates forward in time (+t). Since H is Hermitian, the conjugate wavefunction satisfies

$$i\hbar \frac{\partial \Psi^*}{\partial(-t)} = H\Psi^* \qquad (44)$$

and propagates backwards in time (-t). Interactions are viewed as transactions between transmitter and receiver; the transmitter sends an offer wave Ψ forward in time to the receiver, which itself sends a confirm wave backwards in time to the transmitter. Either process is stimulated linearly by the other, and consequently the amplitude (probability) for the overall interaction is proportional to the product $\Psi^*\Psi$, as required. The mysterious quadratic form is explained. An operator Ξ measured at the receiver projects out of the offer wave one eigenfunction, giving the physical expectation value $\langle\Psi|\Xi|\Psi\rangle$. In Einstein's phrase, it is the offer wave which is "there when nobody looks".

The immediate objection is that acausal signalling could take place via the advanced wave. That this is untrue is demonstrated by adapting the Wheeler-Feynman electrodynamic theory of retarded and advanced potentials. Wheeler and Feynman (1945, 1949) were concerned about the ad hoc manner in which advanced Lienard-Weichert potentials - which are perfectly good solutions of Maxwell's equations - are customarily eliminated "by causality". They reformulated the problem such that a charged particle emits both retarded and advanced waves in a time-symmetric combination; the advanced wave it emits is cancelled by the effects of receiving other advanced waves from the future. The advantage of this procedure is in replacing the ad hoc elimination of advanced waves by boundary conditions in the distant past and future. The disadvantage is that, in order to predict the evolution of a system, an integral must be taken over the whole future light cone - details of which are unknown. But if we are concerned exclusively with interpretation, as we are in (43) and (44), this disadvantage evaporates leaving only the benefit.

For full details, the reader is referred to Cramer's own exposition. This includes such points as the parabolic nature of the Schrodinger equation (43) in contrast with the hyperbolic relativistic equation it approximates (the factor i in (43) keeps the dynamics reversible), and reality of the overall wavefunction at transmitter and receiver but not in between. It also applies the transactional idea to the entire gamut of quantum paradoxes: measurement, Schrodinger's cat, Wigner's friend (the infinite regress of nested observers) and others, resolving them all objectively and realistically. The Bell experiment is included - since the particles continually exchange waves, there is no longer any problem about which affects which - but here we illustrate the idea with a simpler problem: the "delayed choice" paradox.

Suppose we fire a single particle at a pair of Young's slits, and only decide after it has passed them whether to measure the interference pattern (due to both slits) or the position of the particle (indicating which one slit was traversed). How can our choice of what is placed beyond the slits - photographic emulsion for recording fringes, or

collimators to detect particles - influence whether the particle traverses one slit or both, when it has already passed them? This puzzle has the authentic quantum flavour.

To fill in the details: first, it is known that a single particle can interfere with itself; second, we can be as certain as we like when the particle traverses the slits, by measuring its "perpendicular" velocity to arbitrary accuracy. (The uncertainty principle provides no excuse, for the experiment can be repeated many times, with the problem growing ever more acute.) Third, if we choose to measure position, we are not stating in advance which slit was traversed, only that one was. Finally, the paradox is a consequence of the non-commutativity of the two operators - transverse position and momentum - which are alternatives for measurement.

The transactional resolution is that the source of the particle emits offer waves forward in time, which pass unhindered through both slits. Depending on the experiment selected, confirm waves from the future come back through one slit or both. What is observed is the "handshake" between the offer and confirm waves, and this incorporates the chosen measurement while avoiding the paradox. Most other quantum riddles yield just as easily.

One puzzling nonlocal effect, mentioned here for completeness, has a further interpretation. Aharonov and Bohm (1959) predicted quantum-theoretically that the magnetic field inside a region from which a charged particle is excluded nevertheless influences the particle's motion. Experiment confirms this (Chambers 1960). The explanation is due to Peshkin (1981): although the particle is excluded, its electric field still penetrates the region, and the crossed electromagnetic fields there have angular momentum. Quantisation of (total) angular momentum couples this to the motional angular momentum of the particle. No metaphysical discussion of whether the magnetic vector potential is "more real" than the magnetic field is necessary.

Although these ideas are not testable other than at $t = \pm\infty$ - cosmologically - they do suggest where to look for the next generation of physical theories. Despite the accuracy of some quantum predictions to 1 part in 10^7, nature is not in general linear; linearity is usually only a convenient mathematical approximation, and a linear theory of excitation could very well turn out to be a weak-Ψ approximation. Also, the offer and confirm waves Ψ and Ψ^* may prove to be physically measurable quantities, in defiance of causal signalling and gauge invariance.

5. CONCLUSION

Bell's theorem is an analysis in Bayesian inference, incorporating physics only through the assumption of locality/causality. Tested

against experiment, this assumption fails. The theory which correctly predicts the experimental outcome, quantum mechanics, is therefore nonlocal/acausal. A new "transactional" interpretation of quantum mechanics has been built on this observation, which resolves traditional quantum paradoxes; the statistical nature of quantum processes is necessary to preclude acausal signalling. The transactional interpretation suggests where quantum mechanics should be probed for possible breakdowns.

REFERENCES

Aharonov, Y. & Bohm, D. 1959. Phys. Rev. 115, 485.

Aspect, A., Dalibard, J. & Roger, G. 1982. Phys. Rev. Lett. 49, 1804.

Bell, J.S. 1964. Physics 1, 195.

Bohm, D. 1951. Quantum Theory. Prentice-Hall (Englewood Cliffs, New Jersey, USA). §22.16.

Braunstein, S.L. & Caves, C.M. 1988. Phys. Rev. Lett. 61, 662.

Chambers, R.G. 1960. Phys. Rev. Lett. 5, 3.

Clauser, J.F. & Shimony, A. 1978. Rep. Prog. Phys. 41, 1881.

Cox, R.T. 1946. Am. J. Phys. 14, 1.

Cramer, J.G. 1986. Rev. Mod. Phys. 58, 647.

Einstein, A., Podolsky, B. & Rosen, N. 1935. Phys. Rev. 47, 777.

Jaynes, E.T. 1968. IEEE Transactions on Systems Science and Cybernetics SSC-4, p237. Reprinted as Chapter 7 of: E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics, ed: R.D. Rosencrantz. Synthese series 158. Reidel (Dordrecht) 1983.

Mermin, N.D. April 1985. Phys. Today 38, 38.

Pearle, P.M. 1970. Phys. Rev. D 2, 1418.

Peshkin, M. 1981. Physics Reports 80, no 6.

Roberts, K.V. 1978. Proc. Roy. Soc. Lond. A 360, 135.

Tetrode, H. 1922. Zeit. fur Physik 10, 317.

Wheeler, J.A. & Feynman, R.P. 1945. Rev. Mod. Phys. 17, 157.

Wheeler, J.A. & Feynman, R.P. 1949. Rev. Mod. Phys. 21, 425.

Probability, Philosophy and Science: a briefing for Bayesians

A.J.M. Garrett
Department of Physics & Astronomy
University of Glasgow
GLASGOW G12 8QQ
Scotland, U. K.

Abstract

The objective Bayesian view is considered in relation to philosophy and philosophy of science. Carnap's distinction between logical and factual probabilities is rejected, as is an anti-Bayesian argument due to Popper. Reasons for the confusion are advanced. Inductive philosophy of science is defended, and a tentative methodology proposed. Four prominent anti-inductivists are analysed: Popper, Lakatos, Kuhn and Feyerabend. Popper is shown to have been misled by the problem of improper priors, in hypothesis space; his deductive doctrine of falsifiability is replaced by the inductive one of testability. Kuhn's view that successive theories do not approach any kind of limit is criticised, and is traced to his rejection of induction. It is concluded that deductive methodologies of science are untenable, and that inductive methodology is sound.

1. Introduction

This paper divides into two parts: a survey of how our philosophical colleagues view probability, and a critique of the prominent deductivist philosophies of science. Scientists concerned with probability often do not realise that there has long existed a parallel effort in philosophy. The two are related through the connection - indeed identity - of probability theory with inductive logic, a connection denied by one major school of thought. In both parts we shall encounter a major figure of the 20th century: Karl Popper (1902-).

2. Philosophers' views of probability

I shall present this section from the objective Bayesian point of view. I shall not attempt to defend it here since that has been masterfully done by others (Jaynes: collected papers [1]). The objective Bayesian view is that the probability of an occurrence, conditional on information in a given space, measures how likely one believes that occurrence to be, and that the laws of probability are laws of inference. This view is objective to the extent that definite information corresponds to a definite probability. Anybody having the same information but assigning a different probability is therefore guilty of reasoning inconsistently. It is anthropomorphic to the extent that different individuals often possess different information, and therefore (consistently) assign different probabilities to the same event. That the objective Bayesian view corresponds to the familiar sum and product rules ("Kolmogorov axioms") was

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 107–116.

demonstrated by R. T. Cox [2]. Alternative schemes may coincide with the objective Bayesian in particular circumstances, for example the frequentist view given an infinite number of trials. These alternatives are then quite acceptable; but only the objective Bayesian view works irrespective of context. (Have you ever actually seen an infinite number of trials?)

Philosophers do not suffer as badly as some from Frequentist's Disease. The reason is historical: philosophers have for centuries been concerned with that logic to be used when there is insufficient information for certainty, called inductive logic. They were discussing the problem of induction long before probability even became quantitative, and longer still before frequentists hijacked it. Consequently philosophy never completely fell for this aberration. Of course, the frequentist view did influence philosophers, and in fact Popper initially advocated it, in Logik der Forschung (translated as The Logic of Scientific Discovery [3]). He was concerned to oppose non-objectivity, but failed to distinguish properly between the objective Bayesian view and the obviously crazy "subjective" one that anyone may assign any probability to anything. (True; but doing it consistently is another matter.)

Today, thanks to Cox, we know that inductive logic is probability theory, and vice versa. It is a generalization of the deductive logic of certainty, Boolean algebra, from values 0 and 1 to the interval in between. This view is inexorably gaining acceptance among physicists. In philosophy there is a diversity of positions; one hugely influential stance is due to Rudolf Carnap (1891-1970). Carnap distinguishes two kinds of probability, which he calls logical and factual [4].

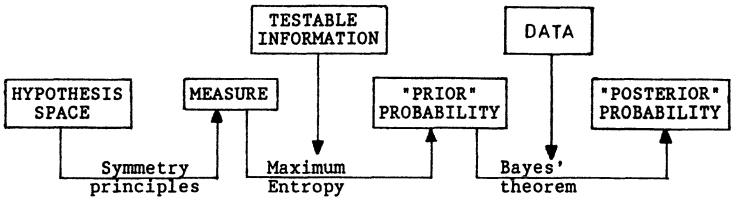


Figure 1: Bayesian Probability Assignments

Let us examine Carnap's claim from an objective Bayesian perspective. The Bayesian procedure for making inference is set out in Figure 1. We begin at the level of the hypothesis space, {x}. The first stage is to determine the measure m(x) on it, using symmetry arguments such as invariance under exchange of two elements, or transformation group theory [5]. Next, testable information is incorporated. These are statements like "the mean has value 2.7" which can be checked, in contrast to data which are the outcome of measurements. (If the mean is to be estimated, a symbol stands for its value at this point.) The probability (density) p(x) is determined by maximising the information entropy

$$- \int dx\ p(x)\ \log\ (p(x)/m(x)) \qquad (1)$$

subject to the constraints of normalisation and any testable information [1]. If there is none, $p(x) \propto m(x)$ without further ado and the entropy maximisation is customarily omitted. Finally, data are brought in to update the "prior" probability using Bayes' theorem. Further data can be incorporated at any time,

and the end result is independent of the order of incorporation. This is in fact one of the consistency conditions from which Cox derived the laws of probability in the Bayesian view [2]. The other is that probabilities shall not depend on what has been marginalized out.

Carnap's logical probability corresponds to the probability before data are incorporated, for this is assigned using logical arguments: symmetry and maximum entropy. But the rationale is not given with anything like the same clarity, needed to tackle real problems. "Factual" probability corresponds to probability assignments after data (facts) are incorporated. Thus, although there is a distinction in the generation of these probabilities, they both have fully Bayesian interpretations as the only ones which can be consistently assigned from the information at hand, be it data or testable.

Philosophers often cite examples like

$$p(\text{John is left-handed} \mid \text{John is a banker and 15\% of bankers are left-handed}) = 0.15 \qquad (2)$$

as logical assignments of probability. Whatever, this is an unfortunate example, because the information is given in the "handedness-space" of all bankers, not the required space of John's handedness. While Bayesians certainly emerge with the answer 0.15, the problem contains extra complications.

Philosophers who have criticised Carnap's position include W. V. O. Quine (1908-) and Popper. Indeed, entire books have been written on the Popper-Carnap controversy [6].

Popper has in addition presented a technical anti-Bayesian argument [7]. Though erroneous, it is still doing damage today, and a refutation is in order. The reader is warned that philosophers often use a notation in which the conditioning information follows after a comma. Thus, what scientists call $p(A|B)$ is written by Popper as $p(A,B)$ - which unfortunately has the distinct meaning to scientists "the joint probability of A and B". Here we use "scientific" notation.

Popper's argument, paraphrased, runs like this: suppose proposition B supports proposition A, given information I. This is held to correspond to the inequality

$$p(A|BI) > p(A|I). \qquad (3)$$

Since $p(A|I) = p(AB|I) + p(A\bar{B}|I)$, marginalizing over B (from the sum and product rules), and since $p(AB|I)$ can be decomposed using the product rule, further inequalities can be derived.

Popper now supposes that the statement "B supports A to the degree z" corresponds to the Bayesian assignment

$$p(A|BI) = z. \qquad (4)$$

It is then easy to conjure up a contradiction of the type "I am likely to drink tea. I an unlikely to drink coffee." (Both type (4).) "But, given a choice, I am more likely to drink coffee than tea." (Type (3).) Popper makes the point with dice.

Popper concludes that the Bayesian view is inconsistent. But this is semantic confusion: the concept of support is different in (3) and (4). Suppose

(3) is taken as the definition of what it means for B to support A; this accords well with intuition. That done, the word means something else in (4), for no inequality is at hand.

Let us illuminate this by showing that "supports" in (4) may not coincide with intuition. Suppose that A, B and I are such that

$$0.99 = p(A|BI) < p(A|I). \tag{5}$$

There is no problem in arranging this: let, for example
A = "there will be a traffic jam in central London today",
B = "there are no road works in central London at present", and
I = "it is a working day". Then B is antagonistic to A; but according to Popper's qualitative statement of (5), it supports it to a degree of 0.99, i.e. very strongly. The lesson is that assignment of probability is distinct from comparisons of probabilities. Only probabilities, not their differences (or differences of their logarithms) satisfy Cox's axioms.

In summary, I believe the confusion prevailing over probability in philosophy is due to two factors. First, the philosopher's disposition is to ask "What is probability?", while the scientist seeks solutions to specific physical problems, asking instead "How can probability help me?" The general is always illustrated by the specific.

Second, it could be that probability has had its day in philosophy. Philosophy bore the torch of Western learning and enquiry for centuries; but, as more became known, specialised areas of knowledge branched off from it. Science itself is the outstanding example; until relatively recently physics was known as Natural Philosophy. But with the underpinning of Cox, recognition of the dominant role of the Principle of Maximum Entropy, and the beginnings (in quantum statistical mechanics) of an operator-valued theory of probability, the day of the amateur - in the best sense, for philosophers are often eminent in several branches of their discipline - may be at an end.

3. Modern philosophy of science

This section critically surveys much 20th century philosophy's view of scientific methodology. For the basis of this material I am indebted to my former colleague at the University of Sydney, David Stove, now retired from its Department of Traditional and Modern Philosophy. David holds to the relevance of Carnap's distinction, but he is a defender of induction, and a formidable critic of deductivist philosophy of science.

Philosophy of science is best described as that which scientific endeavours have in common, but which non-scientific studies do not necessarily share. We should not suppose, though, that there is anything magic about science: it is simply a sustained application of common sense. And since common sense is consistent reasoning (in an appropriately chosen space), we find ourselves staring straight at Bayesian inductive probability.

Bayes notwithstanding, modern philosophy of science has grown into a major pathology associated with the names Popper, Lakatos, Kuhn and Feyerabend. Describing it is my aim here, and as a preliminary arming I present a tentative flowchart of how science is done.

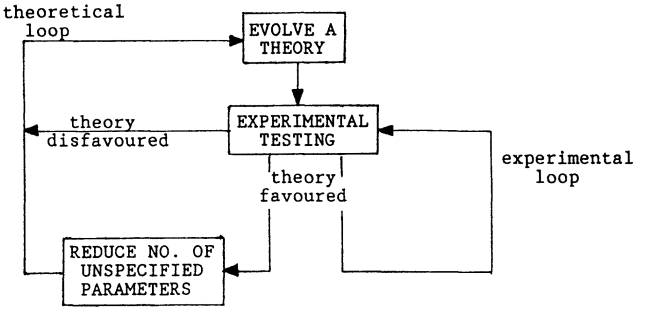


Figure 2: A Broad Methodology For Science

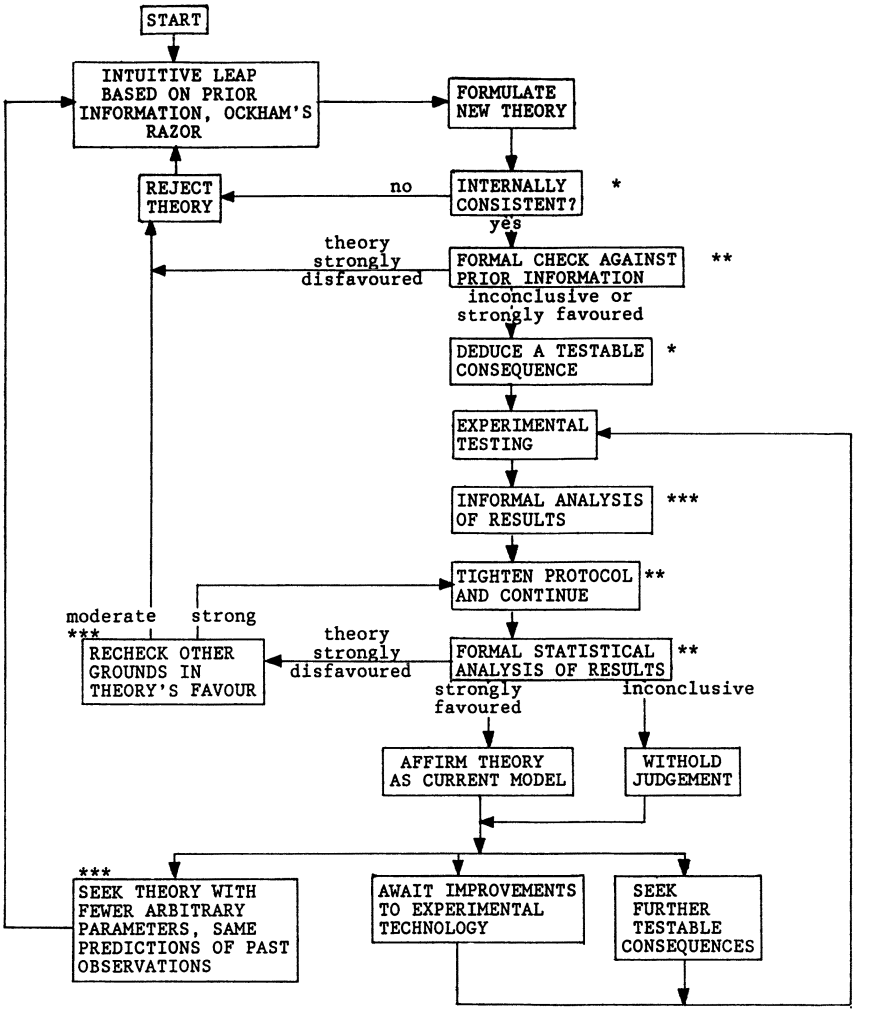
In its basic form, the model is given in Figure 2. There is no endpoint, and so no "final answer". Instead, one continually refines theory and practice. The vexed debate over realism is circumvented by defining scientific truth as the asymptote towards which this process (in practice) converges. Laws of Nature are unknown and never change (assuming, reasonably, that they exist), but our approximations to them improve as we learn more.

Since the loops always contain at least one inductive step, the whole process is inductive. This is no more than it should be: one can be almost certain that the Sun will continue to rise in the east, based on past observations (and the celestial mechanics constructed to explain them), but certainty is absent.

The model can be fleshed out to varying degrees, and that which I have found most illustrative in physical science is displayed in Figure 3. It is here that philosophy can help scientists, although they generally pursue such a strategy implicitly.

A new intuitive leap can throw up a theory at any time. The resulting flowchart is welded to the old by applying the process to the union of the two theories. Unifying demonstrates directly that science is not dogmatically reductionistic: reductionism is simply a convenient way of implementing the strategy of Figure 3. The intuitive leap corresponds to a widening of the region of hypothesis space under consideration, a process for which there is as yet no theory, even in model problems. Meanwhile, the hotchpotch of guess, conjecture and imagination called intuition is precisely what distinguishes great scientists from the rest.

The inductive view of science goes at least as far back as the 12th century scholar Roger Bacon, and thence forward to Elizabeth I's courtier Francis Bacon. (Of course, William of Ockham's famous razor principle "Essentia non sunt multiplicanda praeter necessitatem" - entities should not be multiplied beyond necessity - is essentially Bayesian.) The distinguished British empirical philosopher David Hume (1711-1776) argued against induction, and Stove traces today's movement to this source [8].



*deductive process
**well-defined inductive process
***weakly defined inductive process

Figure 3: A Detailed Methodology For Physical Science

More recently, Pierre Duhem (1861-1916) argued that theory and experiment never meet face-to-face, because in real science a prohibitive number of auxiliary assumptions are involved in reaching the interface [9]. Today this is called the Quine-Duhem thesis. On the inductive picture, extra assumptions are readily incorporated by setting up the prior distribution for their parameters, calculating the joint posterior distribution of these and the desired quantities from the data by using Bayes' theorem, and then marginalizing over the extra parameters to take them out.

Karl Popper opened the modern era with Logik der Forschung in 1934. It is far more about philosophy of science than probability theory. Curiously though, Popper is regarded more by his profession, at least outside England, as a celebrity than a philosopher's philosopher. Summarising him is not easy: as with most cults, many meanings can be read into it. This is due to such contradictions as his acceptance of probability but rejection of induction; David Stove, in the first part of Popper And After: Four Modern Irrationalists [8] exposes the devices by which Popper (among others) lays his smokescreen.

Popper's primary tenets can nevertheless be discerned. One is that all observations are "theory-laden". In Popper's own words, "sense-data, untheoretical items of observation, simply do not exist" [10]. It is difficult for Bayesians to express the depth of their disagreement with this. Data are data, be they a distraction (noise) or tracks in bubble chambers photographed at enormous ingenuity and cost. Whatever, they are incorporated into theory using Bayes' theorem. Of course, theories suggest which data to seek, but that is not at all the same thing; once found, data can be used to update the probability of any hypothesis whatsoever.

Popper also insists that science is deductive rather than inductive. Partly this is a terminological disparity, referring not to the overall process but to a single stage: deduction of the consequences of a hypothesis prior to testing. (Popper's scheme is often described as hypothetico-deductive.) But Popper does reject induction; we have seen already his rejection of the inductive view of probability in favour of other interpretations. Indeed, Popper has asserted that no theory ever becomes more probable when evidence in its favour is discovered, and that every scientific theory not only begins by being infinitely improbable, but always remains so [11].

The first of these statements directly denies Bayes' theorem. Underlying the second is the idea, seldom recognised, of the space in which probabilities are defined. This contains an infinity of competing theories, and before looking at their distinctive features we must assign each equal prior probability $1/\infty$, or zero. It seems that Popper is correct. But Bayesians recognise this as the problem of non-normalisable or improper priors, in hypothesis space. The resolution is the same: though the prior is non-normalisable, Bayes' theorem gives for the posterior ratio a well-defined limit of $0/0$ which may perfectly well be normalisable [5]. Bayesians can open The Logic of Scientific Discovery, and the first volume of its massive Postscript [12], expounding Popper's post-frequentist "propensity" view of probability (much closer to Bayesian, though never made plain), almost at random and illuminate the problems exposed.

The idea by which Popper is best known, and one of which most students are aware, is the doctrine of falsifiability. A hypothesis is only scientific if it is capable of being proved false by observation. This is an important idea, but it is baldly deductivist in restricting the concept to falsity but not truth; for in deductive logic a single counter-example can falsify a theory but no number of examples can prove it. In real science though, theories are not proved false (or true) with certainty. Instead, data are incorporated via Bayes' theorem into the posterior probability, which may approach zero or one. As it gets sufficiently close (a matter of taste), the theory is rejected or adopted. So the criterion is not falsifiability, but testability: that one can conceive of data which alter the probability of the hypothesis. Equivalently, the hypothesis must not be equally disposed to every datum. Stove traces

Popper's dictum of falsifiability to his early distaste, in Vienna, with dogmatic claims that ideas such as Marx's and Freud's were "irrefutable" [11,13]. The word game began here, though Marx at least understood the stakes long before. He wrote to Engels in 1857, concerning a historical event: "One can always get out of [making an ass of oneself] with a little dialectic. I have, of course, so worded my proposition as to be right either way" [14].

Popper does refer to testability although, deprived of inductive logic, he fails to nail down the idea. (Any statement pertaining to a theory, but which is not testable, is part of that theory's interpretation.) He also refers to degrees of falsifiability, and attempts to relate them to probability. This again is word-play: the ease with which a hypothesis is tested is a matter for technologists, not theoreticians.

It was the successor to Popper's London Chair, Imre Lakatos (1922-1974), who pointed out afresh that in real science, theories are never disproved (or proved) with certainty. Lakatos clearly reached this conclusion through contrasting the natural sciences with mathematics, in whose history he had worked [15]. But, following Popper in renouncing induction, Lakatos was left with no framework to hang his observation on. His own attempt to build one, a doctrine of research programmes [16], confuses philosophy with history of science.

Like many ideas given birth in eastern Europe, deductivism has become popular in America. Let us therefore examine the work of Thomas Kuhn (1922-), another avowed anti-inductivist and the author of the hugely influential work The Structure of Scientific Revolutions [17]. Like Lakatos, Kuhn is a first-rate historian of science, who has written on the Copernican revolution in astronomy, and the black-body controversy which gave birth to the earliest quantum hypothesis. The Structure of Scientific Revolutions presents a cyclic view of how science evolves, beginning with a mass of unordered observations and competing theories, going into a quiescent stage after the triumph of one theory over the rest, followed by gradual breakdown into chaos again under the accumulation of anomalies from more stringent testing. Kuhn has bequeathed us one of today's fashionable words, paradigm, to describe the model prevailing during the quiescent stage.

The history of science abounds with examples of this process; trouble again arises when it is combined with anti-inductivism as a philosophy. Close to the end of The Structure of Scientific Revolutions, Kuhn clearly echoes Popper's assertion that every theory is infinitely improbable, when he says that "we may.....have to relinquish the notion that changes of paradigm carry scientists....closer and closer to the truth" [18]. In other words, Kuhn believes that theories come and go as arbitrarily as fashions in clothing. For anti-inductivists, the closer fit to observation of relativistic mechanics than Newton's counts for nothing. This singular ideology is deflated by applying it to progressively simpler problems: it can hardly be no more true that the Moon is made of rock than green cheese.

There is no doubting, though, that Popper, Lakatos and Kuhn all appear respectful to science. Consider finally the ideas of Paul Feyerabend (1924-) [19,20]. Feyerabend describes his approach as "epistemological anarchism", and his slogan is "Anything Goes". Again, this derives clearly from the idea that all theories are equally invalid, though it is also an accurate distillation of the subjective Bayesian stance. And Popper's oldest criticism holds of it: it is not falsifiable!

Feyerabend is on record as stating that normal science is a fairytale, and that equal time should be given to "astrology, acupuncture and witchcraft" [21] (though I do not know what unctions he seeks when ill). He is fond of categorising science with "religion, prostitution and so on" [19]. Feyerabend believes that science is just one of many internally consistent views of the world, and that the consequent choice between them should be made on social grounds. But while many systems are internally consistent, only one plugs consistently into the world of observations, and to reason systematically about that world we must use it: science. Ethical and social considerations may dictate which areas to study, but that is a different matter. Feyerabend's ideas have been brought forth by "the sleep of reason", and while they could probably only flourish in a society disillusioned with science (through its perceived misuse), they represent the logical culmination of the rejection of induction. For that is what Popper, Lakatos, Kuhn and Feyerabend have in common; and despite much mutual repudiation, it far outweighs their differences. The ideas of these four comprise a major stream in contemporary philosophy of science.

It is an odd fact that scientists often quote these philosophers favourably [22]. Science magazine recently lauded Feyerabend's views as "a breath of fresh air" [21]. The explanation is undoubtedly a benign ignorance. Being prepared to revise hypotheses in the light of fresh information (an attitude politicians might heed) makes scientists easy prey. What, by contrast, could alter Feyerabend's opinion?

4. Conclusion

The objective Bayesian view is as capable of resolving problems concerning inductive logic in philosophy as it is in science. Difficulties are not conceptual, but merely technical: the huge spaces used in real problems, and the determination of prior probabilities in a wide variety of contexts [1]. In particular, the Bayesian view, applied to scientific methodology, produces a coherent, inductive philosophy of science. Non-inductive philosophies of science invariably lead to absurdities.

References and Notes

[1] E.T. Jaynes. 1983. Papers on Probability, Statistics and Statistical Physics, ed: R. D. Rosencrantz. Synthese series vol 158. Reidel (Dordrecht).

[2] R.T. Cox. 1946. Am. J. Phys. 14, 1.

[3] K.R. Popper. 1959. The Logic of Scientific Discovery. Hutchinson (London). Translation of: Logik der Forschung (Springer, Vienna, 1934).

[4] R. Carnap. 1950. Logical Foundations of Probability. University of Chicago Press.

[5] E.T. Jaynes. 1968. IEEE Transactions in Systems Science and Cybernetics, SSC-4, p227. Reprinted as Chapter 7 of [1].

[6] A.C. Michalos. 1971. The Popper-Carnap Controversy. Martinus Nijhoff (The Hague).

[7] K.R. Popper. 1954. Brit. J. Philos. Sci. 5, 143. Reprinted in reference [3], revised edition 1980, Appendix ix.

[8] D.C. Stove. 1982. Popper And After: Four Modern Irrationalists. Pergamon

(Oxford).
[9] P. Duhem. 1954. The Aim and Structure of Physical Theory. Second edition, Princeton University Press. (Translation of second French edition, 1914.)
[10] K.R. Popper. 1968. In: Problems in the Philosophy of Science, eds: I. Lakatos & A. Musgrave, p163. North-Holland (Amsterdam).
[11] see: D.C. Stove. June 1985. Encounter, p65-74.
[12] K.R. Popper. 1982-3. Postscript to The Logic of Scientific Discovery, Volume I: Realism and the Aim of Science (1983), Volume II: The Open Universe: An Argument for Indeterminism (1982), Volume III: Quantum Theory and the Schism in Phyics (1982). Ed: W.W. Bartley III. Hutchinson (London). Note the title of volume III; the schism occasioned by quantum theory is between past and present, not in physics as Popper asserts.
[13] D.C. Stove. 1982. 'How Popper's Philosophy Began'. Philosophy, 57, 381. The source is an autobiographical detail by Popper in the summary paper 'Science: Conjectures and Refutations', in: Conjectures and Refutations, publisher: Routledge & Kegan Paul (London) 1963.
[14] K. Marx. 1983. Collected Works, 40, 152. Lawrence & Wishart (London).
[15] I. Lakatos. 1963-4. Various papers, collected as: Proofs and Refutations: The Logic of Mathematical Discovery. Eds: J. Worrall & E. Zaher, Cambridge University Press, 1976.
[16] I. Lakatos. 1978. Philosophical Papers, Vol I: The Methodology of Scientific Research Programmes. Eds: J. Worrall & G. Currie. Cambridge University Press.
[17] T.S. Kuhn. 1962. The Structure of Scientific Revolutions. University of Chicago Press. (Second Edition, enlarged, 1970.)
[18] Reference [17], second edition, p170.
[19] P.K. Feyerabend. 1975. Against Method: Outline of an Anarchistic Theory of Knowledge. New Left Books (London).
[20] P.K. Feyerabend. 1987. Farewell To Reason. Verso (London).
[21] P.K. Feyerabend. 1979. Quoted in: Science, 206, 534.
[22] T. Theocharis & M. Psimopoulos. 1987. Nature, 329, 595, and 331, 384 (1988).

The Statistics of Quantum Mechanical Wavefunctions

R. D. Levine
The Fritz Haber Research Center
for Molecular Dynamics
The Hebrew University
Jerusalem 91904, Israel
and
Department of Chemistry
Harvard University
Cambridge, MA 02138, USA

ABSTRACT. The maximum entropy formalism is used to obtain the distribution of amplitudes of a single quantum state. Such a distribution is required to account for the observed irregular but reproducible spectra at high levels of excitation. The computed distribution agrees well with experimentally determined histograms. The reasons for possible deviations are noted. Special attention is given to the conceptual foundations of the approach and analogies are drawn with classical statistical mechanics. A distinction between the objective and subjective elements in quantum mechanics is made . In particular it is proposed that the amplitudes are objective while their distribution reflects a state of knowledge.

1. Introduction

The technical problem which we address is the nature of the spectrum of highly excited states of systems with few degrees of freedom (e.g., molecules or nuclei). Such a spectrum is typically quite dense with many transitions whose intensities vary in a seemingly erratic manner as we scan the frequency. Often, the density of transitions is comparable to the density of states which implies the near absence of selection roles (which are typical of low excitation spectra). Traditional methods of spectroscopy, which seek to assign each and every transition are thus of limited use.

The problem is not one of data analysis. I shall assume that the spectrum has been measured with low signal to noise. Hence I shall take the reported experimental results as the true, inherent, spectrum of the system. The 'statistics' are therefore not due to external noise. Rather, they are considered as a signature of the 'irregular' character [1] of highly excited states of the isolated system. Such deterministic yet chaotic dynamics is well understood for classical systems [2]. The point is that, strictly speaking, a quantal Hamiltonian with a purely discrete spectrum cannot give rise to chaotic dynamics. This can be argued in general [3,4] since with a purely discrete spectrum, any initial pure state can be expanded as a sum over eigenstates. Hence the time evolution is given by a discrete number of oscillating exponentials so that it is quasiperiodic. It can also be proved in detail by establishing the convergence of perturbation expansions [5] for a purely discrete spectrum. In classical mechanics, where the energy spectrum is continuous, these arguments fail. One can therefore conclude that there is no chaos in discrete quantal systems. Yet both experimental and computational results very clearly demonstrate that also in quantum mechanics it is useful to consider both the 'statistical' limit[6] and the behavior enroute to this limit.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 117–122.

1.1 The Distribution of Amplitudes

In addition to the technical problem we thus have a problem of interpretation. The wave functions of even the highly excited states can be numerically determined. Using enough care and computer time such converged computations have indeed been carried out. How then can we apply statistical considerations to the result? The problem is quite reminiscent of the basic issue in classical statistical mechanics. At a given instant in time, each molecule in a (dilute) gas has a well defined position and momentum. These will evolve according to the (Hamilton) equation of motion. Even for a chaotic system, given precise initial conditions, the future evolution is uniquely specified. Statistics only comes in when we ask for the distribution of velocities of all molecules at a given instant. In otherwords, when we ask for the distribution of velocities irrespective of position.

In constructing the distribution of amplitudes we shall use a similar approach. We shall ask for the distribution of amplitudes irrespective of, say, the corresponding energy. This can be understood in two ways. The first is that we pick a particular 'reference vector' in Hilbert space, say $|i\rangle$ and consider the distribution of the amplitudes $\langle f|i\rangle$ as we go over all the vectors $|f\rangle$ of interest. (As appropriate for a workshop at St John's College, we use the Dirac braket notation). Alternatively, we can specify a complete reference basis set of vectors $\{|n\rangle\}$ and examine the distribution of amplitudes $\langle n|f\rangle$ for a particular state $|f\rangle$ of interest.

In terms of the analogy with classical statistical mechanics, what we are doing is equivalent to our understanding of the Boltzmann velocity distribution circa 1890. Indeed, our invariance argument below (namely that in the limit, the distribution of the amplitudes $\langle n|f\rangle$ must be independent of the choice of the particular 'coordinate set' $\{|n\rangle\}$), is fashioned precisely by analogy with Maxwell's derivation of the Boltzmann distribution. (Recall that he required the velocity distribution to be invariant to the choice of the coordinate system used to define the components of the velocity vector. See e.g., [7])

There is one further amusing analogy with the example of the Boltzmann velocity distribution. When Otto Stern set out an experiment to verify the predicted distribution, the initial agreement was not perfect. The fit was much improved after Einstein pointed out that the experimental distribution must incorporate an additional Jacobian [8] (which is due to the faster molecules being preferentially sampled in the effusion). Here too, we shall find that the distribution that can be observed is not the one that we can most readily predict and differs from it by a Jacobian.

1.2 Quantum Chaos?

There is also a second sense in which the consideration of the distribution of quantal amplitudes is analogous to classical statistical mechanics. The time evolution of a trajectory in classical phase space which originates from sharply specified initial conditions can be traced back precisely. This is not the case when the initial conditions specify only a region and the dynamics are chaotic. Two trajectories which originate quite near to one another will, in time, exponentially diverge. This 'spreading out' or 'mixing' of an initially localized region in phase space is an important ingredient in modern ergodic theory[9]. It would be of interest to pursue a similar point of view with regards to the distribution of amplitudes. There have been preliminary attempts [10] to use such an approach, but much more work remains to be done.

1.3 Objective vs. Subjective in Quantum Mechanics

Beyond the purely technical problem, our considerations have a bearing on the issue of what is 'objective', (i.e., representing the real world), and what is 'subjective' (i.e.,

representing our knowledge of the real world) in quantum mechanics[11]. Our approach is that the amplitudes are objective. What is subjective is the distribution over the amplitudes. A fully specified pure state corresponds to a unique set of amplitudes and as such is analogous to a point in classical phase space. Typically, however, we can only specify a region in phase space (or a distribution therein) and we take the same to be true for the amplitudes which we regard as the 'coordinates' of the system in a Hilbert space. Elsewhere, [12], we have discussed the 'collapse' of the wavefunction upon measurement from the present point of view.

2. The statistical Wavefunction

To specify the wavefunction $|f\rangle$ of the final state accessed in the spectral transition we expand it in a fixed orthonormal basis set $\{|n\rangle\}$ in an N dimensional (Hilbert) space

$$|f\rangle = \sum_{n=1}^{N} x_{fn} |n\rangle \qquad (1)$$

The expansion coefficients x_{fn} are known as amplitudes since it is their squares which are probabilities. For our purpose it is convenient to think of them as the coordinates of the state $|f\rangle$ where x_{fn} is the projection of $|f\rangle$ along the n'th axis $\langle n|f\rangle = x_{fn}$. Strictly speaking the amplitudes can be complex numbers but for the moment we shall take them to be real. Hence, since the state $|f\rangle$ is normalised,

$$1 = \langle f|f\rangle = \sum_{n=1}^{N} x_{fn}^2 \qquad (2)$$

the amplitudes (for a given f) can therefore be regarded as the direction cosines of a unit vector in an N dimensional vector space. Different states $|f\rangle$ are orthogonal,

$$\delta_{f,f'} = \langle f | f' \rangle = \sum_{n=1}^{N} x_{fn} x_{f'n} \qquad (3)$$

which brings in correlations between the amplitudes belonging to different states.

In the chaotic limit we expect the state $|f\rangle$ to have components along many directions. They cannot all be equal to one another (and to $1/\sqrt{N}$) since different states need be orthogonal. We can, however, ask for their distribution, i.e. for the number of components $P_f(x)\,dx$ which have a magnitude between x and x + dx, for the given state f. Of course, the question is operationally reasonable only if the number, N, of possible components is large.

Note that the distribution is introduced in terms of the amplitudes rather than as the distribution of the probabilities. The reason is that one expects the distribution, in the chaotic limit, to be independent of any particular basis $\{|n\rangle\}$. Since it is the amplitudes which are the 'direction cosines' and hence it is the vector $\mathbf{x}$ of amplitudes which will be linearly transformed upon rotation, (i.e., upon change of basis)

$$\mathbf{x}'_f = \mathbf{U}\,\mathbf{x}_f, \quad x_{fn'} \equiv \langle n' | f \rangle, \quad U_{n'n} \equiv \langle n' | n \rangle, \qquad (4)$$

to a different basis $\{|n'\rangle\}$, it is the distribution of amplitudes which is of primary concern.

As discussed in the introduction, the reasoning is quite analogous to the statistical mechanics of a classical ideal monoatomic gas. Any mechanical state of the gas corresponds to a particular point in the 6N dimensional phase space of the N atoms. When however we ask for, say, the distribution of velocities irrespective of which atom, we get sensible results. Indeed, the invariance under rotation of axis is precisely the argument used by Maxwell [7] to derive the Boltzman distribution.

The distribution we are concerned with here is therefore of our own making and hence reflects our own state of knowledge. Any particular state $|f\rangle$ has definite values for the amplitudes along any particular direction $|n\rangle$. The probability density $P_f(x)$ is introduced by our asking for the distribution of the values of the amplitudes irrespective of the direction.

2.1. Maximum Entropy

In the limit of large N we can write the condition that the state $|f\rangle$ is normalised, equation (2), as

$$\frac{1}{N}\sum_{n=1}^{N} x_{fn}^2 = \int_{-\infty}^{\infty} x^2 P_f(x)\, dx \equiv \langle x^2 \rangle \qquad (5)$$

In (5), summation over all basis vectors n is replaced by an integration over all values of the amplitude, with $P_f(x)dx$ as the fraction of amplitudes with values in the range x to x + dx. Hence $P_f(x)$ itself need to be normalised

$$\int_{-\infty}^{\infty} P_f(x)\, dx = 1 \qquad (6)$$

If (5) and (6) are the only two constraints on $P_f(x)$, we obtain as the distribution of maximum entropy [13]

$$P(x) = (2\pi \langle x^2 \rangle)^{-1/2} \exp\left(-x^2 / 2\langle x^2 \rangle\right) \qquad (7)$$

We have dropped the subscript f since the density is universal. Note also that the Gaussian distribution (7) inherently satisfies the constraint that positive and negative values of x are equally probable, $\langle x \rangle = 0$.

2.2 The Distribution of Probabilities

The corresponding density of probabilities $y = x^2$ can now be obtained by the usual rule for change of variable in a probability density

$$P(y) = \left(2\pi \langle y \rangle\right)^{-1/2} y^{-1/2} \exp\left(-y / 2\langle y \rangle\right) \qquad (8)$$

The factor $y^{-1/2}$ comes from the Jacobian dx/dy and implies that low values of the probability $y = x^2$ are strongly favored.

2.3 The Prior Distribution of Probabilities

The probability distribution (8) can be derived directly as the distribution of maximal entropy provided one uses a measure (or a prior probability) for the distribution of y

$$P^0(y) \propto y^{-1/2} \tag{9}$$

One can indeed bring forth very general arguments why $y^{-1/2}$ is the proper prior distribution for probabilities [14]. Independent arguments leading to the same result have been presented at this meeting by Skilling. In the present context we prefer, however, the argument which uses a prior uniform density for the amplitudes because the physics is more obvious.

2.4 Complex Amplitudes

So far we have taken the amplitudes to be real. If they are complex, the required changes are quite straight-forward. Since the normalization is a constraint on $\langle |x|^2 \rangle$, the real and imaginary parts of x are independently distributed with a Gaussian density of width $\langle |x|^2 \rangle$. The essential change is that $y = |x|^2$ now has an exponential distribution

$$P(y) = (\langle y \rangle)^{-1} \exp(-y / \langle y \rangle) \tag{10}$$

Numerical studies (e.g., [15] for an early example) show that, in the statistical limit, the Gaussian density is accurate provided that one considers the amplitudes with respect to basis states of comparable energy. Many more details can be found in [16].

3. **The Optical Spectrum**

An ideal 'stick' spectrum is given by

$$S(E) = \sum_f y_f \delta(E - E_f) \tag{11}$$

where y_f is the intensity of the transition to the final state $|f\rangle$ of energy E_f,

$$y_f = |x_f|^2 = |\langle i | D | f \rangle|^2 \tag{12}$$

D is the transition operator and $|i\rangle$ is the initial state. What we are concerned with is the distribution P(y) of intensities irrespective of the energy of the final state. In otherwords, here too we generate a distribution by collapsing the energy resolved spectrum onto the intensity axis and examining the fraction, P(y) dy, of transitions with intensities in the range y to y + dy.

The intensities y_f satisfy a sum rule:

$$\sum_f y_f = \sum_f |x_f|^2 = \langle i | D^+ D | i \rangle \tag{13}$$

which can be regarded as the analog of the normalization condition (2) for the state $d|i\rangle$. Hence the distribution of spectral intensities in the chaotic limit is given by (8), (or by (10), if the amplitudes are complex).

3.1 Additional Constraints

The simple derivation above does lead to a distribution which sometimes but not always agrees with experimental results. To examine one (out of several) reasons for deviations we note that for spectra which access very highly excited states, each final state |f> in (12) can be expanded as in (1), hence

$$x_f \equiv \langle f \mid D \mid i \rangle = \sum_n x_{fn} \langle n \mid D \mid i \rangle \equiv \mathbf{x}_f \cdot \mathbf{d}^T \qquad (14)$$

where **d** is a fixed (for all final states f) vector whose components are the amplitudes $\langle i \mid D \mid n \rangle$. In otherwords, the set of amplitudes $\{x_f\}$, f variable, is obtained by a rotation of the set $\{d_n\}$. In the strict chaotic limit such rotations will have no effect on the distribution of amplitudes. If, however, it is possible to introduce a privileged basis set such that at the energy range of interest, the amplitudes $\langle n \mid D \mid i \rangle$ have a systematic structure then this will be reflected in the distribution of intensities. This diagnostic tool has served us well in a number of concrete examples (e.g. [17]).

Acknowledgement

I thank Prof. Y. Alhassid for discussion. This work was supported by the Stiftung Volkswagenwerk.

References

1. L. C. Percival, Proc. Roy. Soc. A **413**, 131 (1987); M. V. Berry, ibid, A **413**, 183 (1987).
2. H. G. Schuster, Deterministic Chaos (Physik Verlag, Weinheim 1984).
3. S. Golden and H. C. Longuet-Higgins, J. Chem. Phys. **33**, 1479 (1960).
4. R. Kosloff and S. A. Rice, J. Chem. Phys. **74**, 1340 (1981)
5. T. Yukawa, Phys. Rev. Lett. **54**, 1883 (1985).
6. C. E. Porter, Statistical Theories of Spectra: Fluctuations (Academic Press, N. Y., 1965).
7. M. Born, Natural Philosophy of Cause and Chance (Clarendon Press, Oxford, 1949).
8. D. R. Herschbach, Angew. Chem. **26**, 1221 (1987)
9. V. I. Arnold and A. Avez, Ergodic Problems of Classical Mechanics (Benjamin, N. Y., 1968).
10. F. Dyson, J. Math. Phys. **3**, 1191 (1962).
11. E. T. Jaynes, in this volume
12. R. D. Levine, J. Stat. Phys. (1988).
13. Y. Alhassid and R. D. Levine, Phys. Rev. Lett. **57**, 2879 (1986).
14. R. D. Levine, J. Chem. Phys. **84**, 910 (1986).
15. V. Buch, R. B. Gerber and M. A. Ratner, J. Chem. Phys. **76**, 5397 (1982).
16. R. D. Levine, Adv. Chem. Phys. **70**, 53 (1988).
17. J. P. Pique, Y. M. Engel, R. D. Levine, Y. Chen, R. W. Field and J. L. Kinsey, J. Chem. Phys. **88**, 5972 (1988).

JUSTIFICATION OF THE MAXIMUM ENTROPY CRITERION IN QUANTUM MECHANICS

ROGER BALIAN
Service de Physique Théorique de Saclay
Institut de Recherche Fondamentale
du Commissariat à L'Energie Atomique
91191 Gif-sur-Yvette Cedex, France

Abstract. By relying on the principle of indifference in a form suited to quantum mechanics, we prove that the density operator which should be assigned to a quantum system when only partial information is available has a generalized canonical form. This result provides an indirect justification of the quantal maximum entropy criterion, based on the use of von Neumann's entropy with constraints on the known expectation values.

We present below the main ideas of a work already published in a detailed form. We include some relevant references [1-5], in which a more complete bibliography can be found.

Our aim is to select the least biased density operator D in case the only available information is the set of expectation values $a_i = \langle A_i \rangle$ of some observables A_i. The state D of the system is supposed to be generated by some reproducible statistical device. In equilibrium statistical mechanics, the observables A_i are the constants of the motion; for irreversible processes, they are macroscopic noncommuting quantities, taken either at the initial time or at arbitrary times [3]; we also have in mind small quantum systems, prepared in a systematic fashion, then tested by measuring some observables A_i [5]. We assume that the data a_i have been observed on some samples of the statistical ensemble described by D; these samples must differ from one another if the observables A_i do not commute. We are interested in D before measurement, so as to disregard the resulting quantum perturbation and to make predictions about untested samples.

The available data

$$\mathrm{Tr} D A_i = a_i \tag{1}$$

are not sufficient in general for determining D. The *maximum entropy criterion* advocated long ago by Jaynes may achieve such a determination, through the maximization of

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 123–129.

$$S(D) \equiv -\mathrm{Tr} D \ln D \tag{2}$$

subject to the contraints (1). As in equilibrium statistical mechanics, this yields the generalized canonical distribution

$$D = \frac{1}{Z} \exp\left(-\sum_i \beta_i A_i\right), \tag{3}$$

where the Lagrange multipliers β_i are related to the data a_i by

$$Z \equiv \mathrm{Tr} \exp\left(-\sum_i \beta_i A_i\right), \qquad \partial \ln Z / \partial \beta_i = -a_i. \tag{4}$$

The maximum of (2) defines the *relevant entropy* relative to the data a_i, a quantity useful in irreversible statistical mechanics [3] as well as in measurement theory [5].

In spite of its many successes, this procedure has been criticized, even by Jaynes himself. Within the context of information theory, it is natural to require that the uncertainty due to the statistical description of a state by a density operator D should be measured by a quantity $S(D)$ which is additive and invariant in a change of basis; these conditions justify von Neumann's form (2) for $S(D)$. Moreover, when the observables A_i are constants of the motion, (2) can be identified with the thermodynamic entropy. But the identification of the *information content* (2) as a measure of *bias* in the choice of D is questionable : why should we require additivity for a measure of bias? A direct justification of (3),(4) from (1), by-passing the maximum entropy criterion, is therefore desirable. This has been achieved in various ways in ordinary statistics [1]. *Quantum mechanics* brings in difficulties, due to the non-commutation of the observables A_i and to the operator nature of quantum states D, which we have solved as follows [2].

Hypotheses. We shall rely only on the *principle of indifference* (or of insufficient reason) introduced by Laplace : *equal ignorance* implies *equal likelihood.* In order to apply this priciple, it is necessary to define the possible *elementary equivalent events* which will be considered equally probable. Their choice, which may be a source of difficulties [4], is always (though sometimes implicitly) based on some invariance, for instance under permutations for discrete events. In quantum mechanics, the required equivalence between events is defined by the group of *unitary transformations* in Hilbert space. Then, if our knowledge just amounts to arranging the basic events into two groups, those which are allowed and those which are incompatible with the available information, we should assign to the state a density operator acting in the subspace thus defined and invariant under unitary

transformations, i.e., D should be taken as proportional to the *projector* on this subspace.

The constraints (1) do not define such a subspace of the Hilbert space, and we now recast our problem into a form allowing to use the principle of indifference. To this end, we treat as a unique "*supersystem*" a set of N replicas ($N \to \infty$) of the system under study, labelled as $\alpha = 1, ...N$, and reminiscent of the Gibbs ensemble. This system has not a real existence, but represents a *collection of real or thought experiments* performed on the actual system, with always the same preparation. An elementary event for the supersytem is thus a collection of results for N experiments. With each replica of the original system is associated a Hilbert space h^α (all h^α have the same structure), and the overall Hilbert space $\mathcal{H}$ of the supersystem is the direct product of the spaces h^α. The (statistical) state of the supersystem is represented by a density operator $\mathcal{D}$ in $\mathcal{H}$. The single systems constituting the supersystem may be correlated because they belong to the same population; their individual density operators D^α are all equal to D, and can be derived from $\mathcal{D}$ by taking a *partial trace* over $N - 1$ systems:

$$D = \mathrm{Tr}_{\alpha=2,...N}\ \mathcal{D}. \tag{5}$$

We shall first *select* $\mathcal{D}$ by means of the *principle of indifference*, then *deduce* D *from it* by calculating (5).

Choice of $\mathcal{D}$. In order to transfer to the supersystem our information (1) about each of its constituents α , we introduce in $\mathcal{H}$ the mean observable

$$\mathcal{A}_i \equiv \frac{1}{N} \sum_{\alpha=1}^{N} A_i^\alpha \tag{6}$$

associated with each physical quantity A_i. The state $\mathcal{D}$ satisfies

$$\mathrm{Tr}\ \mathcal{D}\mathcal{A}_i = a_i, \tag{7}$$

consistently with (1),(6). However, $\mathcal{A}_i$ does not behave as A_i. In the limit $N \to \infty$, each $\mathcal{A}_i$ has a nearly continuous spectrum; moreover the commutators

$$[\mathcal{A}_i, \mathcal{A}_j] = \frac{1}{N} \mathcal{A}_{ij} \tag{8}$$

(where $\mathcal{A}_{ij}$ is associated through (6) with $A_{ij} \equiv [A_i, A_j]$) tend to zero. We can thus treat the observables $\mathcal{A}_i$ as *nearly classical variables*, which take the values a_i within vanishingly small errors for N large. This allows us to *identify the*

expectation values (1) for a single system with the *mean values* (6) over the super-system. A careless application of the principle of indifference would then provide for $N = \infty$

$$D \propto \prod_i \delta(\mathcal{A}_i - a_i), \tag{9}$$

which both satisfies the constraints on mean values and is invariant under the unitary transformations compatible with these constraints.

However, in order to work out (5) explicitly, we have to take N large but *finite*, and then (9) is unsatisfactory. Indeed, small fluctuations of $\mathcal{A}_i$ around a_i are required because (8) (where $1/N$ plays the role of $\hbar$) implies uncertainties of order $1/\sqrt{N}$. We perform therefore a coarse-graining on (9), taking for N finite

$$D = \frac{1}{\mathcal{W}} \exp\left[-\sum_i \frac{(\mathcal{A}_i - a_i')^2}{2\lambda_i^2 \varepsilon^2}\right]. \tag{10}$$

The factor $\mathcal{W}$ ensures normalization; when $N \to \infty$, we shall let $a_i' \to a_i$ and $\varepsilon_i \to 0$, while the λ_i's (introduced for dimensional reasons) remain constant. For N large, (10) behaves as (9), but it is meaningful for N finite while (9) is not. Because the final result (3),(4) will not depend on the particulars a_i', λ_i of (10), we can also take for D more general expressions, which involve *other shapes for the widening* of $\mathcal{A}_i$ around a_i and which are obtained by superposition of forms (10) with different parameters a_i', λ_i. We thus choose for D a form approaching a projector on a subspace such that $\mathcal{A}_i \simeq a_i$.

Calculation of D. It remains to derive from (5) and (10) the density operator D of a single system, letting $N \to \infty$ in the end. This calculation, which is not straightforward, turns out to be an amusing and instructive exercise, involving rather unexpectedly many techniques of field theory [2]. We present here just its first stages.

The explicit evaluation of a partial trace such as (5) over many subsystems $\alpha = 2, \ldots N$ is easy only if the contributions of the subsystems are factorized. We wish therefore to rewrite (10) as the integral of a product of factors depending each on $\alpha = 1, 2, \ldots N$. This would be readily done for each i if (10) were a product of exponentials rather than the exponential of a sum over i. Indeed, by carrying out for each i the Fourier transform

$$e^{-X^2/2} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} d\xi \, e^{i\xi X - \xi^2/2}, \qquad X \equiv \frac{\mathcal{A}_i - a_i'}{\lambda_i \varepsilon}, \tag{11}$$

we would then obtain an integrand which, according to (6), is a product of operators indexed by α and i. However, the contributions of the various i's to (10) are entangled. In order to treat each i separately while accounting properly for non-commutation, we use a standard trick in field theory : we introduce a fictitious time $0 < t < 1$, replace $(\mathcal{A}_i - a_i')^2$ in the exponent of (10) by

$$\int_0^1 \mathrm{d}t \, (\mathcal{A}_i(t) - a_i')^2,$$

and order the operators $\mathcal{A}_i(t)$ (now indexed by t) by means of a T-product. We can thus rewrite (10) as

$$\mathcal{D} = \frac{1}{\mathcal{W}} T \prod_{t,i} \mathrm{e}^{-K_i(t)\delta t}, \qquad K_i \equiv \frac{(\mathcal{A}_i - a_i')^2}{2\lambda_i^2 \varepsilon^2}.$$

The Fourier transform (11) can now be performed for each time t and each i, and $\mathcal{D}$ is expressed in the limit of infinitesimal time-steps $\mathrm{d}t$ as the *functional integral*

$$\mathcal{D} = \frac{1}{\mathcal{W}} \int \mathrm{d}\varphi \; T \exp \sum_i \int_0^1 \mathrm{d}t \left\{ -\varphi_i(t) \sum_{\alpha=1}^{N} \left[A_i^\alpha(t) - a_i' \right] + \varphi_i^2(t) N \mu_i / 2 \right\} \quad (12)$$

over variables $\varphi_i(t)$ running from $-\mathrm{i}\infty$ to $+\mathrm{i}\infty$ for each i and t (we have set $\mu_i^2 \equiv N\varepsilon^2 \lambda_i^2$). The precise definition of the measure $\mathrm{d}\varphi$ is given in [2].

Under the T-product, each contribution $\alpha = 1, 2, \ldots N$ now factorizes. The partial traces of (5) are easily performed, and each one brings in a factor

$$\mathrm{Tr}\; T \exp \left[-\sum_i \int_0^1 \mathrm{d}t \; \varphi_i(t) \, A_i(t) \right] \equiv \mathrm{e}^{\Phi[\varphi]}, \quad (13)$$

which is a functional of $\varphi_i(t)$. The resulting expression

$$D = \frac{1}{\mathcal{W}} \int \mathrm{d}\varphi \; \mathrm{e}^{N\Sigma[\varphi]} R[\varphi] \quad (14)$$

involves the functional

$$\Sigma[\varphi] \equiv \Phi + \sum_i \int_0^1 \mathrm{d}t \left[a_i' \varphi_i(t) + \frac{1}{2} \mu_i \varphi_i^2(t) \right], \quad (15)$$

and the functional

$$R[\varphi] \equiv T \exp\left[-\sum_i \int_0^1 \mathrm{d}t\, \varphi_i(t)\, A_i(t) - \Phi\right] \tag{16}$$

which is moreover an operator in the original Hilbert space h. The analogy of (14) with field theory is complete. The integration variables $\varphi_i(t)$ play the role of (Bose) *fields* with an *effective action* $\Sigma[\varphi]$, which includes, besides $\Phi[\varphi]$, a (time-independent) *source term* and a *mass term* (with bare masses μ_i small as $N\varepsilon^2$). We are interested in the *average* (14) *of R over the fields* φ. In (14), N plays the role of $1/\hbar$, and our field theory should therefore be worked out in its "*classical*" *limit* (although the operators A_i do not commute).

The result is easy to guess, since in this limit the weight is expected to be concentrated around the maximum of $\Sigma[\varphi]$, the fluctuations of the fields becoming negligible. In the limit $N \to \infty, N\varepsilon^2 \ll 1$, we have $a_i' \to a_i$, $\mu_i \to 0$, and $\Sigma[\varphi]$ is maximum for time-independent fields $\varphi_i(t) = \beta_i$ satisfying the condition (4) (note that (15) reduces then to $\ln Z + \sum_i a_i \beta_i$). For this "classical" value β_i of the fields, (16) is identical to (3). We thus get the *same result as with the maximum entropy criterion*, by simply starting from *equiprobability for the supersystem.*

In order to make this argument rigorous, we still have to show that the tree approximation which we have just sketched becomes exact in the limit $N \to \infty$. This is hard because the maximum of $\Sigma[\varphi]$ is very flat for $\varepsilon \to 0$. Indeed, the second order contribution in $\varphi_i(t) - \beta_i$ of (15) involves the vanishingly *small mass* term μ_i, plus a term from Φ which may also vanish. This vanishing reflects a kind of *gauge invariance* arizing from *commutation* of (3) with some linear combinations of A_i; actually, if the observables A_i all commute with one another, (13) does not depend fully on $\varphi_i(t)$ but only on its zero-frequency Fourier component (in this case, it is not necessary to introduce a time-dependence, and the set of fields $\varphi_i(t)$ is overabundant). Nevertheless, the estimation of the successive terms in the diagrammatic expansion of (14) has shown [2] that they are negligible in the limit $N \to \infty$, provided

$$N^{-3/4} \ll \varepsilon \ll N^{-1/2}. \tag{17}$$

The *upper bound for* ε was expected. This bound is necessary to fix effectively A_i at a_i in (10), within an error of order $N^{-1/2}$ imposed by the commutation relations (8). The *lower bound* has a purely quantal origin. It ensures that the number of orthogonal quantum states involved in (10) with approximately equal probabilities is large. In other words, $\mathcal{D}$ is approximately equal to a projector over a subspace with large dimension. This dimension, of the same order as $\mathcal{W}$, behaves as an exponential of N. Thus, possible regularities in the spectrum of $\mathcal{D}$ are smoothed

out by the large enough width ε, and (10) nearly describes equiprobability in spite of its rounding-off.

We have shown above how the maximum entropy criterion in quantum mechanics can be proved indirectly through its consequences (3),(4). A *more direct derivation* from the principle of indifference implemented as (10) can also be built [2] by defining the *relevant entropy* associated with the data a_i as $\lim_{N\to\infty} \ln \mathcal{W}/N$ (where $\mathcal{W}$ behaves as the *number of equiprobable states of the supersystem* compatible with the data a_i), then by identifying this quantity with the von Neumann entropy (2) of the state (3),(4). Alternatively, it would be desirable to justify the quantal maximum entropy criterion from consistency requirements (as in the works of Shore and Johnson, and of Tikochinsky, Tishby and Levine, ref. [1]), rather than by using the principle of indifference; but non-commutation of the observables A_i brings in serious difficulties.

As a final remark, note that the *projection of the state by a quantum measurement* enters the above framework [5]. Consider an ideal measurement of A_0 where all samples of the statistical ensemble are retained, irrespective of the result obtained about A_0. Take as observables A_i the ones which commute with A_0 ; their expectation values a_i are specified, being unchanged afin measurement of A_0, while any other information is lost in the process. This is consistent with the fact that the projected state has just the form (3), where the A_i's are the observables commuting with A_0.

References

[1] E.T. Jaynes, *IEEE Trans. Syst. Sci. Cybernetics* **4** (1968) 227; J.E. Shore and R.W. Johnson, *IEEE Trans. Inf. Theory* **26** (1980) 26; **27** (1981) 472; **29** (1983) 942; W. Thirring, "*Quantum Mechanics of Large Systems*", p.59 (Springer-Verlag, New York, 1983); Y. Tikochinsky, N.Z. Tishby and R.D. Levine, *Phys. Rev. Lett.* **52** (1984) 1357; **55** (1985) 336; *Phys. Rev. A* **30** (1984) 2638: Proof of the Maximum Entropy Principle in classical statistics.

[2] R. Balian and N.L. Balazs, *Ann. Phys.* **179** (1987) 97 : Proof of the Maximum Entropy Principle in quantum mechanics.

[3] R. Balian, Y. Alhassid and H. Reinhardt, *Phys. Rep.* **131** (1986) 1: Relevant entropy relative to a set of observables; use of $-\mathrm{d}^2 S$ as a metric in the space of (quantum) states; applications to irreversible statistical mechanics.

[4] R. Balian, M. Vénéroni and N.L. Balazs, *Europhys. Lett.* **1** (1986) 1: Importance of the underlying invariant measure in the definition of entropies.

[5] R. Balian and M. Vénéroni, *Ann. Phys.* **174** (1987) 229; R. Balian, subm. to *Am. J. Phys.*; subm. to *Europ. J. Phys.* : Incomplete preparations and measurements in quantum mechanics; projection of the state; entropy changes in a measurement.

APPROACHES TO NON-EQUILIBRIUM STATISTICAL MECHANICS

J. P. DOUGHERTY

Department of Applied Mathematics and Theoretical Physics
University of Cambridge
Silver Street, Cambridge CB3 9EW, UK

ABSTRACT. This work explores the relationship (if any) between two apparently disparate approaches to non-equilibrium statistical mechanics. These are (1) the methods of the Brussels School, notably that of subdynamics, (2) the methods developed by those working in the field of maximum entropy.

While the technical details of these approaches are both very formidable and quite different, it is possible to see in general terms how they are related. It is suggested that for many dynamical systems the approaches are likely to lead to the same results.

1. Introduction

The aim of statistical mechanics is to understand the macroscopic behaviour and properties of matter in terms of its microscopic (i.e. molecular) physics. While that may seem at first to be largely a technical matter, it can also be regarded as a paradigm for science generally; in that connection one may recall Popper's phrase "...the art of discerning what we may with advantage omit" (Popper 1982, p.44). Clearly, a very great deal of information has to be omitted, and in the present company this suggests that where 'information' is unavailable, maximum entropy procedures should be applied to make the best predictions. Such a development was initiated and pursued vigorously by Jaynes, starting in 1957 (see his collected papers, Jaynes 1983) and by his co-workers; a valuable exposition by Grandy (1988) has appeared recently. The same idea was developed independently by Zubarev (1974).

Another approach to this area is due to the so-called 'Brussels School' of statistical mechanics. An early account of it appears in the book by Prigogine (1962) but it has been much revised and extended since then, notably by the introduction of the concept of subdynamics. One version of it has been expounded systematically in the book by Balescu (1975). A central tenet of this approach is that everything should originate in the microscopic dynamics, implying that entropy maximisation is an additional assumption that should be unnecessary.

Neither of these schools has earned widespread external acceptance; they both appear to be very introvert, and each rigorously excludes reference to the other.

2. Basic Ideas

To expedite the simplification required in science, one wants to consider a macroscopic but relatively small and straightforward system, supposed isolated from the rest of the universe. The starting point is thus Hamilton's or Schrödinger's equations. This leads immediately to the fundamental problem of explaining irreversibility (Loschmidt's paradox).

The observed approach to equilibrium implies the existence of equilibrium macroscopic states even though microscopic dynamics would not recognize that concept. This in turn divides the subject into equilibrium statistical mechanics and non-equilibrium statistical

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 131–136.

mechanics (NESM).

The following are among the ingredients that have been proposed for understanding the foundations of statistical mechanics.

(a) *Open systems* No system can be completely isolated, and extremely small discrepancies involving very remote regions can produce large perturbations eventually. This is illustrated by Borel's well-known example, as refined by Berry (see Denbigh & Denbigh 1985, pp. 32-33). Formal consideration of the system in an infinite universe introduces the question of boundary conditions at infinity (as with the retarded potential in electromagnetism), so linking the 'arrow of time' to cosmology. Mathematically this often appears in manipulations of Fourier transforms, requiring a singularity to be displaced for reasons of 'causality', (e.g. in the Gell-Mann–Goldberger theory of quantum scattering, Landau damping in plasma physics, etc.). This can be referred to as the 'iε-trick' for introducing irreversibility.

(b) *Chaotic dynamics* This recognizes that, in a formally deterministic system, predictability over a long time may be unattainable. There are various levels of chaotic behaviour, but a good example is the 'mixing' condition (introduced intuitively by Gibbs with his illustration of ink and water), where the exponential divergence of orbits in phase space is involved. The result is that the mathematical problem of prediction is technically not a well-posed one because of sensitive dependence on initial conditions.

(c) *Information theory* can also be traced to Gibbs ideas (e.g. the most probable state) and the subsequent development by Shannon, but needs no further elaboration here.

The relevance of these points may well be challenged. In contemplating the diffusion of sugar into hot coffee it is hard to believe that the diffusion coefficient is controlled by events on distant galaxies as might be supposed from (a). Point (b) does not in itself introduce irreversibility, as it applies equally to retrodiction. To some physicists, point (c) introduces an unacceptable subjective element, which is examined in detail by Denbigh & Denbigh (1985).

3. Equilibrium Theory

Even here there appear to be rival versions for the Liouville density $\rho(X)$, the well-known microcanonical and canonical distributions. (X denotes a point of phase space.) These can be connected, respectively, with the strictly dynamical and the information-theoretical approaches to statistical mechanics. However, they lead to equivalent results for all practical problems (a consequence of a version of the central limit theorem, see Khinchin, 1949). There is therefore no opportunity to make a decisive choice either theoretically or experimentally.

We merely observe here that, in equilibrium theory, very general methods exist, but that there is *not* a unique choice of ρ that can be said to correspond to reality.

We note also that the canonical distribution leads to technically more tractable calculations.

4. Non-equilibrium Statistical Mechanics

It is natural to ask whether methods of similar generality exist in NESM. The verdict among physicists generally appears to be negative. For example Kubo (1978, p.10) wrote

> Is it possible to generalize Gibbs' ensembles to NESM? It would be great if we could answer this affirmatively, because we should be able then to start NESM from a basis as general at the microscopic level as we do in equilibrium statistical mechanics. But I think this is too much to be hoped. It cannot be done with such great generality, although I do not deny the possibility of doing it within a certain limitation or as an approximation.

Other quotations are given by Grandy (1988) in his preface. This point of view seems to originate in a feeling that the field of enquiry is too wide, encompassing the whole of the science of matter.

To the proponents of the approaches of the two schools that we have referred to, the answer is naturally positive; but their procedures are apparently unrelated. As noted in the equilibrium case the same macroscopic results may follow from differing methodologies, though such equivalence will be far more difficult to establish in the time-dependent case.

The difference between the two approaches can be traced to deep questions in the epistemology of probability. It seems better at present not to get immersed in ideology, but to note that both groups manipulate a time-dependent probability density $\rho(X,t)$ that satisfies Liouville's equation. We describe these procedures intuitively by regarding ρ as a vector in the appropriate Hilbert space, **H**, as in Figure 1. A level of macroscopic description is selected, with the result that ρ may be split (though not uniquely) into 'wanted' and 'unwanted' parts as shown. Each of the two axes actually represents an infinite-dimensional subspace. Liouville's equation, written symbolically

$$\partial\rho/\partial t = \mathbf{L}\rho$$

where **L** is a linear operator, defines trajectories $\rho(t)$ in **H**.

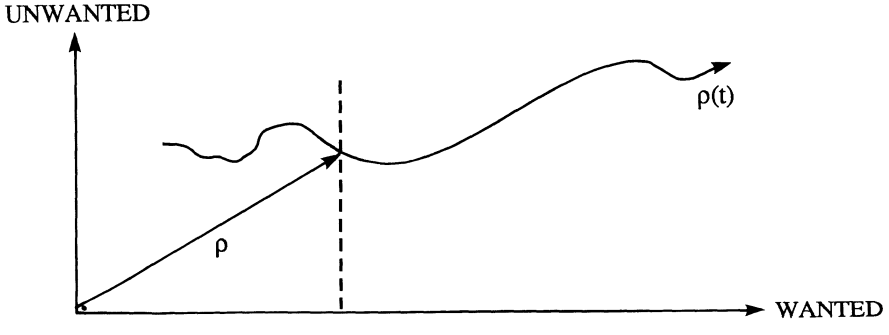


Figure 1

5. Subdynamics

Central to the Brussels school approach is the idea that a linear subspace of **H** exists, having such a dimensionality that to a given macroscopic state there corresponds a unique member ρ_S of the subspace, as in Figure 2. The subspace must have the additional property that any orbit $\rho(t)$ that intersects the space lies wholly in it. Macroscopic physics (at the level of description chosen) then corresponds to this 'priveleged' set of solutions of Liouville's equation. Dynamics thus becomes confined to the subspace, hence the name "Subdynamics". This procedure, (formidably difficult in practice) articulates earlier ideas about unwanted

details decaying rapidly (Bogoliubov's hypothesis). Irreversibility can be traced to the 'iε-trick'. (George, 1970).

This procedure is convincing if (as informally suggested in Figure 2) additional contributions lying outside the priveleged subspace decay, and especially if they do so rapidly. For this to be true, the operator **L** must have an approprate spectral property, which can in turn be related to the level of ergodicity (see for example, Parry, 1981).

The reduction results entirely from the properties of the Hamiltonian, i.e. from dynamics, and is obtained by applying asymptotic analysis to the formal solution of Liouville's equation in terms of its Laplace transform. There is no appeal to entropy or information.

Figure 2 shows how a a general Liouville function ρ can be mapped into $\rho_S=\Pi\rho$ that lies in the subspace and implies the same macroscopic state (i.e. wanted information). The map Π is then a projection operator in **H**, i.e. $\Pi^2=\Pi$. It has the further property

$$L\Pi=\Pi L$$

which ensures that the special choice $\rho=\rho_S$, if made at t=0, is preserved in the time evolution.

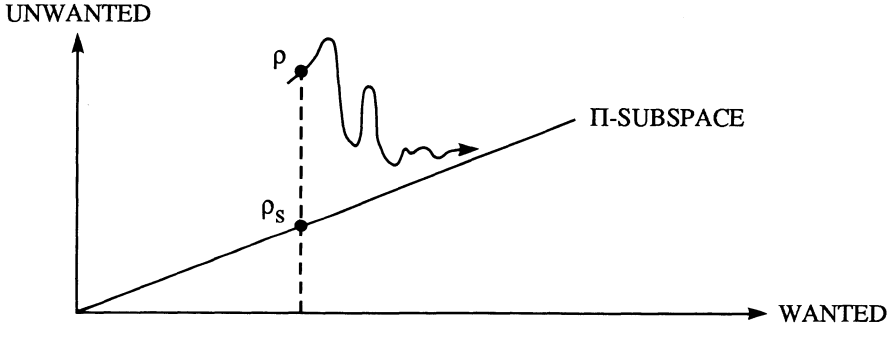


Figure 2

This may be contrasted with early attempts to 'discard information' by means of a projection operator, the well-known example being coarse-graining. There, the projection operator was selected on an intuitive basis, and does not commute with **L**. In subdynamics, Π, if it exists, is already determined by the dynamics of the system.

Balescu (1975) should be consulted for details of this theory, and its applications to a number of physical systems. He also gives the quantum mechanical version of the formalism.

6. Maximum Entropy Methods

To the present audience, an attractive way to select ρ at one instant, with prescribed values of the 'wanted' information, is to maximise the information entropy (which physicists call the

Gibbs entropy). The resulting values $\rho=\rho_m$ then lie on a curved manifold, as in Figure 3. This does not solve the problem, however, since the trajectories $\rho(t)$ satisfying Liouville's equation do not in general lie in the manifold. So the (nonlinear but idempotent) map from a general ρ to ρ_m does not commute with **L**. As Jaynes (1983, p.289) notes in the case of gas dynamics, the adoption of ρ_m at an instant implies the absence of transport effects at that instant, although the subsequent evolution would start them up after a very short time. But by then ρ would no longer be on the maximum entropy manifold. Jaynes and his coauthors modified the maximum entropy approach, adopting the following prescription. ρ must be a solution of Liouville's equation with the property that at each time t it maximizes the Gibbs entropy subject not just to the macroscopic data at time t but to values of that data at all earlier times <t. By means of a Heisenberg-picture calculation, this requirement can be expressed as an integral equation, with integrals over past time. The involvement over past but not future times introduces the symmetry-breaking.

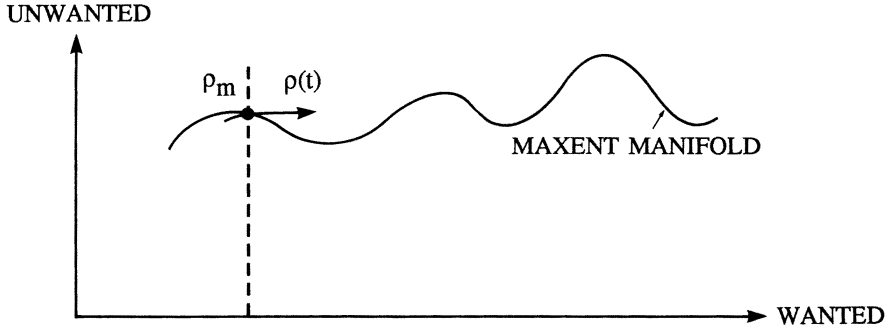


Figure 3

However Zubarev (1974) reaches essentially the same conclusion merely by displacing ρ from ρ_m and using Liouville's equation to solve for the displacement, in a way that is superficially similar to the construction of the master equation, but in terms of $\log\rho$ rather than ρ. Calculations of this type, in ρ itself, were precursors of subdynamics, see Balescu 1975, p.526.

The integral over past time that appears in these calculations must converge if the method is to succeed, and this too leads to an 'iε-trick' situation. This is explained in some detail by Zubarev. Very rapid convergence implies that the macroscopic variables evolve essentially in a Markovian way (no 'memory' effects). But their evolution is accompanied by transport effects, i.e. irreversible phenomena, though as mentioned above these are not present if one tries to base the theory on entropy maximisation at a single instant.

Presumably, a condition is required for this to be valid, just as was the case in subdynamics. However no such conditions are explicitly discussed in this part of the literature. The presentations by Jaynes (1983) and Grandy (1988) start in a more general formulation. They suppose that any set of incomplete data, extending over arbitrary subsets of space-time, is supplied. Their application of entropy maximisation then gives a prediction for

events in any part of space time that is entirely in the future of the regions containing the data. Nevertheless the problem considered here, the construction of closed equations for a given level of description, appears to be wide enough for NESM as generally understood.

7. Comparison

At first sight, the two approaches seem so different, both in philosophical foundations and in methodology, that comparison can hardly start. But underlying each of them is the essential idea of making a special selection from the class of solutions of Liouville's equation, in such a way that macroscopic physics is expressed by the appropriate projection onto the 'wanted' subspace. A unique value of ρ can then be ascribed to each macroscopic state, and this can serve as an initial condition for the special solution of Liouville's equation.

These specially selected solutions are different in the two approaches. We note from the equilibrium case that this is not necessarily a conflict, as the eventual macroscopic results may still agree, at least for large systems. Such very simple cases as can be handled at all indicate that outcome.

The mathematical relationship between the two approaches originates in the fact that subdynamics applies methods of linear analysis to Liouville's equation for ρ itself, while maximum entropy follows somewhat similar procedures for Liouville's equation in $\log\rho$. The detailed investigation is beyond the scope of the present paper.

8. References

Balescu, R. 1975 *Equilibrium and non-equilibrium statistical mechanics.* Wiley.

Denbigh, K.G. & Denbigh, J.S. 1985 *Entropy in relation to incomplete knowledge.* Cambridge University Press.

George, C. 1970 *Bull. Acad. Roy. Belg (Cl. Sci.)* **56**, 505.

Grandy, W.T. 1988 *Foundations of statistical mechanics. Volume 2: Nonequilibrium phenomena.* Reidel.

Jaynes, E.T. 1983 *Papers on probability, statistics and statistical physics.* Reidel.

Khinchin, A.I. 1949 *Mathematical foundations of statistical mechanics.* Dover.

Kubo, R. 1978 *Prog. Theor. Phys. Supp. no. 64,* p.1.

Parry, W. 1981 *Topics in ergodic theory.* Cambridge University Press.

Popper, K.R. 1982 *The open universe: an argument for indeterminism.* Hutchinson.

Zubarev, D.N. 1974 *Non-equilibrium statistical thermodynamics.* Plenum.

APPLICATIONS OF MAXIMUM ENTROPY TO CONDENSED MATTER PHYSICS

D. A. Drabold, A. E. Carlsson, and P. A. Fedders
Department of Physics, Washington University,
St. Louis, Missouri 63130, U.S.A.

ABSTRACT. We describe recent applications of maximum entropy to matter in condensed phases. Applications to spin systems, electronic structure, and calculating interatomic potentials are included.

I. INTRODUCTION

Maximum entropy methods have recently been applied to several kinds of problems in condensed matter physics. In broad outline these applications have fallen into either of two categories: moment problems (useful in spin systems, electronic structure calculations and densities of states for lattice vibrations), and a sophisticated application of maxent to the calculation of interatomic potentials in metals. These approaches have had considerable success, and our expectation is that extensions of these methods, and entirely new applications will be developed in the future. In solid state physics workers sometimes resort to approximations that have no *a priori* justification -- an example we will mention is the use of *ad hoc* functional forms to invert moment problems. Beside the information theoretic advantage of using maxent, another point in favor of this procedure is that it provides concrete functional forms to manipulate and base other approximations on: this can be compared to some large-scale computer calculations. It will be the goal of this paper to familiarize maxent practitioners with recent work in condensed matter and to relate a few rather generally encountered properties that might occur in other applications. We will organize this paper as follows: Section II will discuss the maxent solution of the classical moment problem and physical applications. Section III will include a brief discussion of methods for obtaining interatomic potentials via maxent.

II. MOMENT PROBLEMS IN CONDENSED MATTER

The classical moment problem may be stated as follows: Given the first N power moments of a non-negative function $\rho(x)$ on some interval $a \leq x \leq b$:

$$\mu_n = \int_a^b dx\, x^n \rho(x) \quad , \quad n = 0,1,2,...,N \quad , \tag{1}$$

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 137–142.

develop an approximation for ρ based on the information contained in the moments μ_n. For finite N, it is clear that the solution to the moment problem is not unique: many functional forms can be invented which correctly reproduce the known moments, but which may differ (sometimes radically) in the unknown higer moments. This lack of a unique solution leads us to consider what the "optimal" functional form might be. The answer is provided by the method of maximum entropy.[1] Following Mead and Papanicolaou, we may construct an entropy

$$S = -\int_a^b dx\, \rho(x)[\log \rho(x) - 1] \quad , \tag{2}$$

to be maximized subject to the constraint that ρ should have the required first N moments. Using the usual procedure of introducing an auxiliary functional with undetermined Lagrange multipliers to include the constraints, and functionally differentiating with respect to ρ, we easily arrive at the maxent solution of the problem:

$$\rho(x) = \frac{1}{Z} \exp\left\{-\sum_{i=1}^{N} \lambda_i x^i\right\} \quad , \tag{3}$$

and Z, λ_i are determined by requiring that ρ satisfy the moment constraints. The determination of the λ_i is difficult, owing to the nonlinearity of the maxent ρ and the many quadratures involved in an iterative procedure. A Newton minimization procedure was given in Ref. 1. Bretthorst[2] has developed a more robust algorithm, and Drabold[3] has found some sum rules that speed the original Mead and Papanicolaou code up by more than a factor of two. Despite all of this, it is not hard to find examples for which the maxent code fails. Not surprisingly, this tends to happen for functions which have explicit singularities, or those with discontinous derivatives. Such functions are sometimes of physical interest.

Before discussing examples of moment problems in condensed matter, we note that these moment problems are a recurring theme of the subject. Solid state theorists tend to work with moment formulations of problems because they offer an alternative to the task of diagonalizing large matrices. For a spin ½ problem for example, the dimensionality of the Hamiltonian matrix which contains all dynamical information is 2^K where K is the number of spins -- typically order 10^{23} for a macroscopic system! In contrast, calculation of the low order moments is usually fairly straightforward.[4] Also, the moments tend to contain information about the local environment of a particular site; information one is usually interested in. The complete set of eigenvalues and eigenvectors associated with the Hamiltonian contains vastly more information, most of which is irrelevant to an investigators specific interests. It is also worth noting in passing that the reason why one can readily extract the moments is related to the fact that the trace of a quantum mechanical operator is independent of the choice of basis.[5] In each of the examples we discuss, this is the key to obtaining the moments of physically relevant functions. In fact, with the appearance of reasonably reliable code for solving the maxent moment problem, we may think of

maxent as providing us with an alternate numerical method for diagonalizing Hamiltonian matrices: After all, we can (in principle) *always* calculate traces of powers of such matrices. These traces are to within a normalization exactly the power moments of the density of states for the Hamiltonian matrix. So for cases where symmetries or other considerations allow easy calculation of powers of the Hamiltonian, maxent should be considered as a means of obtaining the density of states. One other potential application of maxent moment methods to solid state physics is the improvement of convergence of certain expansions. For example, a high-temperature expansion takes the form of an infinite series in traces of powers of the Hamiltonian. Information theoretic extrapolation for higher order terms in the series may provide a useful means of extracting physically meaningful results for lower temperatures.

The first physical application of the moment problem we discuss is the calculation of response function $G(\omega)$ for spin systems.[6] This function has the physical interpretation of a spectral density: it may be thought of as indicating the "density of excitation" per unit frequency range. It is clear that such a function must depend upon the detailed dynamics of the spins, which is naturally a many-body quantum mechanical problem. Several kinds of spin-spin interactions have been examined: Mead and Papanicolaou[1] applied maxent to the one-dimensional XY model and Heisenberg exchange. Impressive agreement with the exact solution of the XY model was obtained. Because maxent was well converged in the Heisenberg case (in the sense that the answer did not change appreciably with additional moments), these authors reasonably concluded that they had an essentially exact solution for the spin dynamics of the Heisenberg system. This result is significant, because there is no exact solution known.

For a calculation which may be directly compared to experiment, we turn to the case of a magnetic dipolar coupling between spins ½. Here, careful experiments have been done on CaF_2 where the fluorine nuclear spins are arranged in a simple cubic lattice. Some complicated calculations[7] have produced eight exact moments for the physically measurable "lineshape" function $G(\omega)$ for this system. In Ref. 8 maxent was applied to the theoretical moments and found to agree with experiment to within ~ 2%. In general the function $G(\omega)$ is complex-valued, the real part representing the NMR absorption spectrum, which is clearly positive definite. It has been found convenient in other calculations [9] to introduce a function related to G, the self energy $\Sigma(\omega)$ which satisfies the equation $G(\omega) = i/\{[\omega - \Pi(\omega) + i\Gamma(\omega)]\}$, where Π and $-\Gamma$ are the real and imaginary parts of Σ respectively. It can be shown that Γ is of one sign, and therefore another candidate for the application of maxent. Power moments of Γ are readily related to the known theoretical moments of $G(\omega)$. Although the direct use of maxent on the function G was very satisfactory, optimal agreement was obtained by fitting Γ as an intermediate step. The reason for this appears to be that Γ has less structure than G and is therefore easier to apply maxent to. The utility of the auxiliary function Σ is related to the function-theoretic properties of G and Σ on the complex plane. This point is discussed further in Ref. 8. Another interesting feature of the work on the dipolar lattice was the appearance of an oscillating pattern

of convergence. As others have shown, it is quite possible to find sets of moments for which the maxent procedure does not converge. By this we do not refer to a numerical difficulty, but to an intrinsic limitation of the method for certain sets of input moments. For the dipolar case it was observed that for $N = 4k + 2$ (N the number of input moments, k an integer), no maxent solution existed on the infinite interval. This result is connected with the expected large–ω behavior of the exact G, which on physical grounds is expected to decay like $e^{-a|\omega|}$, for some a. On the other hand, the maxent fitting function behaves like $e^{-\lambda_N \omega^N}$ thus to reproduce the correct behavior, there must be considerable cancellation for large ω, implying that the Lagrange multipliers should not be of one sign. Cases for which $\lambda_N < 0$ can lead to situations for which there is no maxent solution. Numerically this non-convergence is manifested by a dependence of the λ_i on the cutoffs for the numerical integrals (the actual range of integration for G and Γ is $\omega \in (-\infty, \infty)$). It was also found that self-energy and lineshape fits were complimentary in the sense that when the lineshape (self energy) produced a non-converged calculation, the self-energy (lineshape) function converged. So in some cases, one may be forced to use an auxiliary function to obtain a converged maxent fit.

We have also applied maxent to a more complicated version of the previous problem, the case of a nonmagnetic host with spins $-\frac{1}{2}$ randomly diluted throughout a crystal.[10] A particular realization of such a system is ordinary diamond, due to the existence of two isotopic species of carbon: spin 0 (magnetically inert) and spin $\frac{1}{2}$. For high concentrations of magnetic particles it was found that maxent and configuration averaged moments produced good line shapes. For low concentrations of spins, we used maxent as an aid in inferring to what extent spin wavefunctions were localized (in the terminology of magnetic resonance this characterized the dipolar broadening as inhomogeneous or homogeneous).

We have also recently applied maxent to the problem of obtaining theoretical estimates of relaxation times in solid molecular hydrogen. We have observed reasonable agreement between theory and experiment.[11]

Maxent has been used to obtain densities of states in binary random alloys. Here, there have been a wide range of methods applied, from exact diagaonalization of large matrices to recursion methods. For a particular model calculation[12] it was found that maxent offered a real alternative to continued fraction and coherent potential approximation (CPA) methods. While it is certainly true that the CPA method produces very satisfactory results in a wide range of regimes, it is limited in some contexts by mean-field like assumptions underlying its derivation. The formulation of recursion and maxent moment methods do not suffer from this weakness. Maxent also has an advantage over recursion; in the usual implementation of recursion the electronic Green's function takes the form of a continued fraction which must be terminated in some way. This is unfortunately more an art than a science. While an experienced practitioner of recursion would correctly argue that a particular choice of truncation schemes incorporates knowledge of the physics of the problem, we point out that such information could also be included as an additional

constraint on maxent, without the introduction of bias. As in the spin problem it was useful to reconstruct functions related to the electronic Green's function rather than the Green's function directly. This was again because the auxiliary functions were better behaved. In the alloy problem it was best to use a function which bears the same relation to the self energy that the self energy did to the Green's function in the spin problem. This procedure produced the best agreement with a CPA calculation. We should also point out that maxent and recursion are complementary to some extent, as recursion provides the most efficient means of calculating the moments needed for the maxent procedure. We also tried a specific example of one vacancy in crystalline Si, and found maxent to be superior to continued fractions.[13]

Brown and Carlsson[14] provided the first application of maxent to structural energy calculations in the presence of defects such as vacancies. Very recently a comprehensive study of methods for calculating bond energies in a tight binding model has appeared.[15] It was found that maxent was a useful means for computing structural energies in the presence of defects, better than continued fractions with a square root terminator, but only roughly equal in accuracy to a gaussian quadrature[17] approach which was computationally easier for more than six recursion levels. Glanville, *et al.*[15] observed that there were computational difficulties with maxent due to an extreme sensitivity of the maxent fit to the values of the Lagrange multipliers conjugate to the moments. A possible remedy for this difficulty is to solve the moment problem on a different basis. The origin of the trouble lies in the nearly singular nature of the covariance (Hessian) matrix -- this is a consequence of the increasing degree of correlation between higher moments. A possible solution is to take N linearly independent combinations of the moment constraints, solve the moment problem on the new constraints, and transform back. Bretthorst[2] has even gone so far as to construct an orthogonal basis, though any reasonable combinations should help significantly. Turek[16] has independently implemented these ideas and finds that his code is much improved over the original approach of Mead and Papanicolaou.[1]

III. INTERATOMIC POTENTIALS VIA THE MAXIMUM ENTROPY PRINCIPLE

The study of defects and structural energetics in metals is greatly aided by the concept of effective interatomic potentials. This field has suffered from the lack of uniform methods for obtaining such potentials. One of us has recently shown[18] that calculating interatomic potentials can be formulated as a problem of incomplete information: Given knowledge of the changes in the two point density-density correlation function in a condensed matter system, how can one best guess the associated energy changes? To answer this question, one must obtain guesses for higher order correlation functions (*i.e.*, triplet, four-body ...). Knowledge of these functions is necessary for calculating total energies since these depend on clusters of 3,4, $\cdots$ particles. Maxent can be used to estimate these functions, and thus produce a rigorous foundation for future work in the area. The method produces expressions for the effective potential as a functional of the pair density-density

correlation function. For further details of this approach, we refer the reader to Ref. 15.

ACKNOWLEDGEMENTS

We are very grateful to Professor John Skilling for his generous support in making this presentation possible. We would also like to thank Professor E. T. Jaynes for many stimulating discussions. One of us (DAD) would like to thank Professor Jaynes for giving the most lucid and thought-provoking lectures he has ever encountered, and for numerous helpful hints in the course of research. This research was supported in part by NSF Grant No. DMR–88–02160, and DOE Grant No. DE–FG02–84ER45130.

BIBLIOGRAPHY

1. Lawrence R. Mead and N. Papanicolaou, J. Math. Phys. **25**, 2404 (1984).
2. G. L. Bretthorst, Proceedings of the 7th Annual conference on Maximum Entropy (to be published).
3. D. A. Drabold (unpublished).
4. D. A. Drabold, Phys. Rev B **37**, 565 (1988).
5. See for example Albert Messiah, *Quantum Mechanics*, (Wiley, 1959).
6. A. Abragam, *Principles of Nuclear Magnetism,* (Oxford, 1985).
7. S. J. Knak Jensen and E. Kjaersgaard Hansen, Phys. Rev. B **7**, 2910 (1973).
8. P. A. Fedders and A. E. Carlsson , Phys. Rev. B **32**, 229 (1985).
9. See for example C. W. Myles and P. A. Fedders, Phys. Rev. B **9**, 4872 (1974).
10. D. A. Drabold and P. A. Fedders, Phys. Rev. B **37**, 3440 (1988).
11. D. A. Drabold and P. A. Fedders, submitted to Phys. Rev. B.
12. A. E. Carlsson and P. A. Fedders, Phys. Rev. B **34**, 3567 (1986).
13. A. E. Carlsson and P. A. Fedders (unpublished).
14. R. H. Brown and A. E. Carlsson, Phys. Rev. B **32**, 6125 (1985).
15. S. Glanville, A. T. Paxton and M. W. Finnis, J. Phys. F **18**, 693 (1988).
16. I. Turek, J. Phys. C **21**, 3251 (1988).
17. C. M. Nex, J. Phys. A **11**, 653 (1978).
18. A. E. Carlsson , Phys. Rev. Lett. **59**, 1108 (1987).

PROBLEMS OF MAXIMUM-ENTROPY FORMALISM IN THE STATISTICAL GEOMETRY OF SIMPLE LIQUIDS

ROBERT COLLINS
Physics Department
University of York
York YO1 5DD, England

TOHRU OGAWA, TAEKO OGAWA
Institute of Applied Physics
University of Tsukuba
Ibaraki 305, Japan

ABSTRACT. Recent work has revived the approach to the equilibrium theory of classical liquids via coding theory and statistical geometry. Each atomic configuration is specified in terms of its Voronoi honeycomb and Delaunay graph, and the perfect gas distribution is used as a prior for the probability density of a Voronoi bond-length. An entropy maximisation then leads to an equation of state in closed form. Preliminary results in two dimensions agree with computer simulations within the approximations made, and include a qualitatively correct account of the liquid/gas phase transition. Further progress, including a description of ordering, requires the solution of several formal problems. These are discussed here in some detail since some of them seem to involve points of general interest in the formulation of maximum entropy methods.

1. Introduction.

The equilibrium theory of simple monatomic liquids is now generally regarded as well-established, in that it is possible to obtain close numerical agreement between theory and experiment in particular density ranges by using the YBG, PY or HNC systems of equations (see Hansen and McDonald 1986 for a recent review). However, there is still no single formalism applicable to the liquid/vapour system over the *whole* range of fluid densities. In an attempt to supply this, it was decided to re-examine the approach via statistical geometry originally due to J.D.Bernal. This had been later combined with the Jaynes-Shannon maximum-entropy formalism (see for example Collins (1972) for a historical review) but the results gave no liquid/gas phase transition, and the wrong low-density limit for the bond distribution.

The new point of departure of the recent work was to use prior distributions which are those of a perfect gas. Even in two dimensions and with a first-order treatment using relatively crude approximations, the results show an immense improvement (Collins, Ogawa and Ogawa 1987. Since this paper will be cited repeatedly in what follows, for conciseness it will be denoted by I). The liquid/gas transition now appears naturally and of course the low-density limit is now trivially correct. In this paper we summarize the present formalism and discuss the problems of extending it to account for an ordering transition. In two dimensions this might be to a genuine solid crystal, as indicated by some computer simulations (Abraham 1980), or a "hexatic phase" (Kosterlitz and Thouless 1972,1973; Halperin and Nelson

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 143–149.

1978; Nelson and Halperin 1979). In either case the transition will here be termed "ordering".

2. Voronoi Coding of a Two-Dimensional Liquid.

The liquid is assumed to consist of a large number N of identical atoms in an area (2-dim. volume) V. Boundary effects are ignored. The arrangement of atoms is irregular, with no long-range correlation, and statistics independent of position (no gravitational field). In particular the expectation atomic density averaged over any small region is ρ=N/V.

Any given atomic configuration is specified in terms of its *Voronoi honeycomb H* and its topological dual *Delaunay net D*, shown in Figure 1.

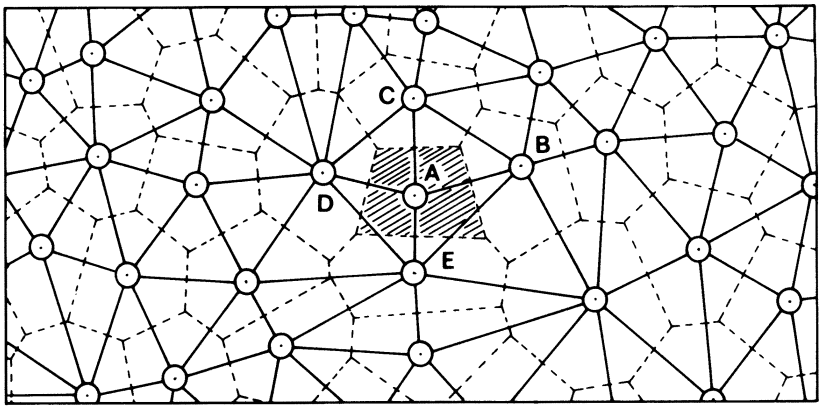


Fig.1. Voronoi honeycomb and Delaunay net for an arbitrary liquid-like atomic configuration.

To a typical atom A is assigned a polygonal *neighbourhood cell* (shaded) consisting of all points nearer to A than to any other atom. The set of these cells constitutes *H* (dotted lines). Except for degenerate cases (of zero total probability) three cells meet at any vertex of *H*. Atoms with adjacent cells are defined to be neighbours, irrespective of their distance apart. To construct *D* we join all neighbour pairs by straight lines termed *bonds* (solid lines), so that *D* consists of an irregular set of close-packed triangles with the atoms at their vertices. An atom where q bonds terminate is said to have (Voronoi) *valency q*, or (for short) to be a *q-atom* (In Figure 1, A is a 4-atom with neighbours B,C,D,E). In general $3 \leq q$, although 3-atoms are rare even in a (two-dimensional) perfect gas. Any bonds from the same atom will be termed *adjoining*.

The atomic configuration is specified if all the bond-lengths and atomic valencies are given (Collins 1967). We now set up an expression for the information entropy H of the system relative to that of a perfect gas (here called the "G-case" for short, to avoid confusion with pressure P) in which there are no interatomic forces and the atoms have a random (Poisson) distribution. The suffix G denotes quantities evaluated for this case. For a first treatment no attempt is made to model the

details of local correlations. The information entropy h=H/N per atom is then given by

$$h = \log(\rho_{ref}/\rho) - 2\int_{-\infty}^{\infty} dp\ f(p)\log[f(p)/f_{ref}] - \sum_{q=3}^{\infty} W_q \log(W_q/W_{Gq}) - j\int_0^{\infty} db\ \psi(b)\log[\psi(b)/\psi_G(b)] \qquad (1)$$

Here f(p) is the probability density of a Cartesian momentum component p of a randomly selected atom. ρ_{ref} and f_{ref} are constant priors. Their actual values do not affect the calculation and for present purposes they may be ignored. W_q and W_{Gq} are (general and G-case) probabilities that a randomly selected atom has valency q. Similarly, $\psi(b)$ and $\psi_G(b)$ are the corresponding probability densities for a randomly selected bond-length b. The parameter j (=3ϛ in I) is defined to be the mean *effective* number of *independently* assignable bond lengths needed to fix the position of an atom. This parameter is essential since to replace j by 3 (the actual number of bonds per atom) would greatly overestimate h by neglecting correlations between adjacent bonds. For any plane *D* containing N atoms there are (neglecting edge effects) a total of 3N bonds (I) so that trivially $j<3$. However *D* can be built up by fixing the positions of each atom in succession. Each position is defined by *two* bonds linking the atom to those already fixed and hence we can write down the stronger inequality $j\leq 2$. Comparison with computer simulation results (I) suggests that the appropriate value of j is very close to its upper limit, i.e. $j\simeq 2$.

The term $\log(\rho_{ref}/\rho)$ is the entropy change of a perfect gas for a density change from ρ_{ref} to ρ. From the information theory viewpoint it expresses, for example, the fact that it takes 1 bit extra of information to locate an atom in a box which is twice as big. The reason that a uniform prior f_{ref} for f(p) is adequate, while the non-uniform prior $\psi_G(b)$ is required for $\psi(b)$ is that the set of b's forming *D* is subject to the topological and geometric constraints of the 2-dimensional space in which it is embedded, while the p's are not. Even when explicit relations for these constraints can be found (e.g. Collins 1968) they are not well-adapted for use in a *variational* calculation. Since the G-case is also subject to these constraints, they can be made implicit in the calculation by using G-case priors. The exact analytic form of $\psi_G(b)$ is known (Collins 1968) and the W_{Gq} can be obtained by computer simulation (Finney 1970).

3. Thermodynamics and the Equation of State.

In I it was argued that the terms in W_q should be of minor importance except near an ordering transition and they were omitted (this view was supported by the eventual results). The interatomic forces were modelled by assigning to each bond of length b an additive pair potential energy $\varphi(b)$. With an average of 3 bonds per atom the expectation internal energy u per atom is given by

$$u = \int_{-\infty}^{\infty} dp\ \frac{p^2}{m} f(p) + 3\int_0^{\infty} db\ \psi(b)\varphi(b) \qquad (2)$$

Here m is the mass of an atom. For given ρ and u, h was then maximised (to give the thermodynamic entropy s per atom) by varying f(p) and $\psi(b)$. The result for f(p) was the usual Boltzmann distribution, while $\psi(b)$ was given by

$$\psi(b) = \frac{1}{z}\psi_G(b)\exp\left[\frac{-2\mu}{j}\left[\frac{P}{\rho T}-1\right]b/\rho - \frac{3\varphi(b)}{jT}\right] \qquad (3)$$

Here z normalizes $\psi(b)$ to unity and $\mu(\simeq 0.9)$ is a geometric factor relating mean bond length to mean triangle area. P is the pressure and T is the temperature in energy units (so that Boltzmann's constant is unity). The equation of state is obtained by substituting from (3) into the relation

$$\int_0^\infty db\ (1-\mu b/\rho)\psi(b) = 0 \qquad (4)$$

This was tested using the standard Lennard-Jones potential

$$\varphi(b) = 4\epsilon\left[(\sigma/b)^{12}-(\sigma/b)^{6}\right] \qquad (5)$$

and comparing with computer simulation results (Abraham 1980). The resulting isotherms agree with the computer results to within about 15% over the whole density range, and the critical temperature T_c is given to within 5%, although the values of P_c and ρ_c are wrong by factors of 2 and 4 (for details see I). This is as much as could be expected with the approximations made, and at least establishes the viability of the general method. The ordering transition (shown by the simulation results) was absent from the theoretical curves, as would be expected following the omission of the valency terms from (1).

3. Inclusion of Valency and Correlation Terms.

The most obvious improvement to be made in maximisation of h is to retain the terms in W_q in (1), subject to the extra constraint

$$\langle q\rangle \equiv \sum_{q=3}^{\infty} qW_q = 6 \qquad \text{(two dimensions)} \qquad (6)$$

which is exact in the limit $V\to\infty$ for *any* two-dimensional *D*. This by itself, however, gives no new physics, since the h maximisation simply gives W_q decaying exponentially with q independently of $\psi(b)$ (which is unaffected by the new terms) and so we obtain the same equation of state as before. It follows that if the formalism is to account for an ordering transition, terms (called "q-b" correlations for short) specifically correlating the W_q with $\psi(b)$ must be included. At the simplest useful level, these are the conditional probabilities $W_q[\psi(b)]$ that a randomly selected atom has valency q *given* a particular $\psi(b)$.

For $T<\epsilon$, $\psi(b)$ consists (see Fig.5 of I) of a sharp peak (representing the condensed liquid or ordered phase) together with a long "tail" (vapour phase). The peak occurs at b_m ($= 2^{1/6}\sigma$ for the Lennard-Jones case) where $\varphi(b)$ has its minimum value $-\epsilon$. Consider for simplicity the case where P is much larger than

the saturated vapour pressure P_{sv} of the condensed phase), so that the vapour phase tail is absent and $\psi(b)$ is negligible except near b_m. The proposed ordering mechanism is then as follows. As T is further reduced, the peak at b_m becomes higher and narrower. Formally we can write

$$\psi(b) \rightarrow \delta(b-b_m) \qquad (T\rightarrow 0,\ P_{s\,v} << P) \qquad (7)$$

so that all the triangles in D are nearly equilateral of side b_m. Their internal angles are then all nearly $\pi/3$ and hence all the atoms must be 6-atoms. Hence we must have

$$W_q[\psi(b)] \rightarrow \delta_{q,6} \qquad (\psi(b) \rightarrow \delta(b-b_m)) \qquad (8)$$

There is a qualitative argument (Collins 1967) that a plane D containing only 6-atoms is necessarily ordered (at least in the hexatic sense). Hence an entropy maximisation leading to a sharp change of the W_q to $\delta_{q,6}$ would by itself denote an ordering transition. The theory would then be (within the approximations made) a unified theory of the 2-dimensional solid/liquid/gas system, and in particular should predict a triple-point. So far however, no analytic functional form for $W_q[\psi(b)]$ has been found.

Correlations ("b-b") between adjoining bonds have already been partially allowed for in the factor j, which is related to the concept of the "entropy power" of an information source (p.60 of Shannon and Weaver 1949). In fact j is probably the best way of expressing b-b correlations. Detailed considerations of local geometry would then "fine-tune" the theory to reveal (for example) any density-dependency of j.

Inclusion of "q-q" correlations (between valencies of neighbouring atoms) may also be regarded as fine-tuning, in that the main qualitative features of the isotherms can be obtained without them. However, they must be included for the theory to predict the correct values of the critical exponents. Without these correlations, although the general shape of the isotherms near the critical point is correct (Fig.4 of [1]) the predicted exponents have their (incorrect) classical values (e.g. γ=1 instead of the correct 7/4 (Fisher 1974)). The difficulty here is that current maxent formalism provides no clear way to include correlations of this type. Since this correlation problem must occur in other applications, there may be a gap in the general theory which could usefully be filled.

4. Extension to Three Dimensions.

In three dimensions, the cells of H are polyhedra. For a general statistical structure, except for a set Q of configurations of total probability zero, 4 cells meet at any vertex of H and hence (again except for Q) D consists of close-packed tetrahedra instead of triangles. With these changes the general formalism can be written down much as before, but the actual implementation is much more difficult. Even for a first-order liquid/gas theory (neglecting correlation and valency terms) two immediate problems in three dimensions are

(i) $\langle q \rangle$ is not a topological constant (cf. (5))

(ii) the form of $\psi_G(b)$ has yet to be found.

In modelling the solid/liquid transition, there is the further serious difficulty that the perfect forms of both the close-packed crystal forms (fcc and hcp) fall into the

degenerate zero-probability set Q previously mentioned. Consequently in the solid state the atomic positions must be given small random perturbations to break this degeneracy. When this *is* done, there is much evidence that (5) can be replaced to a good approximation by $<q> \simeq 14$, at least for high densities (Collins 1968). Point (i) still presents a serious problem in the liquid/gas critical region since there $<q>$ could vary quite rapidly with ρ.

We may summarize the foregoing by noting that all the extra problems in three dimensions arise from the topological and geometric differences between two and three dimensions. Those problems which seem to have wider implications for general maxent problems, occur already in the two-dimensional case.

5. Conclusions.

The combination of statistical geometry with a maximum entropy formulation seems to provide a good qualitative description of the two-dimensional liquid/gas system with a simple interatomic additive pair potential. We can identify the specific problems which need to be solved in order that the theory may account for ordering, and to extend it to three dimensions. Some of these are special to this particular problem, but others may require further developments in general maximum entropy formalism.

The method described here seems to be one of the few maxent calculations in which it is crucially important to use exactly the right prior. It also demonstrates that, in order actually to use the concept of S as the maximised information entropy H, it is not necessary to start from the convential partition function Z of equilibrium statistical mechanics. In fact it seems extremely doubtful if the results described here *could* be obtained in any straightforward way starting from Z. To quote Domb and Green (1972, p.xi) "The theory, although far from the fundamental principles of statistical mechanics, is very near to our intuitive conception of what a liquid really is." The work described here seems to go some way torward removing the reservation while preserving the main comment.

There is still a widely-held view that, although the Jaynes-Shannon formalism provides an improved foundation for equilibrium statistical mechanics, the actual results of JS calculations could equally well be obtained by traditional methods. The translation of Bernal's statistical geometry into liquid thermodynamics seems to provide quite an effective counter-example to this view.

6. Acknowledgements.

Thanks are due to Professor J. Finney, Professor J. Powles, Dr.J. Skilling and Dr. J. Livesey for comments and suggestions at various times, which have been very helpful in preparing this paper.

7. References.

Abraham FF (1980) in *Ordering in Two Dimensions* (ed Sinha SK. N Holland).

Collins R (1967) in *Phase Stability of Metals and Alloys* (eds Rudman PS, Stringer J, Jaffee RI. McGraw-Hill).

Collins R (1968). J Phys C 1,1461.

Collins R (1972) in Domb and Green (1972).
Collins R, Ogawa T, Ogawa T (1987). Prog Theor Phys **78**,83.
Domb C, Green MS, eds (1972). *Phase Transitions and Critical Phenomena 2* (Acad Press).
Finney JL (1970). Proc Roy Soc Lond **A319**,479.
Fisher ME (1967). Rep Prog Phys **30**,615.
Halperin BI, Nelson DR (1978). Phys Rev Lett **41**,121,519.
Hansen JP, McDonald IR (1986). *Theory of Simple Liquids* (2nd ed. Acad Press).
Kosterlitz JM, Thouless DJ (1972). J Phys C **5**,L124.
Kosterlitz JM, Thouless DJ (1973). J Phys C **6**,1181.
Nelson DR, Halperin BI (1979). Phys Rev B **19**,2457.
Shannon CE, Weaver W (1959). *The Mathematical Theory of Communication* (Univ of Illinois Press).

LIQUID STRUCTURE FACTOR DETERMINATION BY NEUTRON SCATTERING - SOME DANGERS OF MAXIMUM ENTROPY

G.J. Daniell and J.A. Potton
Department of Physics
University of Southampton
SOUTHAMPTON SO9 5NH

ABSTRACT: This paper is a tutorial account of how the maximum entropy method has been applied to the determination of pair correlation functions of liquids and amorphous materials using neutron scattering. This is an example where care needs to be taken in the definition of the entropy, in the inclusion of prior information and in the stopping criterion. It is easy to obtain results that are misleading or manifestly wrong and these dangers are particularly stressed. Nevertheless when used with understanding the method is a very satisfactory approach to this unstable inverse problem.

1. Introduction to the problem

In this tutorial paper we consider in detail one problem in neutron scattering: the determination of structure factors or pair correlation functions. This is a problem requiring several of the advanced tricks of the maximum entropy method and we illustrate the use of these.

We are concerned with liquids or amorphous solids where the atoms are not on a regular lattice and want to determine a function g(R) which describes the statistics of the interatomic separations. Suppose we pick one atom and look at a distance R from its centre. There is no chance of finding the centre of another atom very close to the first but at a separation corresponding to one atomic diameter the probability rises to a maximum. Subsidiary maxima occur at distances corresponding to other coordination shells and at large separations the probability of finding another atom becomes constant in the random structure. Figure 2 shows experimentally determined functions g(R) showing these features.

The data obtained by neutron scattering are measurements of a function S(Q) where $Q = (4\pi/\lambda)\sin\theta/2$ and θ is the scattering angle and λ the wavelength of the neutrons. Theory shows that S(Q) is related to g(R) by

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 151–162.

$$S(Q) = 1 + \frac{4\pi\rho}{Q}\int_0^\infty R[g(R) - 1]\ \sin QR\ dR \qquad (1)$$

where ρ is the density of the material. This integral equation for $g(R)$ has an analytic solution: $Q(S(Q)-1)$ and $R(g(R)-1)$ are Fourier sine transforms.

It is clear that for $Q(S(Q)-1)$ to possess a sine transform $Q(S(Q)-1)$ must approach zero sufficiently fast as $Q \to \infty$. If our data has systematic errors so that this limit is violated we might anticipate trouble in finding a solution by Maximum Entropy.

Although the inversion of a Fourier transform is stable this problem is not very stable because of the factor R inside the integral. The value of $g(R)$ for small R gets suppressed and is poorly represented in the data. Therefore although the Fourier inversion proves the existence and uniqueness of the solution it is not a good way of calculating it.

The serious objection to any use of the analytic inverse is that we do not have $S(Q)$ for all Q. In any scattering problem we have a limited range of scattering angles and we cannot resolve detail in $g(R)$ on a scale much less than the wavelength of the neutrons or much greater than a limit set by the angular resolution of the detector.

Figure 1a shows a typical analytic inverse. Note the instability for small R.

2. Use of the Maximum Entropy Method

This is an obvious target for the Maximum Entropy method but we must define a proper probability density before it can be applied. Consider a spherical sample of radius a with n atoms per unit volume, then

$$\frac{1}{\frac{4}{3}\pi a^3}\int_0^a 4\pi R^2\, g(R)\, dR = n.$$

Define

$$p(R) = \frac{3R^2 g(R)}{a^3 n}$$

so that

$$\int_0^a p(R)\, dR = 1\ ,$$

then $p(R)$ is the required probability density. In the absence of data the simplest assumption requires that $p(R)$ is proportional to the volume of a spherical shell, that is $p(R) = 3R^2/a^3 n$ or $g(R) = 1$.

We make the unconstrained maximum of the entropy equal to this by adopting a measure $m(R)$ on R space $m(R) \propto R^2$.

The entropy is therefore

$$S = -\int p(R)\ \log \frac{p(R)}{m(R)}\ dR = -\frac{3}{na^3}\int R^2 g(R)\ \log \frac{R^2 g(R)}{eR^2} \qquad (2)$$

The scale of S is irrelevant so we can drop the premultiplying constant. The result is that in our inverse problem we should determine $R^2g(R)$ with a default proportional to R^2 and get g(R) by dividing at the end by R^2.

We have to discretise the problem to apply a Maximum Entropy computer package and since this is a frequent source of confusion we include the steps in detail.

Our experiment fixes two quantities, Q_{max} which is greater than the largest value of Q in use and ΔQ which is less than the smallest resolvable difference in Q. The first is fixed by the wavelength of the neutrons and the second usually by the resolution of the diffractometer. These numbers imply that we cannot obtain a resolution in R much better than $1/Q_{max}$ or determine g(R) for R much greater than $1/\Delta Q$.

To discretise equation (1) cut off the integral at $R_{max} \gg 1/\Delta Q$, introduce $\Delta R \ll 1/Q_{max}$ and replace the integral by a sum, so that

$$S(Q) = 1 + \frac{4\pi\rho}{Q}\Delta R^2 \sum_r r\,[g_r - 1]\ \sin rQ\Delta R \qquad (3)$$

where we have written $R = r\Delta R$ and $g(r\Delta R) = g_r$ with r an integer. We have to perform this sum for each value of Q for which we have a measured S(Q).

Experiments are rarely done in equal steps of Q, more frequently in equal steps in θ. Nevertheless we will often want to use the Fast Fourier transform to evaluate the sum because of its speed; but we then get S(Q) on a uniform grid in Q. To use the Fast Fourier transform we fill out the r sum with zeros until the number of terms is a power of two and to take $\Delta Q\Delta R = 2\pi/2^N$ to use a 2^N point transform. The inequalities for ΔQ and ΔR imply an inequality for N. In particular $2^N \gg 2\pi Q_{max}/\Delta Q$. In contrast with the use of the FFT to compute the analytic inverse the value of 2^N is not fixed by the number of data points. There is no objection to using too large a value of N apart from the speed of the calculation but as we illustrate below there are dangers in using too small a value.

In our calculations we have used the Fast Fourier transform and interpolated on the Q grid to get S(Q) at the experimental Q values. We also at this point take into account the resolution of the diffractometer. The actual data are averages over a small range of Q. Since we need to interpolate in any case the inclusion of some smoothing in Q corresponding to the diffractometer resolution merely changes the weights that are used.

The problem is now almost in standard form. The Maximum Entropy package that we have used is written for the strictly linear problem so in equation (3) the constant term and Σ r sinrQΔR are moved to the left hand side and combined with S_q producing modified data. Although such preprocessing of data is contrary to the philosophy of modern data processing there is no great objection in this particular case.

Figure 1b shows the Maximum Entropy calculation on the same data as Fig. 1a, which is for a $Y_{63}Cu_{37}$ alloy. The double main peak corresponds to the Cu-Y spacing and the Y-Y spacing. No peak is visible corresponding to the Cu-Cu spacing but this is expected to be weak. The result is very satisfactory. Note particularly the smooth rise of g(R) to 1 at small R in contrast to the erratic behaviour in Fig. 1a. This is due to the absence of large Q data; because there is no information about small distances the solution goes to the default level g(R) = 1.

We do actually have extra information that is not contained in the data; atoms cannot be arbitrarily close together and $g(R) \to 0$ at small R. For aesthetic reasons this was built into the solution by modifying the default level m(R) in equation (2) and the result is shown in Fig. 1c. A striking feature has appeared, the new peak corresponds more or less to the expected Cu-Cu separation. We should however be very suspicious of something that is very sensitive to the default level; the new peak is barely resolved and a smaller value of ΔR is called for. When the calculation is repeated with a smaller value of ΔR, the result of Fig. 2a is obtained. When we try to resolve the peak it has vanished!

It is frequently stated that maximum entropy cannot introduce artifacts into the solution as a result of the data processing. This is true but only if it is used correctly. The discretised problem must adequately describe the real problem to better than the accuracy of the data. There is no objection to taking ΔR too small, Maximum Entropy will correctly give a smooth curve. The only penalty is increased computer time and ultimately poor convergence of the algorithm. Figs. 1c and 2a strikingly illustrate the dangers of taking ΔR too large.

We can also note another artifact that has crept in at large R. This is because R_{max} is too small, and increasing it removes the problem, as illustrated in fig.2b.

3. The automatic correction of background and density values

It was pointed out in section 1 that S(Q) must approach unity correctly as $Q \to \infty$ or there may not exist a solution even for perfect data.

Since large Q corresponds to fine structure in R we might anticipate spurious fine structure in g(R). This is illustrated in Figure 3. This was calculated earlier in our work and suffers from the wrong value of ΔR but it shows rather a lot of fine structure in R that is not physically explicable. Although we want to see fine structure we want only credible fine structure.

We need to remember that in the actual experiment the measured count rate has to be corrected for the background. If we get the background subtraction or calibration wrong then S(Q) will not approach unity but some other value and we expect small changes in the asymptotic level of S(Q) to produce big changes in the fine structure of g(R). We should therefore subtract or add a constant to S(Q) and see the effect on the solution. If we are really not sure of the

background we could choose a shift in S(Q) that maximises the final entropy of the solution. This will remove as much structure as possible from g(R).

Similarly we need the value of ρ, the density. This is bound to be uncertain and we can also adjust ρ to maximise the entropy. One loses structure in g(R) by doing so, but it is better to lose some if what remains is more credible. Figures 3a to 3d show in sequence the results of systematically adjusting the background and density values to increase the final entropy. It is clear that much of the structure in the initial g(R) can be attributed to incorrect values of these parameters. This is a second example of a danger in Maximum Entropy; the data and the theory must be consistent to an accuracy better than the noise in the data or artifacts can be introduced into the result.

4. The Residuals

To convince the experimentalist that we have correctly analysed his data we should look at the residuals in the fit of the predicted data to the observed data. This is shown in Fig. 4 for simulated data. We see that there is a poor fit and that will not endear us to the experimentalist. Other contributors to this conference have described different stopping criteria and better models; these are obviously the sort of things that are needed. Here we just want to explain why this poor fit between the predictions and the data occurs.

The entropy is maximised by making the solution as close as possible to the default level and that is unity for most of the range R. Peaks in g(R) are pulled down and troughs in g(R) are filled in and the amplitude of the oscillations in g(R) is kept as small as possible. Because the operation in equation (1) is roughly a Fourier transform the oscillations in g(R) correspond to the peak in S(Q). So the predicted data has a significantly lower peak in S(Q) than the real data, and this is clearly shown in fig. 4.

Similarly, the main peak in g(R) is pulled down most strongly. Peaks in g(R) correspond to oscillations in S(Q) and therefore the oscillations in S(Q) have too small an amplitude. Again Fig. 4 shows these features in the misfits to the data.

We know the Maximum Entropy method produces biassed results; we put up with this because of its good points, but there is a serious side effect in this problem which is illustrated in fig. 5 where noise has been added to the simulated data. So much χ^2 is used up in the bias that there is little left over for the noise. The consequence is that noise in the data gets through into g(R). The bias, the overfitting of the noise in the data, and the noise in the resulting g(R) are clearly visible in fig. 5. This seems to us to be a fundamental limitation of classic maximum entropy in this problem.

Finally on this problem we show the good points. One of the main reasons for using maximum entropy is to overcome lack of data at large Q. Fig. 6 shows the effect of truncation and it is clear that acceptable results can be obtained even from severely truncated data.

5. Partial Structure Factors

We finally turn to an extention of this problem. The function g(R) contained information about the distribution of interatomic spacings but it said nothing about the different atomic species. Because different isotopes have different scattering lengths, measurements with different isotopes of the same material enable us to determine which bond lengths correspond to which pairs of atoms.

For a binary material there will be four functions $g_{11}(R)$, $g_{12}(R)$, $g_{21}(R)$, $g_{22}(R)$. Roughly speaking $g_{12}(R)$ is the probability of finding an atom of species 2 at a distance R from one of species 1. Obviously $g_{12} = g_{21}$ but both functions need to be included separately in the entropy which is defined as

$$S = - \sum_{ijR} R^2 \, g_{ij}(R) \log g_{ij}(R).$$

The data are 3 sets of measurements which we write as a 3 component vector $\underline{d}(Q)$. This is related to the vector of g functions by an equation of the form

$$\underline{d}(Q) = \underline{\underline{A}} \int_0^\infty \frac{R}{Q} \, \underline{g}(R) \sin QR \, dR \qquad (4)$$

where $\underline{\underline{A}}$ is a 3 × 3 matrix involving the scattering lengths of the nuclei.

The obviously correct approach to this problem is to choose a set of g_{11}, g_{12} and g_{22} that maximises the overall entropy and fits the 3 data sets with a constraint on the overall total χ^2.

Fig. 7 shows some results with simulated data. We expect the superimposed overlapping peaks seen, for example, in fig. 2b to be separated and each peak to appear in one of the three functions g_{11}, g_{12} and g_{22}. Fig. 7a, obtained using the method just described, shows that our expectations are not borne out and each of the functions contains all three peaks. The algorithm has moved "g value" between the three curves and thereby got a higher entropy. We stress that this solution fits the data! The overall χ^2 is correct and there is no doubt this is a correct solution, to the maximum entropy problem formulated above.

If we look at the χ^2 for the three individual data sets that went into this calculation they are not equally well fitted, but we do not normally divide our data into groups and apportion equal amounts of χ^2 to each group. The extreme case where each data point must make one unit contribution to χ^2 is absurd.

We have tried a lot of things to get rid of this failure to separate the three functions. There is nothing to prevent "g value" being traded between them. Of course we want it traded between different points of the same function; that is the whole philosophy of Maximum Entropy. The only way out would seem to be a better default,

but we don't know where the peaks are; the point of the experiment is to find them. The improved default models suggested by other speakers at this conference suggest possible ways forward.

Until such investigations have been tried a practical alternative is to write equation (4) as

$$\underline{\underline{A}}^{-1}\ \underline{d}(Q) = \int_0^\infty \frac{R}{Q}\ \underline{g}(R)\ \sin QR\ dR.$$

We first construct new data $\underline{\underline{A}}^{-1}\ \underline{d}(Q)$ and analyse these three new data sets separately. $\underline{\underline{A}}$ is poorly conditioned and the errors in the data are magnified when $\underline{\underline{A}}^{-1}\underline{d}$ is computed but providing we allow for this we see no objection to this approach. The resulting 3 functions $\underline{g}(R)$ will not of course be independent.

We have obtained very satisfactory results from real experimental data using this method.

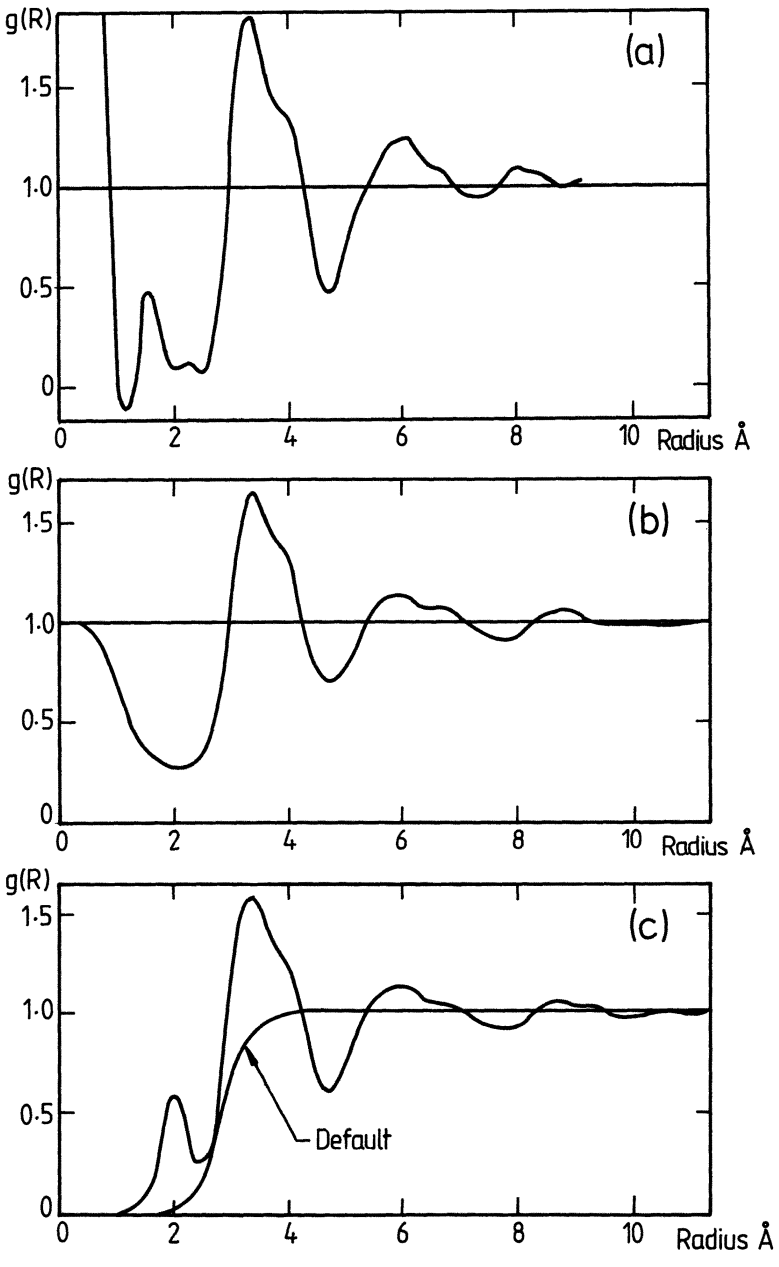


Fig. 1. Pair distribution function g(R) for Cu-Y alloy. a) Fourier Inversion. b) Maximum Entropy, flat default. c) Maximum Entropy, more realistic default but showing spurious line.

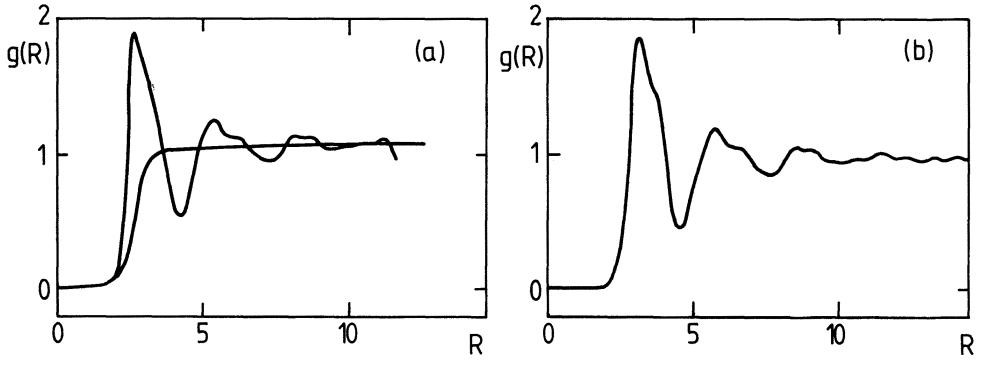


Fig. 2 The same as 1c but with increased resolution.
a) Showing small artifact at large R. b) Artifact removed.

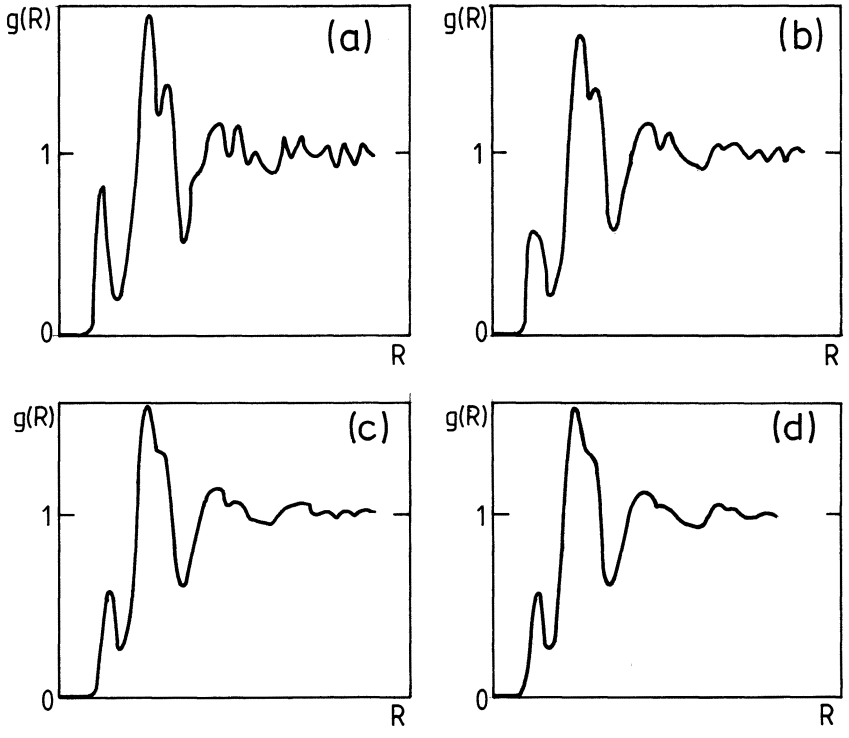


Fig. 3 (a), (b), (c), (d). Successive stages in adjusting backgroun and density values showing removal of artifacts. The spuriou line at small R arises because of insufficient resolution.

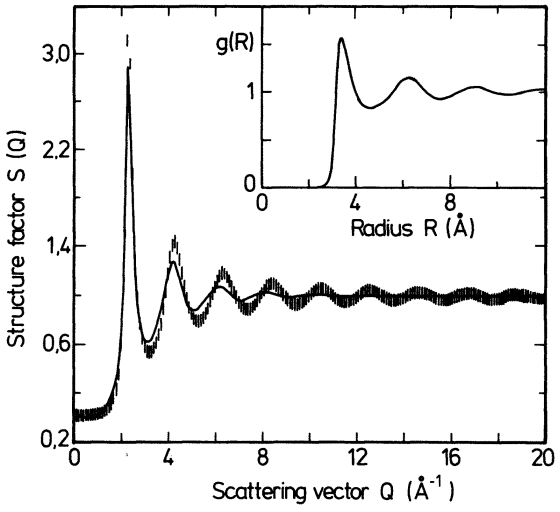


Fig. 4 The residuals. Solid line: values of S(Q) predicted from the maximum entropy solution. Bars: actual data points and errors.

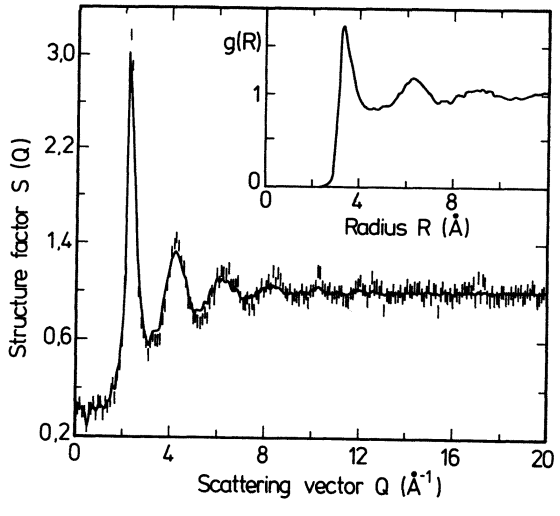


Fig. 5 The addition of noise to Fig. 4. The bias causes noise in the maximum entropy solution for g(R).

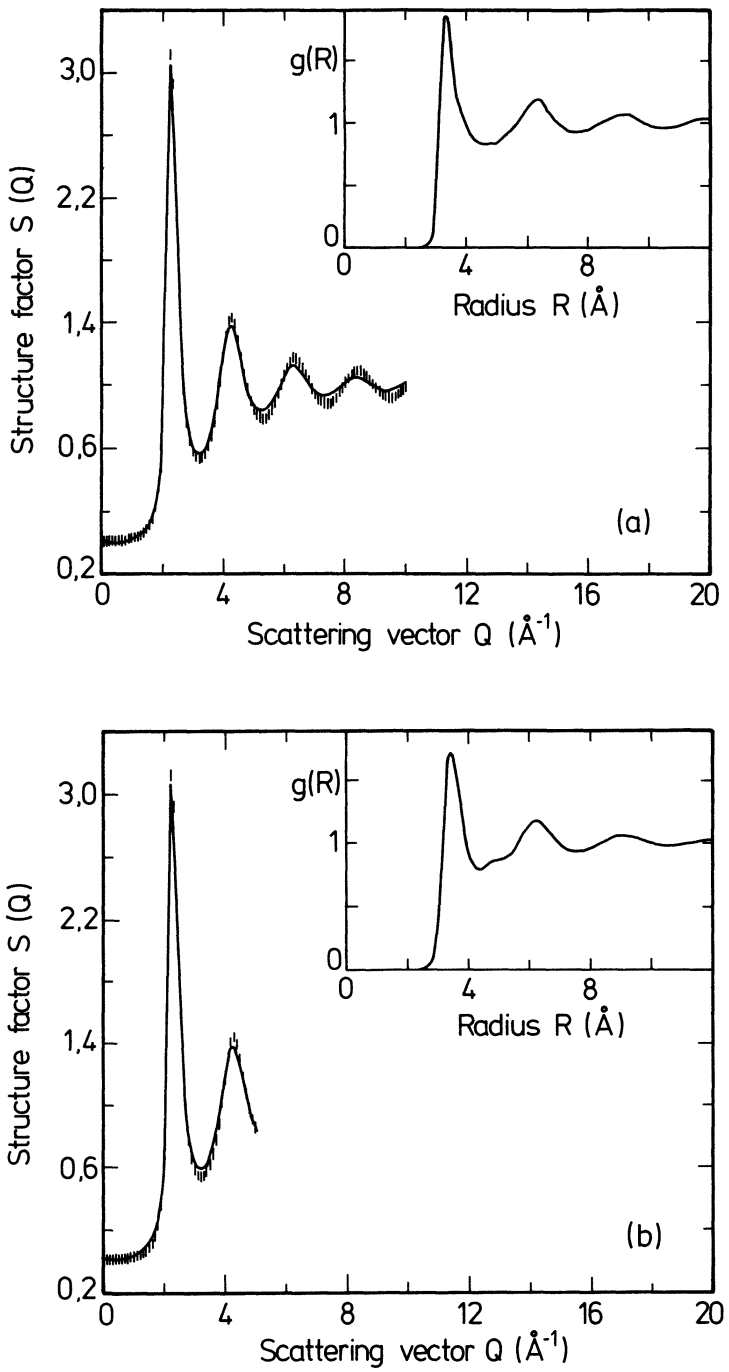


Fig. 6 Effect of truncation of large Q data. a) Moderate truncation and b) severe truncation. The resulting maximum entropy solutions are hardly effected.

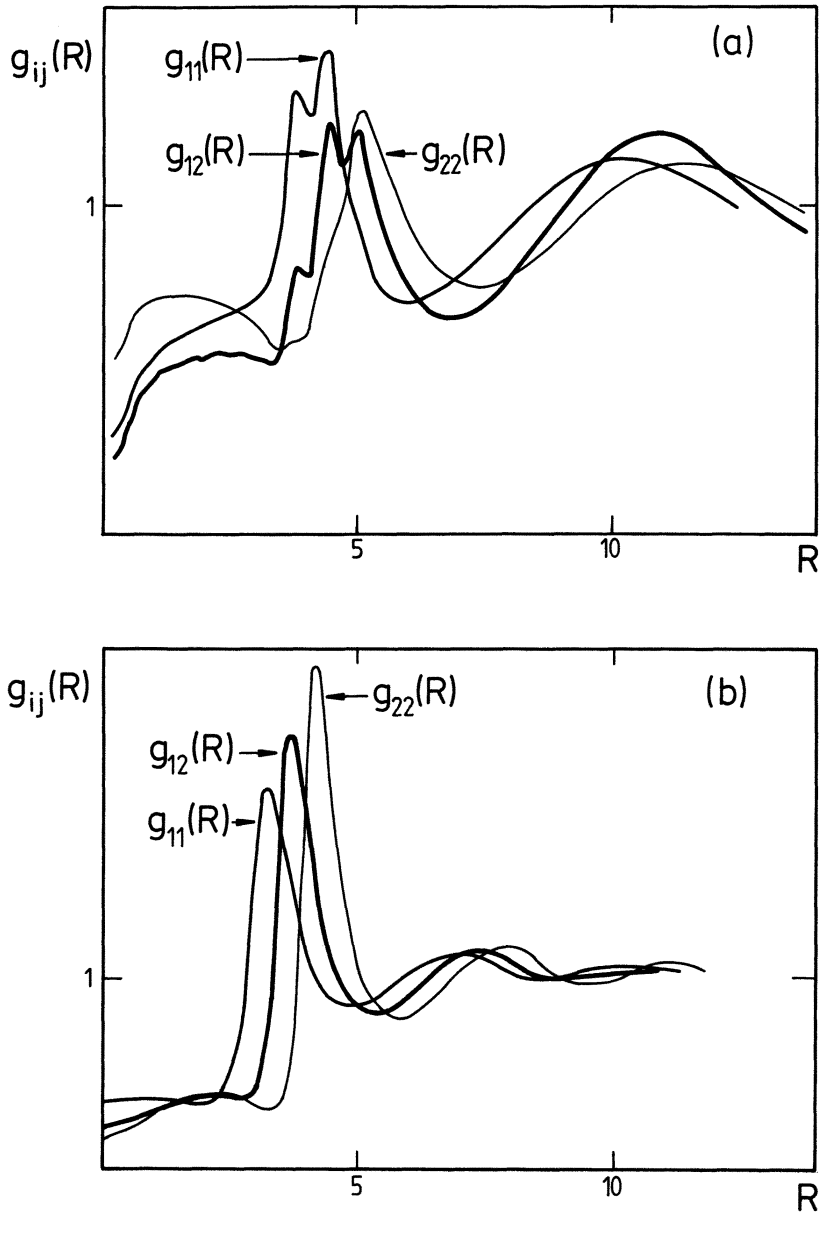


Fig. 7 Partial Structure factors. a) The simultaneous calculation by Maximum Entropy showing failure to separate the functions. b) An alternative method showing acceptable results.

QUASIELASTIC NEUTRON SCATTERING DATA EVALUATION USING THE MAXIMUM ENTROPY METHOD

R.J. Papoular
Laboratoire Léon Brillouin (CEA-CNRS)
C.E.N. Saclay
91191 Gif-sur-Yvette Cedex - France

A.K. Livesey
MRC Laboratory of Molecular Biology
Hills Road, Cambridge CB2 2QH, England
and
Department of Applied Mathematics and Theoretical Physics
Silver Street, Cambridge CB3 9EW, England

ABSTRACT. The central problem in Quasielastic Neutron Spectroscopy (QNS) is the recovery of the scattering function $S(Q,\omega)$ or the recovery of the intermediate scattering function $\tilde{S}(Q,t)$. Either of these two functions characterize the dynamics of the target under investigation. Time-of-Flight (TOF) spectroscopy aims at retrieving $S(Q,\omega)$ by performing a deconvolution involving the Point Spread Function (PSF) of the TOF instrument. The current TOF data analysis involves Least Squares Fitting (LSF) of strongly nonlinear parameters (e.g. linewidths) pertaining to phenomenological models. Neutron Spin-Echo (NSE) spectroscopy consists essentially in measuring $\tilde{S}(Q,t)$ and subsequently Laplace transforming the intermediate scattering function to obtain a distribution of relaxation rates. This is a very ill-conditioned problem, for which LSF is known to yield very poor results. Now, in both the TOF and NSE cases, the data are expressed as linear forms of the sought scattering or distribution functions, for which the Maximum Entropy Method (MaxEnt) is known to yield a unique solution. This method was therefore used to analyze computer-simulated noisy data as well as real experimental data and it is shown to be quite successful for both TOF and NSE spectroscopies.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 163–173.

1. Introduction

Since the beginning of the seventies, the advent of neutron high flux reactors (e.g. the ILL in France, ISIS in the U.K.) as well as novel techniques involving polarized neutrons have pushed further the limits of investigation of condensed matter in such exciting fields as polymer and proteins dynamics, critical phenomena, glasses or structural phase transitions, to name a few. As a rule, the physical/chemical information is deduced from the so-called scattering function $S(Q,\omega)$ or its Fourier transform $\tilde{S}(Q,t)$, with respect to the energy-transfer ω, the so-called intermediate scattering function. None of these two quantities is directly measured. But, as opposed to the crystallographic phased case, the neutron cross-section, which is the observed quantity, relates linearly to the sought information, albeit in a complicated way. The data are always too scarce and too noisy. Most often, the current data treatments involve :

- a convolution with an instrumental point spread function.
- the use of phenomenological models, which depend strongly non-linearly on the parameters of interest : linewidths or relaxation rates. Consequently, the generally non unique extrema exhibit a very strong correlation of the fitted parameters.
- Moreover, in the case where a sum of exponentials or Lorentzians is sought, the problem becomes strongly ill-conditioned since the trial functions are highly non-orthogonal.

By contrast, the continuous development of a powerful method, MaxEnt, which maximizes the Shannon-Jaynes entropy and yields a unique solution for linear problems, has led to :

- efficient algorithms (Skilling and Gull, 1985)
- a wealth of applications and spectacular image restorations (Gull and Skilling, 1984).
- not to mention more theoretical justifications (Levine, 1986)
- and more technical improvements (Skilling, Gull : this workshop)

As a matter of fact, MaxEnt has already broken into elastic neutron scattering (e.g. Daniell, 1988 and Johnson, 1986). The aim of this paper is to show how MaxEnt can be applied to inelastic/quasielastic neutron scattering.

Before turning to these new applications, let us recall that :

- Data space consists of D_k noisy values (k = 1,M) of standard deviations σ_k.

- Image space consists of I_j values (j = 1,N) that we seek.

- D_k's relate to I_j's through $D_k = \sum_{j=1}^{N} R_{k,j} * I_j * m_j$,

an equality which holds only to a noise term of order σ_k.
$R_{k,j}$ is the discretized PSD in matrix form. We encode our prior knowledge through the m_j's, the sum of which is unity.

- We maximize the entropy $S = - \sum_{j=1}^{N} p_j * \log(p_j/m_j)$,

- where $p_j = \dfrac{m_j * I_j}{\sum_{j=1}^{N} m_j * I_j}$.

- This maximization is subject to the constraint $\chi^2 = M$,

- where $\chi^2 = \sum_{k=1}^{M} \left(\dfrac{D_k - \sum_{j=1}^{N} R_{k,j} * I_j * m_j}{\sigma_k} \right)^2$.

- MaxEnt is an iterative procedure, in which the initial guess is a flat image : $I_j = I_{default}$ = constant. We choose that value of $I_{default}$ which minimizes χ^2.

- Our ending criterion is : $\vec{\nabla}\chi^2$ is parallel to $\vec{\nabla}S$.

2. Neutron Spin-Echo (F. Mezei, 1980)

Neutron Spin-Echo (NSE) Spectrometry is an irreplaceable tool to observe minute energy-transfers (down to about 1 nev) and hence large relaxation rates (up to 500 ns) in condensed matter. For instance, in the case of polymers or proteins, Quasielastic Light Scattering (QLS) often cannot be used since a too restricted scattering vector range and/or lack of contrast forbid it. Both techniques measure a counting rate at a detector as a function of time. MaxEnt has already proven to be essential (Livesey et al, 1987, Livesey, this workshop) to analyze QLS data. Can NSE, which does not afford hundreds of data points but at most a few tens on the one hand, and which has a much smaller

signal/noise ratio (> 1000 for QLS, < 100 for NSE), benefit from the use of MaxEnt ? Paragraphs a) and b) will show below that this is indeed the case.

a) Mathematically, the general inverse problem one has to solve in NSE in order to retrieve the relaxation rate distribution $h(Q,\tau)$ is described by :

$$\mathcal{D}(\theta,t) = \int_{\text{all neutron paths}} d\theta_o . f(\theta,\theta_o) \int d\lambda . g(\lambda) . \tilde{S}(Q,t)$$

or

$$\mathcal{D}(\theta,t) = \int d\theta_o . f(\theta,\theta_o) \int d\lambda . g(\lambda) \int d\tau . e^{-t/\tau} . h(Q,\tau)$$

introducing the relaxation rate distribution function $h(Q,\tau)$. In the above formulae,

- 2θ is the detector angle and $2\theta_o$ the scattering angle. The physical origin of the spread function $f(\theta,\theta_o)$ is the finite collimation of the neutron beam. It is most often neglected in standard NSE analysis.
- λ is the incident wavelength and $g(\lambda)$ describes the wavelength distribution. The latter is also often neglected.
- Q is the scattering vector, which is very well approximated by :

$$Q = 4\pi \frac{\sin\theta_o}{\lambda} \qquad \text{in NSE spectroscopy.}$$

The problem of inverting the Laplace transform to obtain time rates has long been known to require a non-flat prior in time-rate space : Jeffreys's prior is called for (Jaynes, 1968, Livesey et al, 1987, and references therein). Using Jeffreys' prior amounts instead to looking for the image in the log time-rate space where our prior knowledge is flat. The underlying idea for that is that total ignorance a priori regarding the relaxation rates should not depend on the time scale which is used. A straightforward consequence is that, in order to have a PSF matrix as well conditioned as possible, one should measure equidistant data points in a logarithmic scale as well.

Finally, let us mention that, because we cannot measure as many data points as in QLS and in particular, because we cannot measure at experimental times large enough for all relaxation components to have died out, our goal is more restricted from the start, namely : i) can we satisfactorily retrieve a portion of the distribution (that seen by the time-window of the spectrometer) ? and

ii) is MaxEnt powerful enough to separate two neighboring time contributions ? The answer is positive indeed.

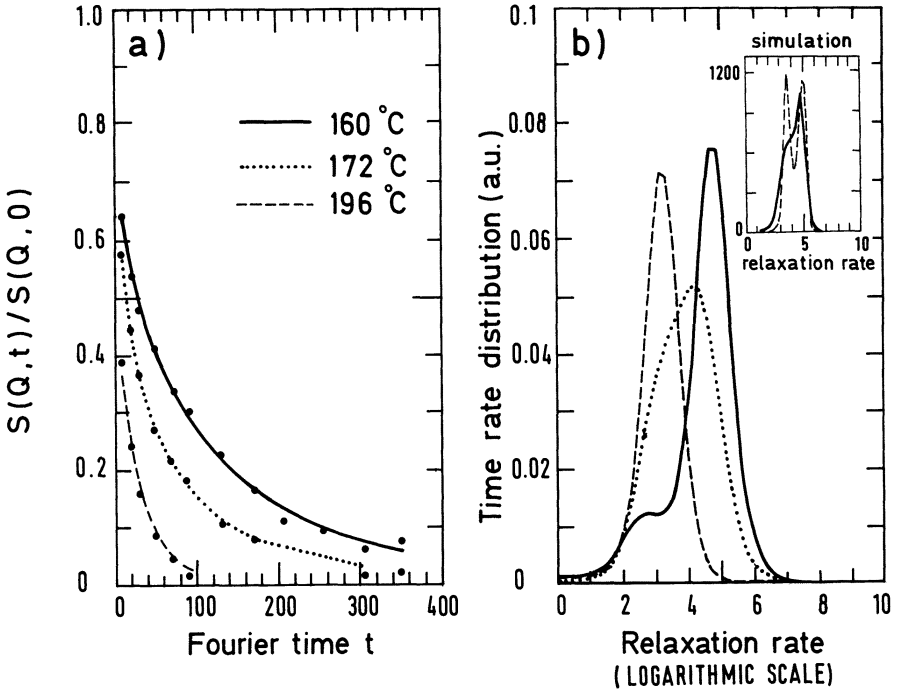


Figure 1. a) Measured NSE spectra (Mezei et al., 1987) evidencing the slowing down of the dynamics of a glass forming system as a function of temperature. (• : data points ; ---, ··· and — : MaxEnt fits).

b) The 3 MaxEnt reconstructions corresponding to the 3 data curves of a). Insert : Numerical simulation using the same noise level and data time range as in a). It shows that increasing the number of points from 10 (typical spectrum) to 30 can help to resolve a multimodal time rate distribution. (— : using 10 data points, --- : using 30 data points).

First, many simulations with noise were made in order to make sure that a single rate could be properly recovered. These simulations were run in conditions close to the real experimental ones, especially as regards the number of data points (about 10) and the dynamical data time range. The latter is defined as the ratio of t_{max}/t_{min} and is generally of the order of 50 to 100. The image is defined using 51 pixels. Provided that i) the sought rate was included between the minimum and maximum data times and ii) that the ratio of the simulated dynamical data intensity range was larger than $e \simeq 2.718$, the recovery was very good. Conditions i) and ii) merely express that the time window is properly set to observe the simulated phenomenon.

A real example (data from Mezei et al., 1987) is shown in fig.1. It can be seen from fig.1b that MaxEnt reconstruction corresponding to 172°C seems to be bimodal. In order to check that it could have been possible experimentally to resolve the distribution better, we ran a simulation in the same conditions (insert of fig.1b). Using 10 data points, we obtained the solid line, whereas using 30 data points with the same dynamic range, we obtained a nice two-peak distribution. This shows that if the physics had involved 2 peaks, these could have been separated by measuring 30 data points and using MaxEnt.

b) A second fundamental quantity to be determined experimentally is the wavelength distribution $g(\lambda)$ of the NSE spectrometer. Besides the fact that it can easily be taken into account using MaxEnt as in a), its precise knowledge is necessary to see if the instrument is properly set and/or to determine Q and data time values with enough accuracy.

It can be shown (Mezei, 1980) that we can easily measure directly its Cosine Fourier transform. Moreover, we are looking for a positive distribution. This is an another case for which MaxEnt is best suited : no need for equi-spaced data points and no need to complement the data points by fake points before Fourier transforming back a noisy data set. Let us just mention an extra technicality involved, autocalibration (Gull and Skilling, 1984), which does not pose any problem in our case.

A real data example is given in fig.2. The recovered image is almost noise-free. In order to check that we could believe in this result, numerical simulations were also run, yielding an excellent agreement between the original images and MaxEnt reconstructions.

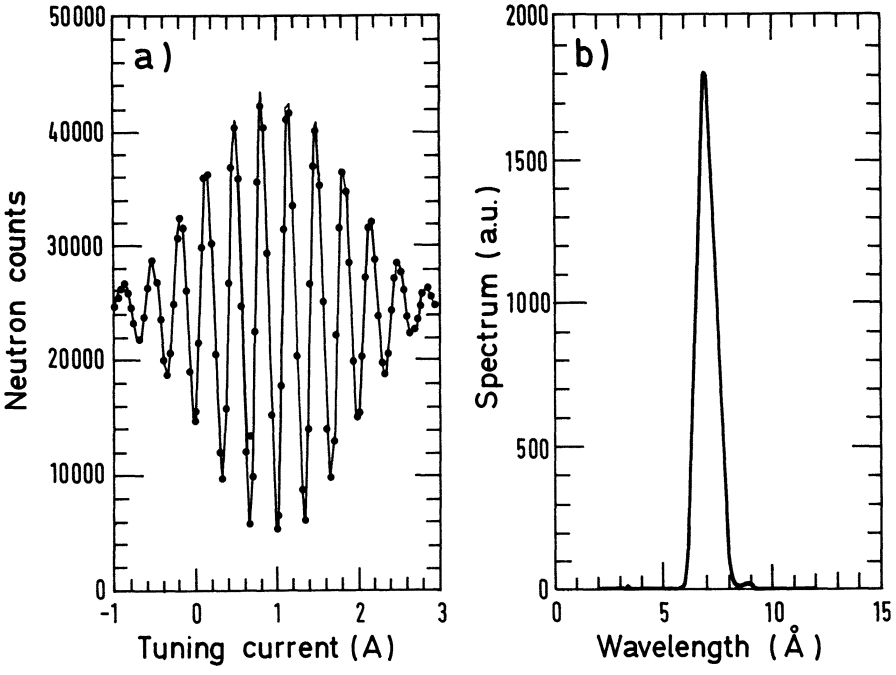


Figure 2. a) Measured neutron spin-echo group.
b) The corresponding MaxEnt wavelength distribution.

3. Time-of-Flight neutron spectroscopy (Lechner, 1984)

Here, the very general problem involved is that of separating an elastic (or Bragg) contribution from an inelastic or quasielastic one. The image space is the neutron energy-transfer ω-space and the data space is the time of arrival t-space (of the neutrons at the detector). The sought image consists of a spike plus a well-behaved bell-shaped distribution. The difficulty here is that, while the sought image is discontinuous, MaxEnt provides us with the smoothest

image compatible with the data. Skilling addressed this problem in 1986 and Skilling and Gull in 1987. Here, we use a numerical simulation of the data and image to illustrate the pitfalls and describe our preferred procedure.

Let the simulated image consist of N = 301 pixels. It is the sum of a broad line centered in pixel 151 and a spike located in the same pixel. The latter corresponds to a zero energy-transfer. Numerically,

$$I(\omega) = 20000*\delta(\omega) + 30000*\left[\frac{25/\pi}{\omega^2+625}\right]$$

The TOF data space is divided into channels of known width and central time values. The image is convolved with the PSF, which can be represented by :

$$R(n,\omega) = \left[1 + \frac{\hbar\omega}{E_o}\right]^{1/2} * \exp\left[-\frac{1}{2}\left\{\frac{n-n(\omega)}{\sigma(\omega)}\right\}^2\right] \Big/ (\sigma(\omega)*\sqrt{2\pi})$$

where n applies to the n-th TOF channel, $n(\omega)$ and $\sigma(\omega)$ are known well behaved functions of instrumental parameters and E_o is the average incident neutron energy of a suitably monochromatized neutron beam. Finally, a gaussian random noise is added, resulting in fig.3a. A first run of MaxEnt using no prior at all yielded the solid curve in fig.3b, pointing out the location of the spike, but clearly not the ideal original image. As a consequence, prior knowledge should be used. Note, in passing, that the left side of the reconstruction starts with a constant equal to the default level : no simulated data point provides information over the related energy-transfer range ! The following prior was then used (Skilling, 1986) : the pixel containing the spike is so many times more intense than the average of the remaining pixels. Applying this criterion to define our m_j's, our reconstruction, evidencing some huge ringing about the spike, is shown in fig.3c. Something must have gone wrong ! Still keeping in mind a uniform prior except for one pixel, we then added the following requirement : CONTINUITY. Once divided by the prior, the resulting image (the I_j's) must be continuous. Mathematically, the only relevant parameter is P, the fraction of the spike intensity in the related pixel. After some elementary algebra, one obtains :

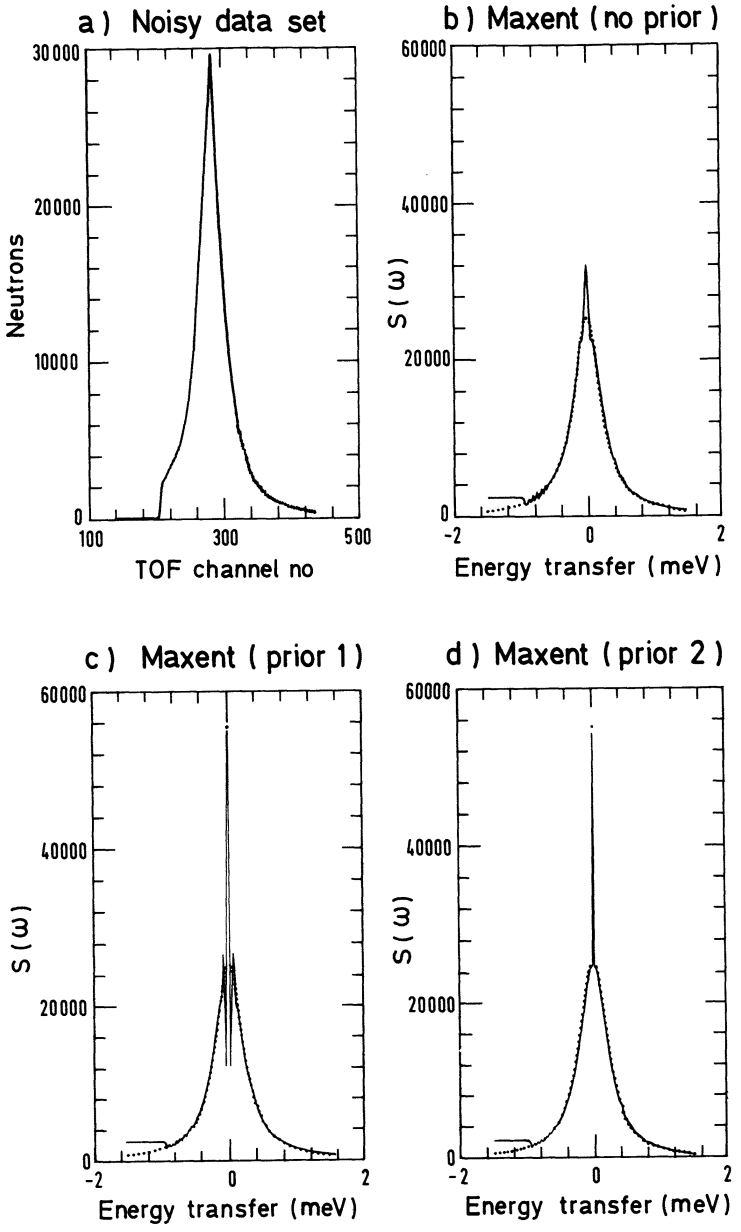


Figure 3. a) Simulated time-of-flight data spectrum. The PSF corresponds to that of the Saclay TOF instrument at 7.07 angstroms.
In b), c) and d) the dotted line corresponds to the original image. The solid line represents the MaxEnt reconstruction.

$$m_{spike} = 1/(P + N*(1-P))$$
$$m_{else} = (1-P)/(P + N*(1-P))$$

Using our new criterion, one finds P = .5416171 as opposed to P = .8939298 using Skilling's. The reconstruction is improved indeed, as can be seen from fig.3d. Finally, note that in Skilling's paper (1986, p.176, fig.15), BOTH criteria are obeyed, which can never be the case in Quasielastic Neutron Scattering since the spike sits on top of the maximum of the quasielastic line !

4. Conclusion

In this paper, we have demonstrated the ability of MaxEnt to recover incident wavelength distributions and to separate two relaxation rates in NSE spectrocopy, as well as to recover the proper lineshapes and linewidths without making any assumptions using an a priori model in TOF spectroscopy.

Moreover, Quasielastic Neutron Scattering provides MaxEnt with straightforward examples illustrating typical inverse problems (Fourier, Laplace) as well as the importance of using Prior Knowledge.

Acknowledgments

The authors are grateful to J. Skilling and S. Gull for the use of the Cambridge code. One of us (R.J.P.) wishes to thank M. Lambert and G. Jannink for constant encouragements and support. He also thanks F. Mezei and B. Farago for his use of some of their experimental data.

References

Daniell, G.J. (1988) "The Maximum Entropy Technique Applied in Neutron Scattering", *Neutron Scattering at a Pulse Source*, R.J. Newport, B.D. Rainford and R. Cywinski, eds., Adam Hilger, pp. 144-154.

Gull, S.F. and J. Skilling (1984), "The Maximum Entropy Method in Image Processing", Proc. IEE **131F**, pp. 646-659.

Jaynes, E.T. (1968), "Prior Probabilities", IEEE Trans., SSC-4 pp. 227-241.

Johnson, M.W. ed. (1986), *Neutron Scattering Data Analysis*, Institute of Physics Conference Series 81.

Lechner, R.E. (1984), "TOF-TOF Spectrometers at Pulsed Neutrons Sources and at Steady-State Reactors", *Proc. Workshop on Neutron Scattering for SNQ, Maria Laach*, R. Scherm and H. Stiller, eds. Jül-1954, pp.202-223

Levine R.D. (1986), "Theory and Practice of Maximum Entropy Formalism", *Maximum Entropy and Bayesian Methods in Applied Statistics*, J.H. Justice, ed., pp.59-84.

Livesey, A.K., M. Delaye, P. Licinio and J.C. Brochon (1987), "Maximun Entropy Analysis of Dynamics Parameters via the Laplace Transform", Faraday Discuss. Chem. Soc. **83**, pp. 247-258.

Mezei, F., ed. (1980), *Neutron Spin Echo*, Lecture Notes in Physics 128, Springer-Verlag.

Mezei, F., W. Knaak and B. Farago (1987), "Neutron Spin-Echo Study of Dynamics Correlations near the Liquid-Glass Transition", Phys. Rev. Lett. **58**, pp.571-574.

Skilling, J. and S.F. Gull (1985), "Algorithms and Applications", *Maximum Entropy Methods in Inverse Problems*, C. Ray Smith and W.T. Grandy, Jr., eds., D. Reidel, pp.83-132.

Skilling, J. (1986), "Theory of Maximum Entropy Reconstruction", *Maximum Entropy and Bayesian Methods in Applied Statistics*, J.H. Justice, ed., pp.156-178.

Skilling, J. and S.F. Gull (1987), "Prior knowledge must be used", *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, C. Ray Smith and G.J. Erickson, eds., D. Reidel, pp.161-172.

Maximum Entropy Reconstruction in Magnetic Resonance Imaging

R.T. Constable, R.M. Henkelman
Department of Medical Biophysics
University of Toronto
500 Sherbourne Street, Toronto
Canada, M4X 1K9

Abstract: The maximum entropy method of reconstruction is applied to magnetic resonance images. The results indicate that MEM is not a good measure of image quality and that the maximum entropy image is not neccesarily the image we desire.

In conventional Magnetic Resonance (MR) Imaging the data is collected in the time domain and a two-dimensional Fourier transform is used to convert the time domain signal into its spatial frequency representation in the formation of an image. However, two problems exist with this method of image reconstruction. First, it assumes that the data is band limited, in the sense that the highest frequency of the sampled signal is less than half the sampling frequency. This assumption is invalid for human anatomy and as a result truncation artifacts arise in images reconstructed using the Fourier transform. These artifacts can be important in the clinical setting where they may be mistaken for pathology[1]. Furthermore, truncation artifacts increase in severity with reduction in the size of the data set collected precluding reduction in imaging time by aquiring less data. These artifacts also confound 3D volume reconstructions where imaging time further restricts the amount of data that may be collected.

Secondly, the Fourier transform makes no distinction between noise and valid signal. Thus attempts at reducing high frequency noise in an image through filtering in either the data or image space lead to an unsatisfactory loss of resolution in the final image.

With these inadequacies in the Fourier transform, new methods of reconstruction have been considered[2,3]. The application of the Maximum Entropy Method (MEM) to MR imaging data is the subject of this work. To date, a detailed analysis of the behaviour of the entropy regularizer on MR imaging data has not been presented. In this paper, characteristics of the method and the basic problems encountered implementing the maximum entropy technique in MR imaging are presented. A counter example is shown demonstrating that the maximum entropy solution is not necessarily representative of the image we desire.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 175–180.

Theory:

MEM has found considerable success in many diverse fields such as radio-astronomy[4], NMR spectroscopy[5], x-ray diffraction analysis[6] and geophysics[7]. In these fields MEM has increased the conspicuity of signals and suppressed background noise making interpretation of spectra and images a much simpler task. With this success in mind MEM was investigated as a possible approach to the reconstruction problem in MR imaging.

Justification of the MEM stems from both a probability theory approach, as in Jaynes[7], and from an information theory approach; see Ulrych[8] and Shannon[9]. MEM applied without any constraints to model the measured data, yields a completely flat image with no information. As constraints are applied, the solution moves away from that of a totally flat image and begins to gain information with a resultant decrease in entropy. Thus any information in the image arises directly from the data and the final image will be as smooth, in the multiplicity sense, as allowed by the data.

The application of MEM to the reconstruction of MR images proceeds by maximizing the entropy in image space subject to a χ^2 fit to the data[10]. That is, maximize

$$Q(\lambda) = - \sum_{j=1}^{N} m_j \log m_j - \lambda \chi^2$$

$$\text{where,} \quad \chi^2 = \sum_{k=1}^{M} (M_k - E_k)^2 / \sigma^2$$

and, m_j represents the intensity of a pixel in the proposed image, M_k the Fourier transform of m_j and E_k is the actual data measured in the MR experiment. λ is a Lagrangian multiplier and σ is the standard deviation of the noise. The actual solution to this maximization problem was found using the algorithm of Skilling and Bryan[11]. The solution is reached when χ^2 equals the number of data points.

Results:

Other author's[12,13] have shown that MEM decreases the background noise and leaves noise on top of the image unaltered. Figures (1) and (2) show 1-D reconstructions of a series of top-hat functions from a noisy time domain data set. Figure (1) is a magnitude image reconstructed using the conventional FT method while Figure (2) is the MEM solution from the same data set. Note that in the MEM reconstruction the background noise has been reduced significantly but the noise on the top-hat functions remains essentially unaltered. Figures (3) and (4) show FT and MEM reconstructions of a coronal MR brain image. Intensity differences are seen in the MEM image as a result of line by line processing. Noise measurements indicate that the background noise has been reduced in the MEM image while the noise within the high intensity regions of the brain remains unaltered.

Such an 'improvement' in the image makes no contribution to MR imaging because the diagnostic value of the image is not increased. This has been the major complaint with the technique and is the reason for the methods failure in MR imaging. It has not been clear until now however, whether this problem was due to improper application of the technique or if it is in fact intrinsic to the method.

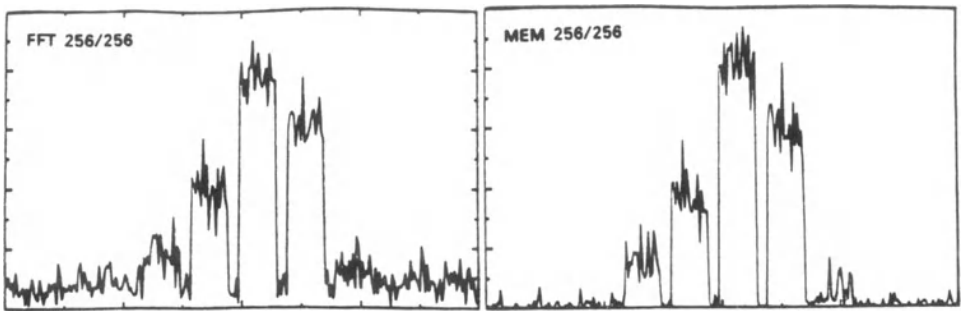


Figure 1. A series of noisy top-hat functions reconstructed using the conventional Fourier tranform method. Figure 2. The same series of top-hat functions reconstructed using MEM.

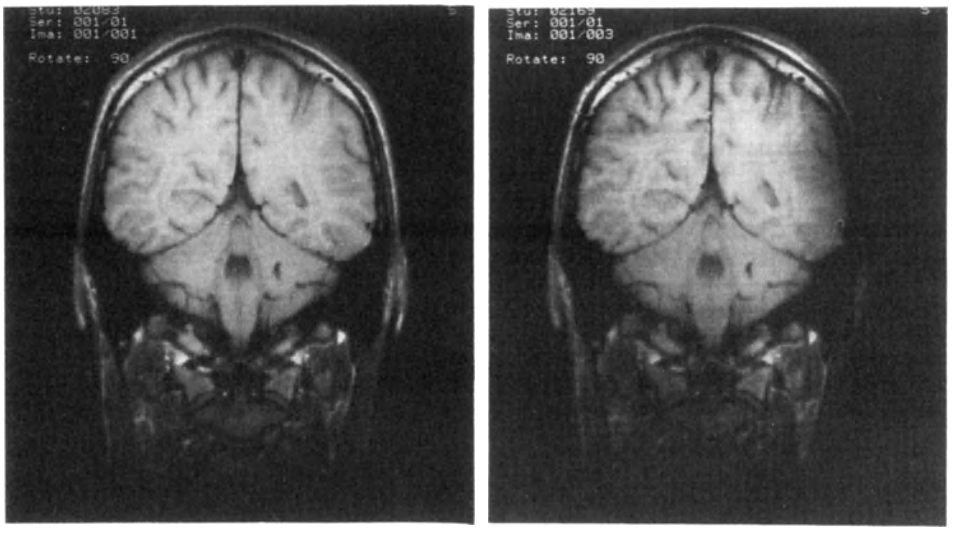


Figure 3. Magnitude image of the brain reconstructed using the conventional FT method of reconstruction. Figure 4. Same image reconstructed using MEM.

Discussion:

That MEM flattens the background can be attributed to two factors. Firstly, the shape of the entropy curve favours changing the lower amplitude background pixels as more entropy is gained per unit change in these pixels than in the high intensity pixels. This would imply

that preferential weighting is given to the lower amplitude values. In addition, some author's[13] have applied MEM (incorrectly) without the constraint that the sum of the probabilities must be 1. If this constraint is ignored, the shape of the entropy curve for intensities much greater than one is essentially linear with a negative slope. The entropy regularizer in this case will simply try to pull all pixel values down to lower amplitudes in an effort to increase the entropy. However, the entropy criteria necessitates the reconstruction of positive images only. Therefore, any background components which are made negative through the entropy regularizer must be set to some default positive value. If enough of these points are flipped up, the background becomes completely flat with no visible structure.

To understand why the noise on top of the signals remains unchanged one must consider how the entropy term influences the χ^2 fit to the data. The entropy term biases the χ^2 fit in such a way as to gain the maximum increase in entropy. Little gain in entropy is obtained by flattening out the high frequency noise on top of the signals. Therefore, the high frequency noise is fit closely and the freedom in the χ^2 goes into simply compressing the dynamic range of the image.

However, the noise in MR is distributed uniformly across all frequencies[15]. Therefore any fitting criteria, such as χ^2, should not allow some data components to be fit exactly while putting more freedom into other components. Fitting ordered residuals could be applied as an alternative constraint although this would not lead to a significant change in the final answer as the residuals in the biased solution are in fact very close to gaussianly distributed. In addition, a typical MR imaging experiment collects only 128 or 256 data points per line; this leads to poor statistics in fitting residuals. Furthermore, applying constraints to counteract the properties of the applied regularizer does not address the real problem.

The difficulty lies with the entropy regularizer itself. The entropy expression contains no neighbour to neighbour information. Thus alternating high frequency fluctuations are quite acceptable to the entropy regularizer. Figure (5a) shows a perfect profile of a series of top hat functions with a total entropy measure of S = 1.95. This represents the ideal reconstruction. Figure (5b) meanwhile shows another series of top hat functions with high frequency fluctuations on top of the signals and a slightly elevated baseline with an entropy of S = 1.96, representative in part, of the noisy images we wish to improve. Both profiles have been normalized to the same intensity. From this example it is apparent that entropy is not a good measure of image quality and that the maximum entropy image is not necessarily the image we desire. Furthermore, this example also indicates that MEM cannot be expected to flatten the noise on top of the signals as this does not lead to the maximum entropy solution. This example does not suggest that using the data of Figure (5a) will yield a maximum entropy solution of the form of Figure (5b); it will not. What it does demonstrate is, if there is evidence for those high frequency fluctuations in the data, MEM cannot be asked to remove them as the

desired solution has a lower entropy than the unacceptable image we wish to improve.

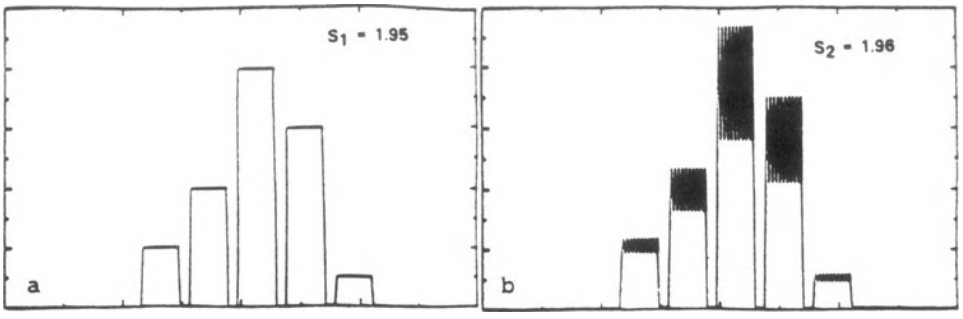


Figure 5a. Perfect profile of 5 top-hat functions. Figure 5b. Counter example with high frequency noise present and a higher entropy than in Figure 5a. This example demonstrates that entropy is not a good measure of image quality.

Conclusions:

MEM works well if the reconstruction is required to have distinct peaks with a flat baseline as in spectroscopy[14]. While the justification of MEM given in section 2.1 is sound for statistical processes, difficulties arise when noise is introduced and constraints must be applied to the data. The point has often been made that the entropy favours smooth images since the global (unconstrained) maximum is given by the uniform image. The final image, when constraints are applied, is in fact smooth in the multiplicity sense, but since no neighbour to neighbour information is contained in the entropy term, high frequency fluctuations may remain. It is therefore misleading to say that the reconstruction is as smooth, in the conventional sense, or featureless as possible. Finally we may conclude that MEM as applied above is inappropriate for use in MR imaging as the maximum entropy solution does not select the most desireable image for medical interpretation.

References:

1. Henkelman R.M., Bronskill M.J., Artifacts in Magnetic Resonance Imaging. Reviews of Magn. Reson. in Med.,2(1), 1-126, 1987.

2. Smith M.R., Nichols S.T., Henkelman R.M., Wood M.L., Application of Autoregressive Modelling in Magnetic Resonance Imaging to Remove Noise and Truncation Artifacts. Magn. Reson. Imaging, vol.4,257-261, 1986.

3. Haacke E.M., Liang Z.P., High Resolution, Limited View Reconstructions. 1987 Topical Conference on Fast Magn. Reson. Imaging Techniques, Cleveland, Ohio, 1987.

4. Bryan R.K., Skilling J., Deconvolution by Maximum Entropy as Illustrated by Application to the Jet of M87. Mon. Not. Rad. Astron. Soc., 191, 69-79,1980.

5. Sibusiso S., Skilling J., Brereton R.G., Laue E.D., Stauton J., Maximum Entropy Signal Processing in Practical NMR Spectroscopy. Nature, 311(4), 446-447, 1984.

6. Collins D.M., Electron Density Images From Imperfect Data by Iterative Entropy Maximization. Nature, 298, 49-51, 1982.

7. Jaynes E.T., Prior Probabilities. IEEE Trans. Systems,Science and Cybernetics, ssc-4(3), 227-241, 1968.

8. Ulrych T.J., Bishop T.N., Maximum Entropy Spectral Analysis and Autoregressive Decomposition. Reviews of Geophysics and Space Physics, 131(1), 183-200, 1975.

9. Shannon C.E., A Mathematical Theory of Communication. Bell Syst. Tech. J., 27, 379-423, 1948.

10. Gull S.F., Daniell G.J., Image Reconstruction From Incomplete and Noisy Data. Nature vol.272, April 1978.

11. Skilling J., Bryan R.K., Maximum Entropy Image Reconstruction: General Algorithm. Mon. Not. R. Astron. Soc., 211, 111-124,1984.

12. Chapman B., Mansfield P., Doyle M., Enhancement of Echo-Planar Images by the Maximum Entropy Method. Proc. SMRM 5^{th} Annual Meeting, vol. , 261-262, 1986.

13. De Simone B.C., DeLuca F., Maraviglia B. Maximum Entropy in Phase-Encoding NMR Imaging. Magn. Reson. Med., 4, 78-82, 1987.

14. Martin J.F., The Maximum Entropy Method in NMR. J. Magn. Reson., 65, 291-297, 1985.

15. McVeigh E., Henkelman R.M., Bronskill M.J., Noise and Filtration in MRI. Med. Phys., 12(5), 586-591, 1985.

SOLUTION OF AUTOCORRELATION FUNCTION CONSTRAINED MAXIMUM ENTROPY PROBLEMS USING THE METHOD OF SIMULATED ANNEALING

N.A. FARROW AND F.P. OTTENSMEYER
Department of Medical Biophysics,
University of Toronto,
500 Sherbourne Street, Toronto,
Ontario, M4X 1K9 Canada

ABSTRACT. An algorithm is presented that constrains the autocorrelation function of noise removed when a maximum entropy model of noisy data is selected. The algorithm uses the method of simulated annealing and is applied to a model data set. The algorithm is successful in finding solutions to both unconstrained and constrained maximum entropy problems.

INTRODUCTION

Dark field electron microscope images are often corrupted by noise, the main source of which is the random scatter of electrons by the near random structure of the carbon film used to support the biological sample. This noise often obscures structures of interest within the micrograph. Various image processing techniques have been used in the past in an attempt to try and remove the noise from these images but many of the conventional methods, e.g. averaging or low pass filtering, inevitably lead to a drop in the spatial resolution of the images. This inherent loss of resolution led us to consider the application of maximum entropy methods to the problem of noise removal. We have reported earlier (1,2) preliminary results with these techniques. Concern about the distribution of the residuals, differences between the data and model solution, led us to use an error fitting statistic (3), E^2, rather than the more conventional Chi squared. E^2 is a measure of the similarity between the residuals u, the differences between the data and "models" produced by the algorithm in the search for the maximum entropy solution, and the distribution of the noise corrupting the data ν. E^2 is defined,

$$E^2 = \sum_{i=1}^{n} (u_{(i)}-\nu_{(i)})^2 \qquad (1)$$

where the residuals between the data set $D=\{d(i):i=1,n\}$ and the model $M=\{m(i):i=1,n\}$ are defined,

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 181–189.

$$u(i) = (d(i)-m(i))/\sigma \qquad (2)$$

and in the definition of E^2 $N=\{\nu(i):i=1,n\}$ is the sample noise distribution. σ is a measure of the standard deviation of the noise present in the data. The subscripted parentheses in the definition of E^2 indicate that the residual distribution and the noise distribution are compared after ordering, e.g. the largest experimental residual is compared to the largest expected noise value in a sample of similar size and variance.

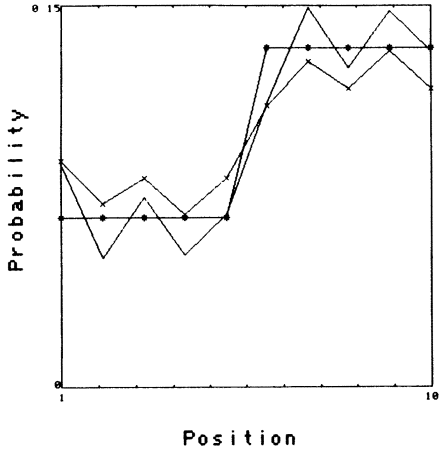


Figure 1. Illustration of the maximum entropy combination of residuals and data. The original data -------, the maximum entropy solution ---+--- and the step function that we require the entropy maximisation algorithm to return ---*---.

The method by which we adapted and applied the concept of error fitting to the electron microscopy problem has been described elsewhere (1). The method of entropy maximisation using an E^2 constraint was found to produce an amplitude distribution of residuals in agreement with that of the noise corrupting the data but lack of control of the position of the residuals led the algorithm to introduce an artefact into the final solution (2). The artefact results from the tendency of the maximum entropy method to produce solutions in which the points in the processed image all move towards the mean of the data. It was found that a maximum entropy solution was always achieved when all data values above the mean were brought down towards it and all those below the mean were adjusted upwards towards it. The distribution of residuals was obviously strongly affected by the gross features in the data. Had some smaller section of the data set been considered then a different solution would have resulted, not a desirable property for this type of processing algorithm. Indeed if the algorithm is applied to a noisy but ergotic data set then the results of the error fitting maximum entropy process

were much more satisfactory.

An example of the effect of error fitting and entropy maximisation on non-ergotic data is shown in figure 1. To generate the figure a 10 point data set was used. The original data consisted of a step function to which was added noise with a known amplitude distribution. Model distributions were then created by the addition of the same noise distribution to the data in an attempt to return the original step function. As the noise distribution is symmetric, one combination of the noise and the data will produce the step function. Dealing with as few as 10 points we were able to examine all the possible solutions that resulted from combining the known residuals with the data. From the resulting models we were able to select the one that had the maximum entropy, S, defined,

$$S = - \sum_{i=1}^{n} (d(i)+u(i))\log(d(i)+u(i)) \qquad (3)$$

The artefact described above can clearly be seen in figure 1.

The noise removed from the data is found to have an autocorrelation function similar in form to that of the step function rather than the noise that was added to the step function. The autocorrelation function, ACF, of a function $F=\{f(i):i=1,n\}$ was defined,

$$ACF(r) = \frac{1}{n-r} \sum_{i=1}^{n-r} f(i)f(i+r) \qquad (4)$$

where r is the lag and the autocorrelation function is defined for r from 1 to n-1.In figure 2(a), (b) and (c) the normalised autocorrelation functions of the step function, the noise added to the step function, and the residuals removed from the noisy data to give the final model are plotted. Ideally the distribution in figure 2(c) should be the same as that in figure 2(b) indicating that the autocorrelation of the noise removed was the same as that added. In fact, the distribution of figure 2(c) has a stronger similarity to that of figure 2(a) the autocorrelation function of the step function.

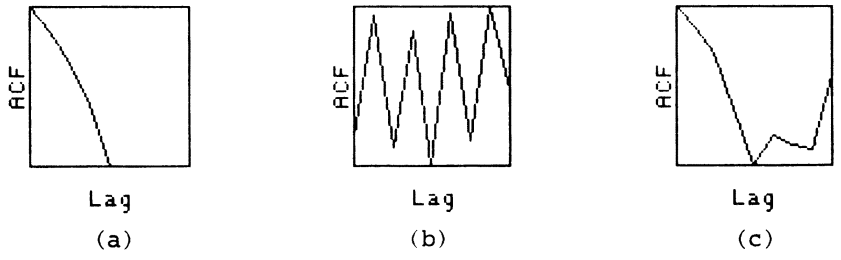


Figure 2. The normalised autocorrelation function of (a) the step function without noise, (b) the noise added to the step function to produce the data set, (c)the noise removed from the data set to produce

the maximum entropy solution.

To ensure that the spatial distribution of the residuals removed by the maximum entropy method was the same as that present in the noise corrupting the true signal, we compared their autocorrelation functions. The autocorrelation function as defined in equation 4 was calculated for both the noise added to the step function, ACF^{+}, and the residuals between the data and the model, ACF^{-}. These two functions were then compared using the following measure, k, of the differences between the two functions squared.

$$k = \sum_{r=1}^{n-1} (ACF^{+}(r) - ACF^{-}(r))^2 \qquad (5)$$

We can now state what is demanded of the maximum entropy procedure when it is considered applied to the combination of known noise residuals and a data set corrupted by the same residuals. The maximum entropy process must not select the combination with maximum entropy resulting from all the possible combinations but from only those that have an autocorrelation function of their residuals not significantly different to that of the noise present in the data. In other words we want a process yielding a model with high entropy S, and a low autocorrelation dissimilarity k.

In figure 3 we have plotted the distribution of k between the the autocorrelation function of the noise added to the step function and all the possible autocorrelation functions that could be generated by reordering the residuals. The position of the maximum entropy solution's value of k is 3.43×10^{-3}. At this value of k random configurations of the residuals would yield autocorrelation functions more dissimilar to that of the maximum entropy solution 28% of the time. This degree of dissimilarity is not extremely significant. However the chances of getting a combination with only a single zero crossing (change from positive to negative residuals) as this configuration has is only 0.4%. It would not be unreasonable to conclude that the step function form of the data had affected the spatial distribution of residuals.

THE METHOD OF SIMULATED ANNEALING

Maximisation of entropy in the manner described for the generation of figure 1 becomes completely infeasible for data sets with more than 10 points. The entropy maximisation problem considered in this way is not a maximisation in an n dimensional space of continuously varying parameters but rather an n dimensional configurational space, where the configurations are the combinations of different noise residuals with data. The number of different combinations possible within this space is n!. It quickly becomes impossible to explore them all. To tackle this problem we have applied the method of simulated annealing (4,5) in a manner similar to that used to solve the most famous of the set of combinatorial problems, that of the "travelling salesman" (4). In the travelling salesman problem one has to decide upon a path between n

cities which minimises the total distance travelled but passes through each city only once and returns to its starting point.

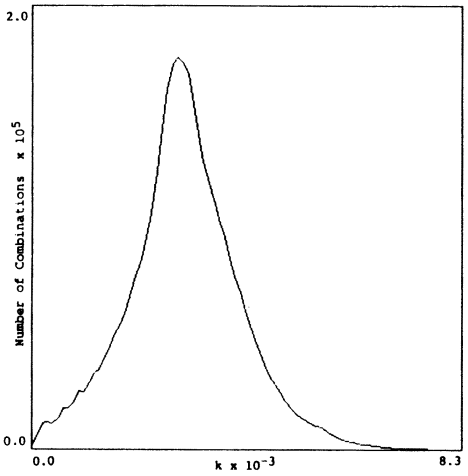


Figure 3. Distribution of the parameter k comparing noise distributions for all possible combinations of residuals to that of the original noise.

The method is called annealing because of the parallels between the way a final configuration is settled upon and the way a slowly cooled melt will find a low energy crystal configuration. When molten, the elements in a melt are able to move freely throughout the body of the material due to the high temperatures. As the temperature begins to drop crystal structures of lower energy will begin to form. The probability that an element of the system is in a given state with energy E is given by the Boltzmann distribution $P(E) \alpha \exp(-E/kT)$. Thus as a system cools there is a possibility that the energy of a part of the system may increase. As the temperature falls the chances of moving to a higher energy state become smaller. Because the chance of being in a higher energy state does exist, the system is able to move out of local minima configurations and eventually form a stable low energy crystal structure throughout the entire material. In contrast rapid cooling, quenching, leads to less ordered states in which crystal structures are often locally minimum. To employ the method of simulated annealing to solve the travelling salesman problem the configuration of the cities on the path was equated to the state of the crystal lattice and the distance between the cities to the energy of the crystal configuration. To use the method to maximise entropy we will equate the the configuration of the residuals to be removed with the crystal lattice state and the entropy of the resulting model to the energy of the structure. Obviously we will be maximising the entropy rather than minimising the energy.

The algorithm to find the maximum entropy combination of data and removed noise now proceeds in the following manner. The data is

considered to be a fixed array $D=\{d(i):i=1,n\}$, the residuals to be removed are in an array of the same size $U=\{u(i):i=1,n\}$. The order of the residuals in the array may be altered at random, e.g. element i and element i+3 may be interchanged. The algorithm considers two types of rearrangements, randomly deciding between the two choices. One choice, reversal, involves the reversal in the order of a random length section of the array. The other choice, transport, involves the removal of a random length section of the array and its reinsertion into the array at another randomly chosen position. Having made one of these changes the entropy, S, of the new combination of U and D is calculated as in equation 1.

The entropy of the new configuration is compared to the previous entropy prior to the rearrangement, the difference between the two entropies is referred to as the "entropy cost", ΔS, of the rearrangement. It is at this point that the concept of a "temperature" must be introduced into the algorithm, this will allow the creation of a parallel to the Boltzmann distribution. The distribution used is given by,

$$P(\Delta S) \propto \exp(\Delta S/T) \tag{6}$$

If the entropy cost is positive, i.e. the entropy is higher after rearrangement, then the new configuration is always accepted. If the cost is negative the chance of acceptance of the new configuration is governed by the above distribution.

To begin, a sufficiently large value of T is employed so that all rearrangements are accepted, which is equivalent to the situation of a free flowing melt. The algorithm then searches for 10n successful changes in configuration. At the initial high temperature it will find 10n successes in 10n trial rearrangements. Once 10n changes have been made then the temperature is reduced by a factor of 0.9. As the temperature falls, more and more rearrangements will be rejected. If the algorithm has not found 10n successful rearrangements in 100n tries the temperature is reduced, this stage is equivalent to the beginnings of crystallisation. The process is deemed complete when no successes are found in 100n rearrangements.

Figure 4 shows the results of applying the annealing process to the same data as that shown in figure 1. The algorithm was able to find the same solution as that shown in figure 1 in 212 CPU seconds when run on a Microvax II computer. The algorithm examined 54485 combinations of data and residuals and made 6716 rearrangements.

INTRODUCTION OF THE AUTOCORRELATION FUNCTION INTO THE ANNEALING PROCESS

Introduction of the autocorrelation function into the cost function is achieved by redefining the entropy cost to include the measure k. The new cost is denoted by $\Delta S'$, and is defined,

$$\Delta S' = \Delta S - \lambda \Delta k \tag{7}$$

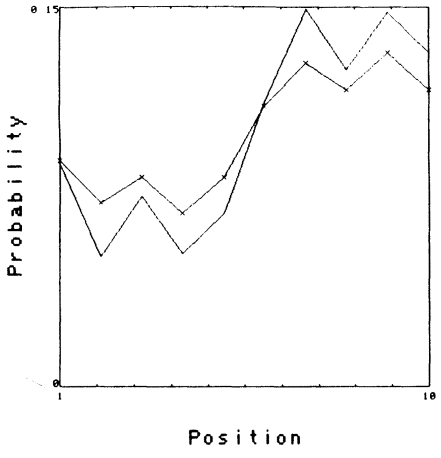


Figure 4. The maximum entropy solution selected by the method of simulated annealing.

where k is the difference between the two autocorrelation functions squared and Δk is the difference in k before and after the rearrangement of residuals. If λ is set to zero then the annealing process is equivalent to maximising entropy alone. As λ becomes larger then the similarity between the two autocorrelation functions becomes more important and the algorithm will tend to yield solutions with the required autocorrelation functions. However for a given "goodness of fit" between the autocorrelation functions, the algorithm should select the highest entropy configuration of residuals with such a fit.

To examine the effect of increasing λ we have plotted the entropy of the solution against its value of k. Figure 5 shows such a plot in which we again used the fact that for 10 data points we can examine all possible outcomes of combining the data with the noise residuals. The contours in figure 5 represent the number of combinations (using a log scale) which have a given value of entropy and k. It will be recognised that distribution, in its y axis, is related to the distribution depicted in figure 3. For λ of zero the annealing algorithm returns the unconstrained maximum entropy solution. As λ is increased from zero to 10^3 the solutions returned gradually decrease in entropy and their values of k also decrease. At λ of 10^3 the algorithm returns the noise free step function, the noise has been removed exactly and the value of k for this solution is zero. There are two possible solutions that have a k of zero, one is the step function, the second occurs when the residuals are added in exactly the same configuration as that used to generate the noisy data from the step function. The algorithm has correctly selected between these two functions and found the one with the higher entropy.

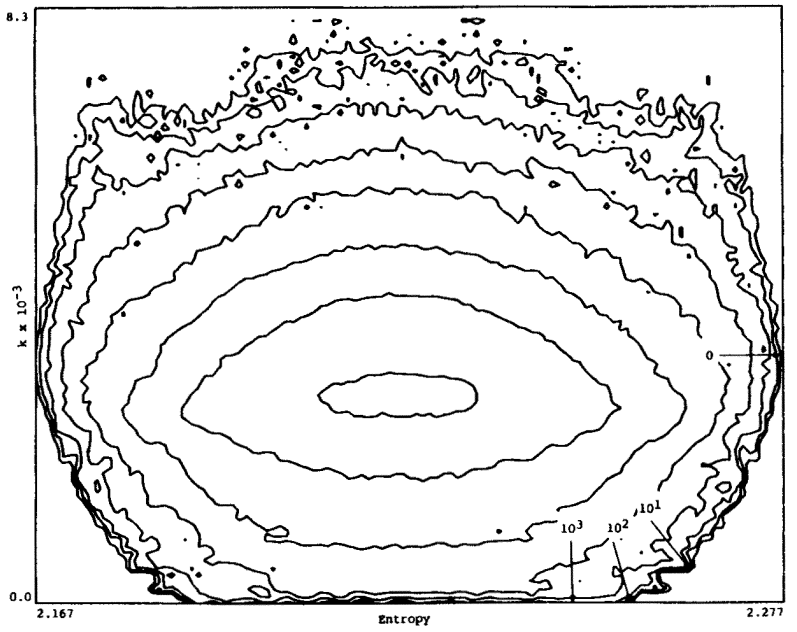


Figure 5. Distribution of entropy and the parameter k derived from all the possible combinations of residuals and data. The contours represent the log of the number of combinations with a particular value of entropy and k. Plotted on this distribution is the locus of the solutions selected by changing λ from 0 to 10^3.

CONCLUSIONS

In this paper we sought to demonstrate the use of the method of simulated annealing to the solution of noise autocorrelation function constrained maximum entropy problems. Whilst we have demonstrated its value in finding solutions for 10 data points the algorithm is still far too slow for problems with large numbers of data points. Calculation of the solution for a 256 point problem required 31 CPU hours on the Microvax II computer. To be useful in electron microscopy we will require manageable processing times for at least 10^3 data points.

The method of simulated annealing implemented as described above allows for the speed of solution to be increased. The increase in speed however will usually imply a less thorough annealing and a not necessarily optimal solution. If we 'cool' the system more rapidly a solution will be found more quickly but it may only be close to the optimum, for large data sets this will often be sufficient. We have found that beginning at very high 'temperatures' and 'cooling' slowly will often give an equivalent solution to that achieved when using a lower starting 'temperature' and rapid 'cooling'. Careful investigation of the effects of the parameters of 'temperature' and 'cooling' rate

will be required to prevent the algorithm running too fast and finding a non-optimal solution, or running unnecessarily slowly, yielding no improvement in solution.

Having shown that the method will work for model data we will now have to examine the implementation of similar algorithms to real data sets. Two remaining problems will be the determination of the value of λ and the algorithmic annealing parameters which provide the correct balance between the fitting of the autocorrelation function and the maximisation of the entropy of the model.

ACKNOWLEDGEMENTS

This work was supported by grants from the National Cancer Institute of Canada, the Medical Research Council of Canada and the Ontario Cancer Research and Treatment Foundation.

REFERENCES

1) Farrow N.A. and Ottensmeyer F.P.; Image and Signal Processing, Ed. Hawkes P.W., S.E.M. Inc. in print (1987)

2) Farrow N.A. and Ottensmeyer F.P., Maximum Entropy and Bayesian Spectral Analysis, Ed. Erickson G.J. and Smith C.R., Kluwer Academic Publishing. in print (1987)

3) Brian R.K. and Skilling J.; Mon. Not. R. Astr. Soc. (1980) **191**, 69-79

4) Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.D., 1983 Science, **220**, pp 671-680

5) Press W.H. et al. Numerical Recipes. Cambridge University Press (1987)

SOLUTION OF LAPLACE TRANSFORM EQUATIONS (SUM OF EXPONENTIALS) BY MAXIMUM ENTROPY.

A.K. Livesey[1], J-C. Brochon[2] & P. Licinio[3]

1 MRC, Lab. of Mol. Biol., Hills Road, Cambridge, and Dept. Appl. Math. & Th. Phys., Silver Street, Cambridge.

2. LURE, CNRS,-MEN-CEA, Bat 209D, Universite Paris Sud, F-91405, Orsay, France.

3. Physique des solides, Bat 510, Universite Paris Sud. F-91405, Orsay, France.

ABSTRACT. References are given to the main work in this field which has already been published. Further progress is reported in quasi-elastic light scattering QLS, and pulse fluorescence. In the former, the program was successfully re-written to handle the measured data directly, which are a non-linear function of the spectrum. This led to an increase in computing speed of up to a factor of 10. In the latter, the program was developed to analyse both the rotational or flexural time constants and the decay rates, simultaneously.

1. INTRODUCTION

Most of our work on the application of maximum entropy to the Laplace transform has already been published. An overview of the use of maximum entropy method in Quasi-elastic light scattering (QLS) and pulse-fluorescence can be found in reference 1. The details of applying the technique to QLS is given in reference 2 and an example of its use in helping to solve a complex biological problem is given in reference 3. Similarly, the details of the application to pulse-fluorescence are contained in reference 4, and a detailed example of its use and comparison with non-linear least squares techniques in an experimental environment is presented in reference 5. The rest of this paper outlines two recent advances in these fields which have occurred since the publication of these papers.

2. IMPROVEMENTS TO THE QLS CODE

Following the notation of reference 2, the measured autocorrelation signal $C(\tau)$ in a QLS experiment using "homodyne" detection (self-beating of the scattered field on the photocathode) is given by

$$C(t) = B + y(t)^2 = B + \left[G(\tau) \exp - t/\tau \, d\tau\right]^2$$

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 191–194.

where we wish to obtain an estimate of the positive spectrum of decay times $G(\tau)$, having measured an inevitably noisy and incomplete representation of the autocorrelation of the scattered light $C(t)$. Since $y(t)$ must always be positive (because of the exponential in equation 1), we estimated $G(\tau)$ from the derived data

$$y'(t) = C(t)-B$$

$y'(t)$ is thus a linear function of $G(\tau)$ and the standard Cambridge package could be used to determine $G(\tau)$. As reported in reference 2 this was successful although convergence was frequently slow. Such ill-conditioned problems were at the limit of the already powerful Cambridge algorithms. We also encountered problems with our estimates of the errors in $y(t)$ at long delay times when the signal is very small with respect to the background. Since the variance of y is

$$\sigma_k^2 = \frac{C_k}{4|C_k-B|}$$

when $|C_k-B|$ is small, large (statistical) fluctuations in B lead to erroneous estimates of the errors preventing final convergence of the algorithm. The algorithm did allow the user to recognise and remove the aberrant points, allowing these (few) data points to be removed, and the algorithm re-run. This solution gives better estimates for $|C_k-B| = y_k^2$ allowing the error estimates to be corrected, the erroneous points reinserted and the algorithm re-run. However, this increased user intervention and triple running of the algorithm was a poor feature of the program.

The solution to this problem was to re-write the algorithm to work with the measured auto-correlation data, C, directly for which the errors are well-defined. C is not, of course, a linear function of the contents of $G(\tau)$ of the spectrum.

To encode this, we calculated the mock data by the correct non-linear function. However, the quadratic models of both the entropy and chi-squared functions about the true position were built using the local differential response of the data. This could, of course, result in slower convergence of the program. Nevertheless, the increased power of the new algorithm and the removal of the need to re-run the program to obtain better estimates of the error bars, more than compensates for this. Indeed, a single run of the program runs 3-4 times faster than the old program, and since it now only needs to be run once (without the need of any user intervention) the total saving is about a factor of 10. Currently we find a typical spectrum of 100 data points with a signal-to-noise of 1000:1 takes about 2 minutes CPU on a VAX 780.

3. EXTENDING PULSE-FLUORESCENCE TO ANALYSE ROTATIONAL AND FLEXURAL MOTION

Following the notation of reference 5 we note that with a vertically polarised excitation the parallel I// and perpendicular $I_{\perp}$ components of the fluorescence intensity at time t after the start of the excitation flash are

$$I_{II}(t) = \frac{1}{3} E_{\lambda}(t) * \left(\int_0^{\infty}\int_0^{\infty}\int_{-0.2}^{0.4} \gamma(\tau,\theta,A)\, e^{-t/\tau}\,(1+2Ae^{-t/\theta})\, d\tau d\theta dA\right.$$

$$I_{\perp}(t) = \frac{1}{3} E_{\lambda}(t) * \left(\int_0^{\infty}\int_0^{\infty}\int_{-0.2}^{0.4} \gamma(\tau,\theta,A)\, e^{-t/\tau}\,(1-Ae^{-t/\theta})\, d\tau d\theta dA\right.$$

where E(t) is the temporal shape of the flash and $\gamma(\tau,\theta,A)$ are the number of fluorophores with fluorescence decay τ, rotation time θ, and initial amplitude of anisotropy A (related to angle between absorption and emission dipoles). * denotes a convolution with time.

In reference 5 we saw that using a particular sum of the parallel and perpendicular components ($I_{//} + 2I_{\perp}$) we could reduce our problem to the one-dimensional problem of determining the distribution of lifetimes $\alpha(\tau)$. We have now extended our program to calculate this full 3-dimensional density. That is we wish to determine the numbers of fluorophores $\gamma(A,\theta,\tau)$ with a particular lifetime τ, rotational constant θ and initial anisotropy A. We present here our first experimental result from a protein, namely the apocytochrome C protein. Since this protein has only a single tryptophan residue we know that all the fluorophores have the same initial anisotropy A although their lifetimes τ and rotational constants θ change due to local (in time and space) fluctuations in time and space. This reduces the problem to 2-dimensions, easing both the computation and display.

Data were measured at the synchrotron on a single tryptophane residue of apocytochrome in aqueous buffer at 20°C at the concentration of 2.5 mg/ml. The contour plot of a section through $\gamma(\tau,\theta,A)$ for A = 0.258 is presented in figure 1b. We can clearly distinguish four lifetime components centered at 0.15, 1.2, 3.14 and 5.4 ns as found previously in a one-dimensional analysis of T(t) (5).

Along the θ axis we can see two high peaks from lifetime τ=3.1 ns centered at θ_1=0.14 ns and θ_2=1.4 ns which reflect the major contribution to the depolarisation process. An identical minor contribution is given by the longest τ of 5.4 ns.

On the contrary, the shortest lifetime (τ=0.15) does not play any role in the fast motion (θ_1=0.14 ns). Its corresponding θ value remains uncertain due to the weak contribution to the signal of a such short lifetime.

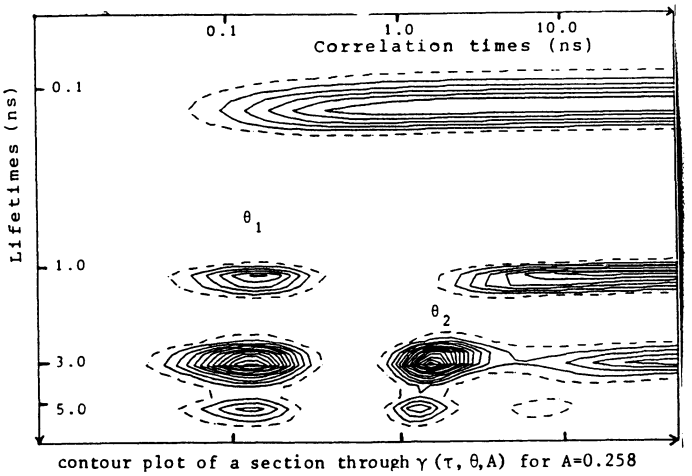
The intermediate τ of 1.2 ns monitors the fast flexibility and the overall motion of the protein (MW=11 900) but is clearly not involved in the intermediate flexibility (θ_2).

The long tail component of correlation times has probably no physical meaning and could be related to an anisotropy of noise on $I_{//}$ and $I_{\perp}$.

In conclusion this example fully demonstrates th ability of Maximum Entropy Method of analysis to resolve both structural heterogeneity and complex protein dynamics from pulse fluorometry data.

REFERENCES

1. A.K. Livesey, P. Licinio, M. Delaye & J-C. Brochon 'Maximum Entropy Analysis of Dynamic Parameters via Laplace Transform'. Proc. of 83rd Faraday Discussion Meeting April 1987, pp 247-258, 317-319, 325.
2. A.K. Livesey, P. Licinio, M. Delaye 'Maximum Entropy Analysis of Quasi-Elastic Light Scattering from Colloidal Dispersions'. J. Phys. Chem. 84 (1986) 5102-5107.
3. P. Licinio, M. Delaye & A.K. Livesey 'Dynamics of colloidal dispersions of alpha-crystallin proteins: a maximum entropy analysis of photon correlation spectroscopy data'. J. de Physique (Paris) 48 (1987) 1217-1223.
4. A.K. Livesey, J-.C. Brochon 'Recovering Distribution of Decay Constants in Pulse-Fluorometry using Maximum Entropy'. Biophysical Journal 52 (1987) 693-706.
5. F. Merola, P. Rigler, A. Holmgren 11, and J-C. Brochon 'Pico-second tryptophan fluorescence of Thioradoxi: Evidence for discrete species in slow exchange'. Submitted to Biophysical Journal.



contour plot of a section through $\gamma(\tau, \theta, A)$ for A=0.258

MAXIMUM ENTROPY AND BAYESIAN APPROACH IN TOMOGRAPHIC IMAGE RECONSTRUCTION AND RESTORATION

Ali Mohammad-Djafari and Guy Demoment
Laboratoire des Signaux et Systèmes (CNRS-ESE-UPS)
Plateau du Moulon, 91192 Gif-sur-Yvette Cédex, France.

ABSTRACT. In this paper we propose a Bayesian approach with Maximum Entropy (ME) priors to solve an integral equation which arises in various image restoration and reconstruction problems. Our contributions in this paper are the following: i) We discuss the *a priori* probability distributions which are deduced from different *a priori* constraints when the principle of ME is used. ii) When the *a priori* knowledge is only the noise covariance matrix and the image total intensity, and when the maximum *a posteriori* (MAP) is chosen as the decision rule to determine the values of image pixels, we show that the solution may be obtained by minimizing a criterion in which the structural entropy of the image is used as a particular choice of a regularization functional. The discussion is illustrated with some simulated results.

1. Introduction

In many problems of image reconstruction and restoration one must solve an integral equation of the form :

$$g(\mathbf{x}) = \int_D f(\mathbf{x}')\, h(\mathbf{x}, \mathbf{x}')\, d\mathbf{x}' \qquad (1)$$

in which $\mathbf{x}, \mathbf{x}' \in \mathbf{R}^2$ are the space variables, g is the observed data, f the image to be determined, h a known function which is the kernel of the imaging system, and D is the support of the image f, a compact region in $\mathbf{R}^2$. In *image restoration* problems f is the *original image*, g is the *degraded image* and h is the *point spread function* (PSF) of the imaging system. In image reconstruction problems f is the *object*, g is called the *projections* (in fact g is not a continuous function of $(x,y) \in \mathbf{R}^2$ but a finite set of functions in one space variable $r \in \mathbf{R}$ parameterized by the other space variable $\theta \in [0,\pi]$) and h is the *kernel* of the imaging system. In the two cases, the inversion of these equations is an ill-posed problem. By ill-posed problem we mean that it does not satisfy the three conditions of *existence*, *uniqueness* and *stability* of the solution.

The numerical solution of these equations needs a discretization procedure which can be done by a quadrature method. The linear system of equations resulting from the discretization of an ill-posed problem is, in general, very ill-conditioned if not singular. So the problem is to find a unique and stable solution for this linear system. The general method which permits us to find a unique and stable solution to an ill-posed problem by introducing an *a priori* information on the solution is called the *regularization* . The *a priori* information can be either in a deterministic form (positivity,...) or in a stochastic form (some contraints on the probability density functions). The unavoidable existence of errors and noise on the measured data leads us to adopt a stochastic approach. The Bayesian approach is a coherent one for solving inverse problems because it allows us to take into account in a coherent way the *a priori* information both on the solution and on the data as well as the errors on the data.

This approach involves the following steps:

i) Assign a probability distribution function (pdf) to the unknown parameter to translate our incomplet *a priori* information about these parameters;

ii) Assign a pdf to the measured data to translate the lack of total precision and the unavoidable existence of the measurement noise;

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 195–201.

iii) Use the Bayesian rule to calculate the *a posteriori* pdf of the unknown parameters;
iv) Define a decision rule to determine the values of these parameters, for example, the values that have the *maximum a posteriori* (MAP) probabilities.

One must note that in this approach :
i) One is able to solve the inverse problems which are described by a finite number of parameters, for example when one has discretized the integral equation (1).
ii) Assigning a probability to a parameter value does not mean forcibly that this parameter is a random variable or that the probability is a limit to its realization frequency. The probability is just a measure of our confidence to that value of parameter.
iii) If it is easy to assign a pdf to the measured data to take into account the noise, it is more difficult to assign a pdf to the unknown parameters of the problem to translate our *a priori* knowledge about them, because this knowledge is not given to us directly in probabilistic terms, and it does not permit us to determine an unique pdf for those parameters. One can use the principle of maximum entropy (ME) to choose one (which has maximum entropy) between all possible pdf satisfying the a priori knowledge constraints.

In this paper we first discuss the basic ideas of of the bayesian approach and principle of ME and then show how these ideas can be used in image reconstruction and restoration problems. Then we show some results simulating the X-ray and diffraction tomography image reconstruction problems.

2. Bayesian approach of the resolution of inverse problems

Once the integral equation (1) is discretized one has to solve a linear system of the form:

$$\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b} \tag{2}$$

where $\mathbf{x}$ is a vector containing all the unknown parameters (image pixel values for example), $\mathbf{y}$ is a vector containing all the measured data (degraded image pixel values or projections), A is a known matrix whose components and structure depend on the imaging system, and $\mathbf{b}$ is a vector containing the measurement errors.

In a statistic approach, $\mathbf{b}$ and $\mathbf{y}$ are considered to be the random vectors which is a natural way to represente the random errors and noise. In a Bayesian statistic approach $\mathbf{x}$ is also considered to be the unknown vector of parameters of a random process which we want to determine. In imaging applications this means that the image is considered as a random process, i.e. each pixel of the image is a random variable and each x_i is a parameter defining its probability distribution. This can have a realistic physical meaning. For example in Positron Emission Tomography (PET) the image is considered as a Poisson random process and x_i is the mean value of the number of positrons emitted in each pixel. But it can be just a mathematical tool to translate our a priori knowledge about these parameters.

Now we suppose that we are able not only to assign a pdf to $\mathbf{y}$ and $\mathbf{b}$ but also to assign a pdf to $\mathbf{x}$ to describe our pior knowledge about it (as mentioned this does not mean forcibly that x_i are random variables but that our knowledge about them is incomplete). This means that we can assign the probability densities $p(\mathbf{y}|\mathbf{x})$, $p(\mathbf{x})$, $p(\mathbf{y})$ and $p(\mathbf{x}|\mathbf{y})$ which are related by the Bayes' formula :

$$p(\mathbf{x}|\mathbf{y}) = p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x}) / p(\mathbf{y}) \tag{3}$$

Thus, in terms of the Bayesian approach $p(\mathbf{x}|\mathbf{y})$ is the solution of our problem. But in practical applications two major difficulties are encountered :
i) How to determine $p(\mathbf{x})$ (how to translate our *a priori* kowledge about the parameters $\mathbf{x}$ by a probabilty density function), and
ii) How to give numerical values to x_i when $p(\mathbf{x}|\mathbf{y})$ is calculated.

The second is the easier one to solve: Define a decision rule, for example the *maximum a posteriori* (MAP) and the solution is $\mathbf{x}^*$ which maximizes $p(\mathbf{x}|\mathbf{y})$. The first is more difficult. The problem is how to translate our prior knowledge about $\mathbf{x}$ which is often of the form : the image is support limited, or is positive and has a known expected gray level, etc. These knowledges are not normally sufficient to define an unique pdf $p(\mathbf{x})$. It is here that the ME principle can be used to choose one possible distribution which is coherent with this prior knowledge and which does not introduce any other extra information.

3. Prior probabilities and maximum entropy

We consider an image as a finite number of pixels and suppose that the pixel values are independent random variables of the same probability law $p(x_i)$. So for the whole **image** (all the pixels) we have :

$$p(\mathbf{x})=p(x_1,\ldots x_n) = \prod_{i=1}^{n} p(x_i) \tag{4}$$

If we note by H_i the entropy of the **pixel** i and by H the entropy of the **Image** we have:

$$H_i = -\int p(x_i) \log p(x_i)\ dx_i \quad \text{and} \quad H = \sum_{i=1}^{n} H_i \tag{5}$$

Consider now the following examples :

a) We are given the average values $[\lambda_1,\ldots, \lambda_n]$, and the variance values $[\sigma_1^2,\ldots, \sigma_n^2]$ of the pixels and we don't known anything else and we want to determine the probabilty density function of the image **x**. If we apply the ME principle we obtain for each pixel :

$$p(x_i) = \mathbf{N}(\lambda_i\ ,\ \sigma_i^2) \tag{6}$$

which is a gaussian probabilty density function. The corresponding maximum entropy is :

$$H_i = (1/2)\ \log(2\pi e\ \sigma_i^2) = (1/2)\ \log 2\pi e + \log \sigma_i \tag{7}$$

The image has the probability density $p(\mathbf{x})$ given by :

$$p(\mathbf{x}) = \mathbf{N}(\mathbf{m}, R) \quad \text{avec} \quad \mathbf{m} = [\lambda_1,\ldots,\lambda_n]^t \quad \text{et} \quad R = \text{diag}\ [\sigma_1^2,\ldots, \sigma_n^2] \tag{8}$$

and the corresponding maximum entropy of the image is :

$$H = \sum_{i=1}^{n} H_i = \sum_{i=1}^{n} \log \sigma_i + (n/2)\ \log (2\pi\ e\ ) \tag{9}$$

What we may note in (9) is that the entropy of the image depends on the spatial distribution of the pixel variances σ_i and we have an expression in the forme $\sum \log \sigma_i$.

b) We are given only the average values $[\lambda_1,\ldots, \lambda_n]$, and know that the pixel values are **positive**. Now maximizing the entropy H_i subject to these constraints we obtain :

$$p(x_i) = (1/\lambda_i)\ \exp[\ -(x_i/\lambda_i)\ ] \tag{10}$$

The corresponding maximum entropy for each pixel is :

$$H_i = 1 + \log\lambda_i \tag{11}$$

and the corresponding maximum entropy of the image is :

$$H = \sum_{i=1}^{n} (\ 1 + \log\lambda_i) = n\ +\ \sum_{i=1}^{n} \log\lambda_i \tag{12}$$

In this case the entropy of the image depends on the spatial distribution of the pixel mean values λ_i .

c) Suppose now that we are given only one global constraint which is in the forme :

$$-\sum_{i=1}^{n} x_i\ \log x_i\ = s\ = S(\mathbf{x}) \tag{13}$$

If the image is normalized it can be considered as a probability distribution function and $S(\mathbf{x})$ is then its *structural entropy* . The name structural entropy is due to Skilling et al [1,13]. This is also equivalent to a knowledge of the total intensity of the image. With this only global constraint the ME principle gives us an exponential pdf in the forme :

$$p(\mathbf{x}) = \exp\ [\lambda_0 + \lambda_1\ S(\mathbf{x})\ ] \tag{14}$$

We conclude this by noting that for different *a piori* knowledge one obtains different expressions for the *a piori* probability distribution of the image. When we use a positivity constraint and just the knowledge of the total intensity of the image we obtain an exponential probability distribution as given in (14).

4. Bayesian solution with ME prior

We have seen that the principal mathematical problem of image reconstruction and image restoration is to solve the equation $\mathbf{y} = A\mathbf{x} + \mathbf{b}$. Suppose that the *a priori* knowledge that we have on the noise **b** and on the solution **x** are :

i) b_j , j=1,..., m are independent identical zero-mean random variables with variance σ^2;

ii) x_i , i= 1,..., n are independent identical discrete and positive random variables ;

iii) the average value of $-x_i \log x_i$ is known; and **iv)** A is a deterministic known matrix.

By applying the principle of ME we can thus conclude that :

i) The ME density function for **b** is a Gaussian one. We then have

$$p(\mathbf{y} \mid \mathbf{x}) = \text{cte} \exp[- Q(\mathbf{x})/\sigma^2] \quad \text{with} \quad Q(\mathbf{x}) = [\mathbf{y} - \mathbf{A}\mathbf{x}]^t [\mathbf{y} - \mathbf{A}\mathbf{x}] \tag{14}$$

ii) The ME density function for **x** is an exponential one so that we have

$$p(\mathbf{x}) = \text{cte.} \exp[\lambda_1 S(\mathbf{x})] \quad \text{with} \quad S(\mathbf{x}) = - \sum_{i=1}^{n} q_i \log q_i \quad \text{and} \quad q_i = x_i / N = x_i / \sum x_i \tag{15}$$

We can now apply the Bayes' formula to find

$$p(\mathbf{x} \mid \mathbf{y}) = \text{cte.} \exp[- Q(\mathbf{x}) + \lambda S(\mathbf{x})] \quad \text{with} \quad \lambda = \lambda_1 \sigma^2 \tag{16}$$

and if we decide that the *best estimation* for the pixel values are those which have the *maximum a posteriori* probability $p(\mathbf{x} \mid \mathbf{y})$ the problem becomes

$$\min J(\mathbf{x}) = Q(\mathbf{x}) - \lambda S(\mathbf{x}) \tag{17}$$

The problem is now to minimize a non-quadratic function J(**x**) subject to the constraint **x**>0. We may give some interpretation about this result :

i) If we compare J(**x**) with the deterministic regularization criterion of Tikhonov, Phillips and Twomey [10] we see that S(**x**) plays the role of a regularizing functional and λ is the regularization parameter.

ii) One can arrive at the same result which is to minimize J(**x**) by other explanations, for example, to use a χ^2 statistics and choose between the solutions **x** which satisfies $Q(\mathbf{x})<\varepsilon$ the one which maximizes the entropy S(**x**).

5. Algorithmic difficulties

We have seen how the Bayesian approach with ME priors results to the problem of minimizing a function in the form $J(\mathbf{x}) = Q(\mathbf{x}) - \lambda S(\mathbf{x})$ subject to the constraint **x**>0. Whatever the interpretation, arrived at this stage, one has a mathematical problem which is a constraint minimization of a non-quadratic function which can be achieved only by an iterative method.

Among the iterative methods of minimizing a non-quadratic function we considered those who search the minimum by a series of monodimensional searches. These methods can be classified in terms of different local informations which one disposes about the function to be minimized. The zeroth, first and second order methods use respectively the function; the function and its gradient; the function, its gradient and its Hessian matrix; to determine the search directions. For applications of image processing it is not possible to use the second order methods due to the huge cpu memory and time needed.

We used a conjugate gradient technique which can be considered as a quasi-second order method. The principal properties of this technique are now well established [16].

5.1. ALGORITHM IMPLEMENTATION

We have implemented a general algorithm to solve the following problem:

Given $\mathbf{y} = \mathbf{A}\,\mathbf{x} + \mathbf{b}$ with $\mathbf{y} = [x_1,\ldots, x_m]^t$, $\mathbf{b} = [b_1,\ldots, b_m]^t$, $\mathbf{x} = [x_1,\ldots, x_n]^t$ and $A=[A_{ij}]$; determine **x** which minimizes $J(\mathbf{x})=Q(\mathbf{x})-\lambda S(\mathbf{x})$, where S(**x**) is the entropy expression and Q(**x**) is a quadratic expression. λ is a constant which must be choosed so that the solution $\mathbf{x}^*$ satisfies the constraint $Q(\mathbf{x}^*)\leq M$. The optimum choice of λ needs more complicated and time-consuming algorithm [13,14]. In our algorithm we choose it empirically at the initialization of the algorithm. Some discussions about this choice in differents applications are given in [7, 8]. The algorithm follows these sequences:

1) An initial estimate is calculated using $\mathbf{x}^{(0)}= (1/M)\, \mathbf{A}^{*t}.\mathbf{y}$

2) At any iteration (k) do:

a) apply the positivity constraint, i.e. if $x_n \leq 0$ then $x_n=\varepsilon$ with ε a small positive value.

b) calculate $J(\mathbf{x}) = Q(\mathbf{x}) - \lambda S(\mathbf{x})$ and his gradient $\nabla J(\mathbf{x}) = \nabla Q(\mathbf{x}) - \lambda \nabla S(\mathbf{x})$

c) deviate the gradient; i.e. if $x_n \leq 0$ then $x_n=\varepsilon$ and if $\partial J(\mathbf{x})/\partial x_n > 0$ then $\partial J(\mathbf{x})/\partial x_n = 0$

d) calculate a new estimate of x_n by the Conjugate Gradient.

e) Some criterion are calculated to ensure the normal execution of the algorithm [8].

6. Simulations and results

6.1. X-RAY TOMOGRAPHY WITH LIMITED PROJECTIONS

In these simulations we choosed a mathematical phantom as an object, calculated the projections, added some noise (S/N=10 dB), and considered a difficult situation where the number of projections are limited to 12. Figure (1) shows the reconstruction results obtained by our **ME** method and some other classical methods as : **ART** (algebraic reconstruction techniques), **SIRT** (simultaneous reconstruction techniques), **ILS** (iterative least squares) and **ILSP** (iterative least squares with positivity constraint applied in each iteration). As we can see all linear methods give the results which are not acceptable. It is the same for the ILSP method in which one can see many abnormal point like sources outside the support of the image. We can conclude also that just the positivity constraint is not sufficient to insure a stable solution.

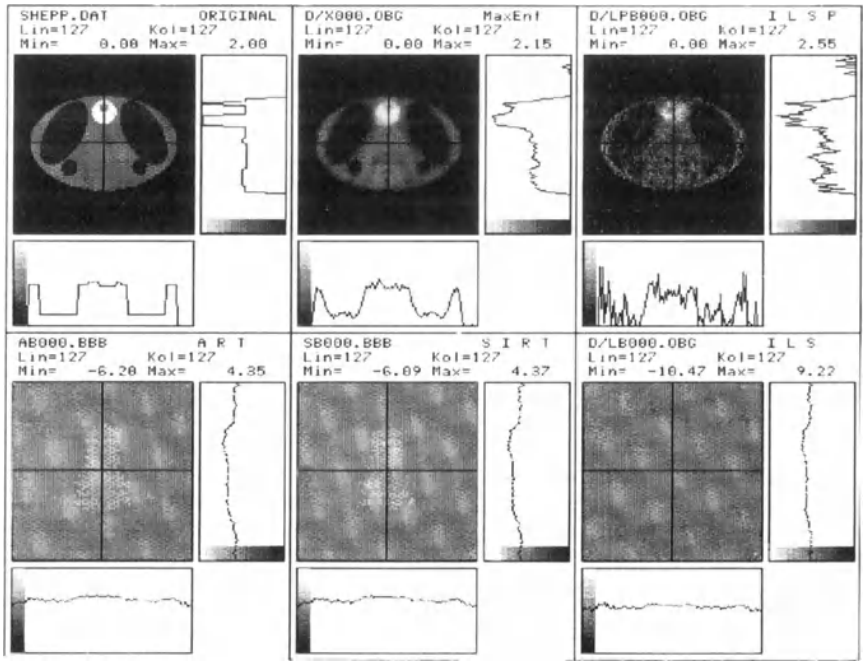


Figure 1: X-ray Tomography: **a**) original, **b**) ME, **c**) ILSP **d**) ART , **e**) SIRT , and **f**) ILS

6.2. DIFFRACTION TOMOGRAPHY AND FOURIER SYNTHESIS

To simulate the Fourier synthesis part of the diffraction tomography imaging we used the same fantom as in X ray tomography and calculated its FT on semi-circles and added some noise on these data. The S/N ratio was about 10dB (separatly for the real and imaginary parts). Then, given these data, we proceeded the reconstruction either by the two classical methods of interpolation used currently in diffraction tomography (Fourier domain interpolation : **M1** and spatial domain interpolation : **M2**) or by our **ME** algorithm [8].

For these simulations we considered two cases: **i**) Np=12, θ=360° and **ii**) Np=8, θ=90° ; where Np is the number of projections and, θ is the angle restriction.

In these cases also the results obtained by our **ME** method have better resolution both in spatial extent and in amplitude. Thes results are also obtained after 20 iterations.

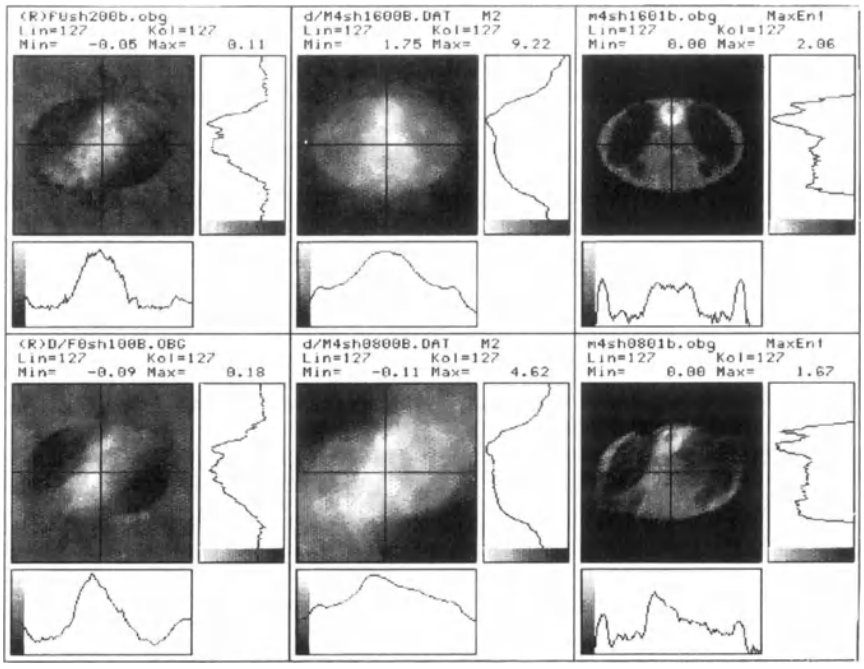


Figure 2: Diffraction Tomography :Upper row: Np=16 and θ=360°, Lower row: Np=8 and θ=90°.
a) reconstruction by **M1**, **b**) reconstruction by **M2**, and **c**) reconstruction by our **ME**.

6.3. IMAGE RESTORATION BY DECONVOLUTION WITH SPARSE DATA

In these simulations we considered an image which blured by a linear PSF and degraded by a gaussian noise and considered the case where we dispose only about 10% of the data samples (1 raw and 1 colomn over 3). Then given these data we restored the original image either by our ME method or by ILS or ILSP methods. Figure (3) shows these results.

7. Conclusions

We proposed a bayesian approach with ME prior method to solve the ill-posed problems of image reconstruction in tomography. We first presented this Bayesian approach, discussed about the different expressions of the *a priori* pdf and the entropy of an image. Then, we saw how the structural entropy of an image can be used as a regularization function. This method is used: **i**) to reconstruct the objects in X ray tomography using directly the projection data in spatial domain, **ii**) to reconstruct the object in diffraction tomography from the data in Fourier domain (Fourier synthesis problem), and **iii**) to image restoration in the situation of sparse data. We focus now our attention on the real application of the method. In a real world we only have the data **y** and a qualitative knownledge of the noise and the image that we want to find. The real problem is how to determine the variance of the noise and the regularization parameter λ only from the data.

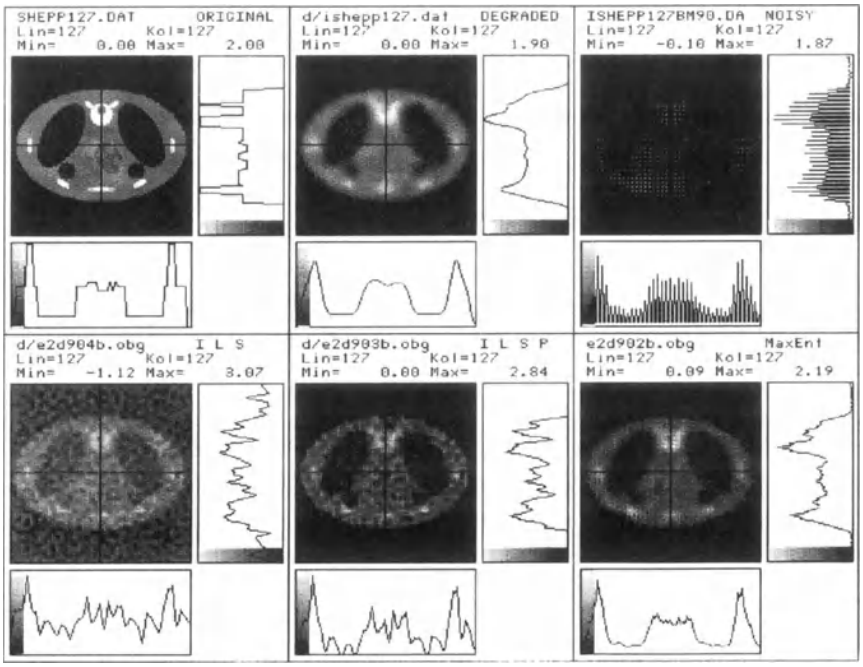


Figure 3: Image restoration by deconvolution in the case of spares data :
a) original, **b**) blured , **c**) noisy sparse data, **d**) deconvolution by ILS, **e**) by ILSP, and **f**) by ME .

8. References

[1] Burch S.F. , Gull S.F. and Skilling J., "Image restoration by a powerful maximum entropy method," *Comput. Vis. Graph. Im. Process.*, **23**, pp. 113-128, 1983.

[4] Frieden B.R. and Zoltani C.K., "Maximum bounded entropy: application to tomographic reconstruction," *Applied Optics*, **24**, pp. 201-207, 1985.

[5] Jaynes E.T. , "On the Rationale of Maximum Entropy Methods," *Proc. IEEE*, **70**, pp. 939-952, 1982.

[6] Jaynes E.T. , "Where do we go from here?," *Maximum-Entropy and Bayesian Methods in Inverse Problems*, C.R. Smith & T. Grandy, Jr. (eds.), pp. 21–58, 1985.

[7] Justice J.H., *Maximum entropy and bayesian methods in applied statistics*, Cambridge Univ. Press., 1986.

[8] Mohammad-Djafari A. and Demoment G., "Maximum entropy Fourier synthesis with application to diffraction tomography," *Applied Optics*, **26**, pp. 1745-1754, 1987.

[9] Mukherjee D. and Hurst D.C., "Maximum Entropy Revisited," *Statistica Neerlandica* **38**, pp. 1-12, 1984.

[10] Phillips D.L. , "A technique for the numerical solution of certain integral equations of the first kind," *J. Assoc. Comput. Mach.*, **9**, pp. 84-97, 1962.

[11] Powell M.J.D. , "Restart procedure for the conjugate gradient method," *Math. Prog*, **12**, pp. 241-254, 1977.

[12] Skilling J. and Gull S.F., "The Entropy of an Image," *SIAM-AMS Proceedings*, **14**, pp. 167-189, 1984.

[13] Wernecke S.J. and D'Addario L.R., "Maximum entropy image reconstruction," *IEEE Trans.*, **C-26**, pp. 351-364, 1977.

Maximum-Entropy-Based Approaches to X-ray Structure Determination and Data Processing.

S. Steenstrup
Physics Laboratory, H. C. Ørsted Institute,
Universitetsparken 5,
2100 Copenhagen Ø, Denmark

S.W. Wilkins
CSIRO, Division of Materials Science and Technology,
Locked Bag 33,
Clayton, Vic. 3168, Australia

ABSTRACT. A brief survey is given of some maximum-entropy-based approaches to x-ray structure refinement and determination of macromolecules. Particular emphasis is placed on those approaches which primarily seek to operate in direct space and involve combining various types of information. Some comments on current work proceeding along these lines are offered.

In addition, a powerful maximum-entropy-based algorithm for combined background subtraction, deconvolution and filtering of one-dimensional profiles is outlined and illustrative results presented for processing of synchrotron data for energy-dispersive powder diffraction from materials under high pressure.

1 Introduction

Crystal structure determination is not a single well-defined subject. It ranges form the determination of the atomic arrangement of large to very large structures like macromolecules such as proteins and viruses, to the simple determination of crystal-class and unit-cell size. For the former, single crystal diffraction is needed while for the latter, powder diffraction is sufficient. Each technique poses its own problems. For the determination of large structures one main problem is that the phases of the structure factors are essentially unknown. In the case of powder diffraction, a key problem (apart from the phases also being unknown) is that only a few structure factors are measured and the observed peaks in the profile often overlap each other.

These two aspects are briefly discussed in the following paper starting with some remarks on recent advances in maxent approaches to the macromolecular structure problem. In a second part, a specific example of treating powder diffraction data is given.

2 Large structures

2.1 The problem

The physics of the problem is well known. The relation between the electron density, $\rho(\mathbf{r})$, in the unit cell and the (complex) structure factors, $F_{\mathbf{k}}$, is given by

$$\rho(\mathbf{r}) = \sum_{\mathbf{k}} F_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \tag{1}$$

Visually each term in the Fourier series in (1) is a sinusoid with a wavelength $d/sin\theta$, with d the distance between planes of atoms and θ the angle between these planes and the direction of the incident wave, and is generally phased such that the wave has crests at the atomic planes. When superposing these waves (with the correct phases) in the sum the waves tend to reinforce each other at the atomic positions and cancel to near zero at positions far from the atoms. In a general sense the amplitudes determine the size of the atoms while the phases determine their positions. However, the phases are essentially unknown, since it is not the structure factors but the intensities $I_{\mathbf{k}} = |F_{\mathbf{k}}|^2$ that are measured, so the problem is not determinate.

The ability of the maximum entropy principle to handle underdetermined problems encounters some diffuculties in this case. The reason for this is seen as follows: What is known is a set of intensities $I_{\mathbf{k}}; \mathbf{k} \in \Omega$, ($\Omega$ being some subset of the reciprocal lattice). The uncertainty relates to: which electron density $\rho(\mathbf{r})$ is the best estimate for the actual one corresponding to the measured intensities. Whithin specific models as discussed earlier[1] this uncertainty is expressed in terms of a discretized electron density ρ_i as $S = -\sum p_i \ln p_i$; $p_i = \rho_i / \sum \rho_j$ or (as suggested by Skilling[2] and by[3]) by $S = \sum(\rho_i - m_i) - \rho_i \ln(\rho_i/m_i)$, with m_i a set of default values. The usual maximization of S with the intensities as constraints (including errors) involves some difficulties since the constraints do not form a convex set, so there is in general no unique solution for the optimization problem. (e.g. see[13]).

Nonetheless, there are advantages in using the maxent principle since, as mentioned below, there is often additional a priori information available and also additional experimental techniques provide some information on few phases, so that a convex or more nearly convex constraint set may result, leading in turn to a single "best" solution. Failing this a multisolution approach would appear necassary and methods for tackling this problem via maxent and Bayesian methods have recently been described by Bricogne[8].

2.2 Attempts at solution

The problem of *ab initio* determination of crystal structures from the measured x-ray diffraction intensity data remains essentially unsolved. For practical purposes structures consisting of up to the order of a hundred non-hydrogen atoms may usually be solved by traditional direct methods, see e.g. the recent review of Woolfson[4]. These methods are based on probabilistic relations for phases of invariants (quantities independent of the choise of origin in the unit cell) and semi-invariants (quantities which do not change value by transfer from one special origin to another). In the case of larger molecules such as proteins and viruses, crystallographers have developed a wide variety of procedures for

helping to elucidate their structure, including: the use of isomorphous replacement of atoms in the structure (the so-called "heavy atom method"), the use of anomalous dispersion together with multiple-wavelength measurements, the use of solvent flattening, and the use of partial fragments. However, with extension into the area of increased complexity and size, conventional methods have been found to become increasingly time consuming and unreliable. There is therefore a need to seek new, more efficient and less subjective approaches to crystal structure estimation.

Bayesian inference combined with the maximum entropy method provide a logically consistent method for incorporating such information in order to make predictions of crystal structure. Recently, various approaches to crystal structure estimation based on maxent principles have been described in the literature. For example, Gull & Daniell[5], Collins[6], Wilkins Varghese & Lehmann[7], Bricogne[8], Navaza[9], Bryan, Bansal, Folkhard, Nave & Marvin[10], Wilkins and Stuart[11], Gull, Livesey & Sivia[12], Bryan & Banner[13], and Bryan[14]. The precise application of maxent in each case tends to be different although the guiding philosophy remains the same.

Even more recently, Bricogne[15] has described a very general approach to crystal-structure estimation based on Bayesian inference and invoking the saddle-point method (maxent) to establish prior joint probability distributions of structure factors. The proposed procedure potentially incorporates all the types of information listed above and so may provide a powerful and general approach to the x-ray structure determination of macromolecules.

3 Powder diffraction

Even for cases where the information is so incomplete that not even intensity data is available, the maximum entropy method is valuable for obtaining partial information such as the crystal class and unit cell parameters. In order to appreciate the way in which the maxent principle enters in some of our work, the essentials of the experimental technique for doing high-pressure X-ray diffraction studies with the energy dispersive technique and using synchrotron radiation, are briefly described[16].

3.1 High pressures with a diamond anvil cell.

The high pressures (up to $\sim$100GPa) are obtained in a diamond anvil cell (for a general review see[17]), in which two diamonds, ($\sim$1/2 carat each) placed in a suitable steel-holder and cut in a brilliant shape with the two smallest flat faces $\sim$1/2 mm in diameter, are pressed together. The sample, in the form of a powder, is placed in a hole in a thin metal sheet (the gasket). The hole is usually $\sim$100 μm in diameter and the sample thickness is $\sim$80 μm. Combined with the sample are small pieces of ruby which serve to monitor the pressure and a liquid to ensure hydrostatic pressure.

3.2 The Energy-dispersive technique.

In the energy-dispersive technique a white beam is shone through one diamond and is scattered from the sample. The diffracted beam is detected by an energy resolving detector

after having passed through the second diamond and a set of angle defining slits. A multichannel analyser records the spectrum.

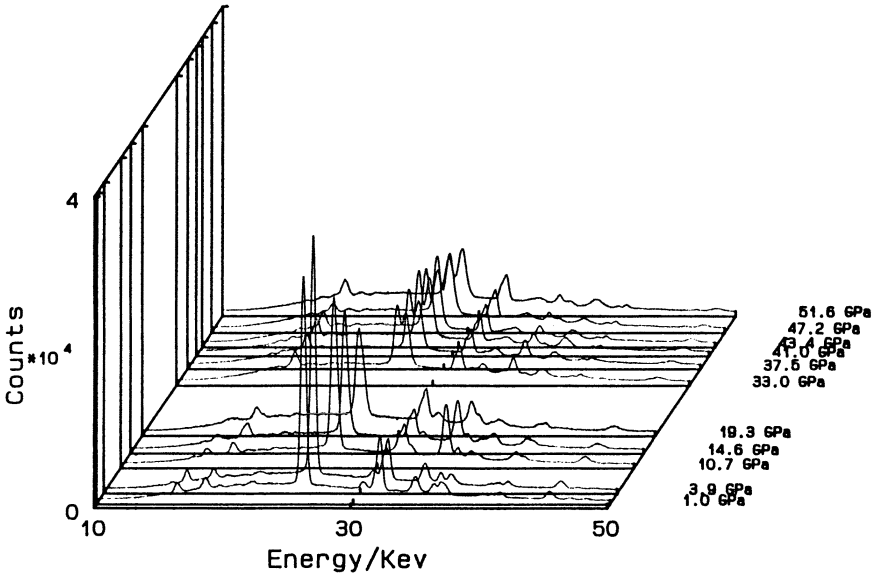


Figure 1: Raw diffraction data from a sample of $EuBa_2Cu_3O_{9-\delta}$

A number of points are important:

- The high pressure cell limits the scattering angle to values less than ~15 degrees.
- The available sample and the beam size limits the number of grains contributing to the diffracted intensity to a small number, making texture almost unavoidable.
- The energy resolution of the detector, even with the best Ge cooled detectors, is limited to around 200 eV at 20 keV[18].
- The only practical source of white radiation is synchrotron radiation since the diffracted intensity is only a small fraction of the incoming intensity due to the small size of the sample.
- The high intensity of the synchrotron beam gives rise to an important amount of scattered radiation yielding a high background intensity.

To illustrate these points, a set of measurements is shown in fig. 1[19]. The raw data from a sample of $EuBa_2Cu_3O_{9-\delta}$, scaled to the same total number of counts, is shown for pressures from ambient pressure to 53.6GPa.

At ambient pressure a conventional structure determination is possible, and is one way of obtaining the indexing of diffraction peaks. At higher pressures no such possibility exists but invoking the maximum entropy principle, spectra comparable in resolution to those measured with the more conventional angle-dispersive technique can be obtained. (Note

though, that with the actual high-pressure cell with its limited scattering angle range an angular dispersive measurement would not be possible. It is also worth noticing the fact that the time of acquisition for a spectrum as a whole is quite small, typically of the order of minutes).

3.3 DATA PROCESSING

As pointed out before, the use of all the available information in a data treatment analysis is extremely important. In the present context, additional pieces of information (apart from the data themselves) are available. There are at least three useful additional pieces of information:

1. There is a fairly high background, and the physics indicate that it is a smooth slowly varying function of energy.
2. The detector system degrades the spectrum in at least two respects.
   (a) The fluorescent Ge K_α X-ray quanta generated in the detector can escape detection giving rise to escape peaks.
   (b) The spectrum is broadened by the detector resolution function.
3. The number of counts in each channel is non-negative.

The most important of these additional pieces of information are the background and the broadening. The escape correction is not extremely important and is in any case fairly straightforward[20].

The problem then reads in terms of the measured number of counts Y_i in a channel i, the background b_i, and the undistorted spectrum f_i:

$$Y_i = \sum R_{ij} f_j + b_i + \epsilon_i, \tag{2}$$

with ϵ_i some (unknown) error term, and R_{ij} the point-spread function.

There is no à priori distribution between signal and background, so some further information is needed. Typically, the background is smooth and slowly varying with energy which suggests setting

$$b_i = \sum_{j=0}^{m} c_j p_j^i, \tag{3}$$

with p_j^i the value of an (orthogonal) polynomial at channel i.

A direct maxent solution of (2) does not seem promising. It turns out that there is a simpler way which follows from the observation that there are regions in the spectra in which obviously there is only background and no signal. A simple least-squares determination of the polynomial coefficients and their appropriate maximal degree as well as an automatic detection of the purely background sections, has been shown to be possible[21]. The result of this background subtraction is shown in fig. 2 for the same data as shown in fig. 1. It is to be noted that this background subtraction may lead to negative counts in

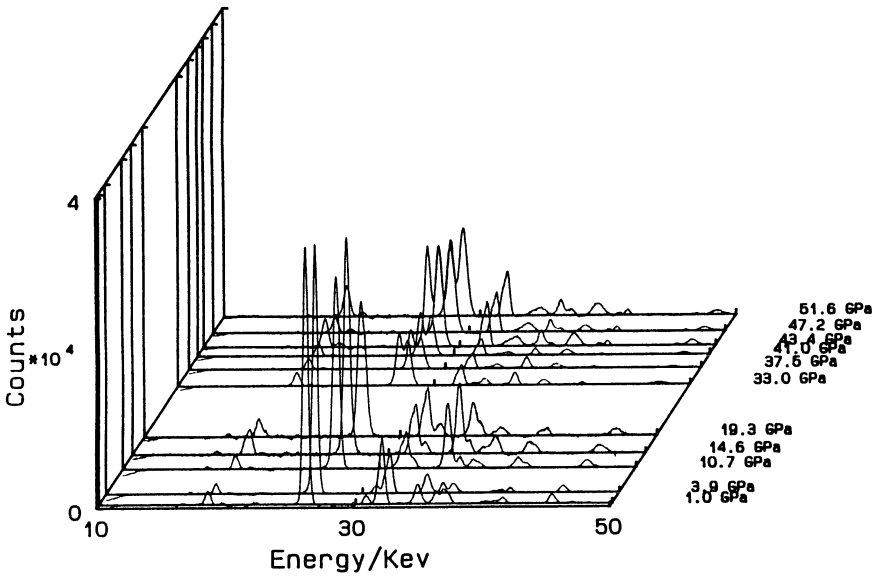


Figure 2: The data of fig. 1 with background subtracted

some channels reflecting the fact that the error in the resulting count number in a given channel is much larger than the one corresponding to the Poisson counting error for counts close to zero.

After backgroundsubtraction a straight forward maxent solution of the still underdetermined resulting equations:

$$y_i = \sum R_{ij} f_j + \epsilon_i, \tag{4}$$

with $y_i = Y_i - b_i$ is easily performed, especially if the statistical proporties of the errors are known (viz. $E(\epsilon_i) = 0$ and $E(\epsilon_i \epsilon_j) = \sigma_i^2 \delta_{ij}$ with $E(\cdot)$ being the statistical expectation), and in principle they are as σ_i^2 can be measured. Likewise, the point-spread function R_{ij} is measurable. The point-spread function turns out to depend on energy, and the expression $\sum R_{ij} f_j$ is not really a convolution. However, in the implementation[3] it is assumed that the point-spread function is only slowly varying with energy and the spectrum is treated in sections assuming a constant point-spread function in each section, so one may take advantage of the simpler convolution method in each section.

The resulting much nicer looking spectra are shown in fig. 3. The scaling is the same as in figs. 1 and 2 but the peak heights are seen much greater, by almost a factor of ten. In fact, the main peak at 1 GPa and 3.9 GPa actually exeeds the maximum height bound of the figure.

The determination of the positions of the peaks as judged from controlled "experiments" and from the fluorescent lines seems to be quite precise (e.g. the position of the fluorescent lines agree to whitin ~50eV compared with the corresponding tabulated values). The intensities have not yet been used, and in the present experiments are not trustworthy anyway due to the possibility af texture.

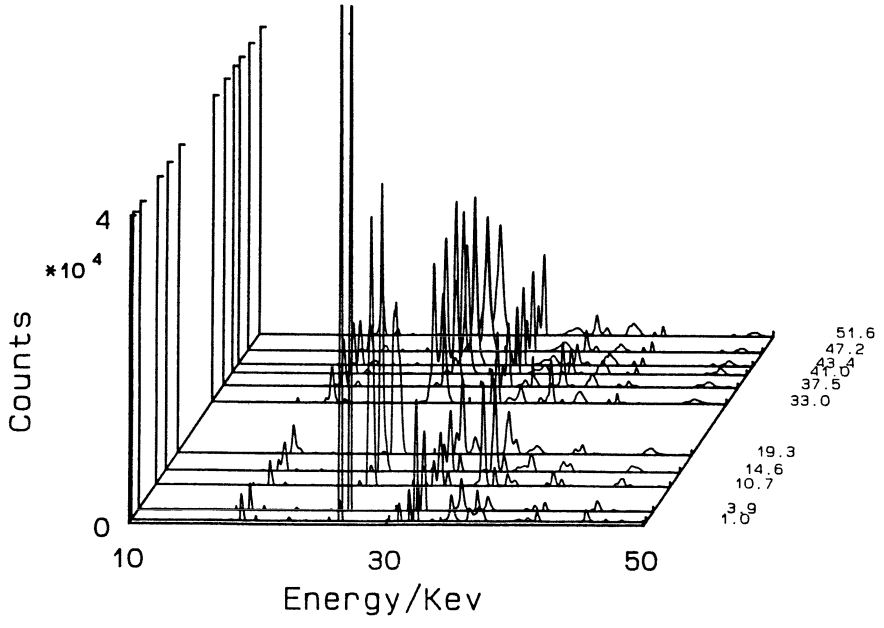


Figure 3: Data as in fig. 1 after deconvolution

3.4 Results.

The main advantage of the present data processing procedure is of course the easier identification of diffraction peaks and fluorescent lines. For example, after deconvolution the Ba K_α line at around 32 keV becomes double, i.e. resolved into a K_{α_1} line and a K_{α_2} line.

An even more interesting observation on the deconvoluted spectrum is that the (1 0 0) peak, (the one just below 20 keV) which at 1.0, 3.9 and 10.7 GPa obviously is a single peak, is split at 14.6 and 19.3 GPa into two which are identified as (0 0 3) and (1 0 0) peaks, with the intensities changing drastically on going from 14.6 to 19.3 GPa. This change in intensity can be due either to texture or to an actual change in the atomic positions in the unit cell - which could well be the case since there seems to be a phase-change occuring between 19.3 and 33 GPa. To decide which of these possibilities is correct, some way of getting rid of texture must be introduced, for instance by rotating the sample. The intensities can then be used for the determination of the atomic positions, and work is in progress using the positions, widths and intensities obtained from the deconvoluted spectra as starting values in a nonlinear least-squares refinement of these parameters.

The importance of the background is illustrated in fig. 4 in which a small part of a spectrum from a $GdBa_2Cu_3O_{9-\delta}$ is shown. Both the raw data and the background subtracted data have been deconvoluted. It should be noted that all the peaks in the d) curve can be identified. The two most prominent peaks just below 50 keV are the Gd K_α lines.

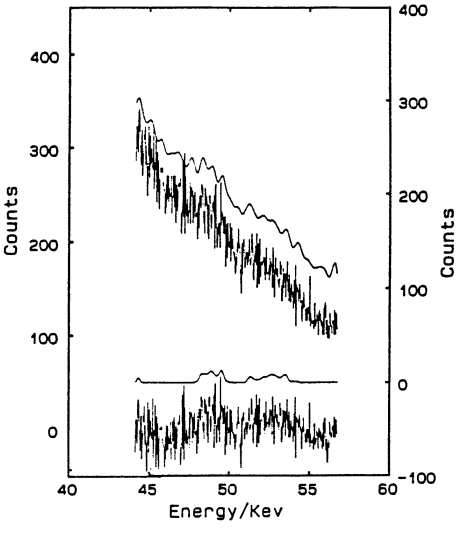


Figure 4: a) Raw data, b) background subtracted data, (both left-hand scale) c) deconvolution of the raw data, and d) deconvolution of the background subtracted data (both right-hand scale).

4 Conclusion

The *ab initio* determination of macromolecular structures is unsolved and is a difficult problem with intensive work going on in various centres, as evidenced by the references. For the powder diffraction problem only one aspect of the application of maxent has been touched upon here, although other aspects could certainly be handled with advantage by Bayesian and maxent methods. One such case is the refinement of atomic positions making use of the intensity data, which is usually done by a Rietveld-type (least-squares) procedure, where a Bayesian approach would probably be more appropriate.

References

[1] Steenstrup S. and Wilkins S.W. On information and complementarity in crystal structure determination. *Acta Cryst.*, **A40**:163–4, 1984.

[2] Skilling J. Classical maximum entropy data analysis. *These proceedings.*

[3] Steenstrup S. Deconvolution in the presence of noise using the maximum entropy principle. *Aust. J. Phys.*, **38**:319–27, 1985.

[4] Woolfson M.M. Direct methods - from birth to maturity. *Acta Cryst.*, **A43**:593–612, 1987.

[5] Gull S.F. and Daniell G.J. Image reconstruction from incomplete and noisy data. *Nature*, **272**:686–90, 1978.

[6] Collins D.M. Electron density images from imperfect data by iterative entropy maximization. *Nature*, **298**:49–51, 1982.

[7] Wilkins S.W., Varghese J.N., and Lehmann M. S. Statistical geometry I. A self-consistent approach to the crystallographic inversion problem based on information theory. *Acta Crys.*, **A39**:47–60, 1983.

[8] Bricogne G. Maximum entropy and the foundations of direct methods. *Acta Cryst.*, **A40**:410–45, 1984.

[9] Navaza J. On the maximum entropy estimate of the electron density function. *Acta Cryst.*, **A41**:232–44, 1985.

[10] Bryan R.K, Bansal M., Folkhard W., Nave C., and Marvin D.A. Maximum-entropy calculation of the electron density at 4 Å resolution of pf1 filamentous bacteriophage. *Proc. Natl. Acad. Sci. USA*, **80**:4728–31, 1983.

[11] Wilkins S.W. and Stuart D. Satistical geometry IV. Maximum-entropy-based extension of multiple isomorphously phased x-ray data to 4 Å resolution for α-lactalbumin. *Acta Cryst.*, **A42**:192–202, 1986.

[12] Gull S.F., Livesey A.K., and Sivia D.S. Maximum entropy solution of a small centrosymmetric crystal structure. *Acta Cryst.*, **A43**:112–7, 1987.

[13] Bryan R.K. and Banner D.W. Maximum entropy calculation of electron density with native and single isomorphously phased data. *Acta Cryst.*, **A43**:556–64, 1987.

[14] Bryan R.K. The maximum entropy method applied to intensity data. *Proc. 6th Pfefferkorn Conference*, Niagara Falls, May 1987. to appear in *Scanning Microscopy Suppl. 2.*

[15] Bricogne G. A Bayesian statistical theory of the phase problem I. A multichannel maximum-entropy formalism for constructing generalized joint probability distributions of structure factors. *Acta Cryst.*, **A44**:517–45, 1988.

[16] Staun Olsen J., Buras B., Gerward L., and Steenstrup S. A spectrometer for x-ray energy-dispersive diffraction using synchrotron radiation. *J. Phys. E: Sci. Instrum.*, **14**:1154–7, 1981.

[17] Jayaraman A. Diamond anvil cell and high-pressure physical investigations. *Rev. Mod. Phys.*, **55**:65–109, 1983.

[18] Buras B., Niimura N., and Staun Olsen J. Optimum resolution in x-ray energy-dispersive diffractometry. *J. Appl. Cryst.*, **11**:137–40, 1978.

[19] Staun Olsen J., Steenstrup S., Johannsen I., and Gerward L. High-pressure studies of the high-temperature superconductors $RBa_2Cu_3O_{9-\delta}$ with R: Y, Eu and Ho up to 60 GPa. *Z. der Physik. In press.*

[20] Steenstrup S. Correction for escape in x-ray spectra measured using a Ge detector. *J. Appl. Cryst.*, **16**:641–4, 1983.

[21] Steenstrup S. A simple procedure for fitting a background to a certain class of measured spectra. *J. Appl. Cryst.*, **14**:226–9, 1981.

MAXIMUM ENTROPY IN CRYSTALLOGRAPHY.

R. K. BRYAN
European Molecular Biology Laboratory,
Meyerhofstrasse 1,
6900 Heidelberg,
West Germany.

ABSTRACT. Measurements of diffracted X-rays give information on the intensities, but not the phases, of the Fourier transform of the electron density of a crystal. Conventional methods for solving macromolecular structures collect data for isomorphous derivatives, and solve explicitly for the phases. Such phases are often of indifferent quality, so interpreting the resultant map in terms of molecular structures is not always straightforward. Using maximum entropy one could hope to produce improved electron density maps from poor isomorphous phases, or, more ambitiously, to use native intensity data only. However, phase ambiguities may not be completely resolved, and the non-convexity of intensity constraints makes considerable demands on numerical algorithms. Moreover, it seems that the underlying molecular structure tends to weaken the assumption of *a priori* uniformity and independence. An alternative strategy for solving the problem would be to incorporate stereochemical information at a much earlier stage than usual, which it is proposed to do by use of a triple correlation function.

1. The Crystallographic Phase Problem.

A crystal is an object which is identical if translated by any of the basis vectors $\mathbf{a}_i$, $i = 1,2,3$, which thus define the unit cell. If $\mathbf{x}$ denotes fractional unit cell coordinates (*i.e.*, the position $\mathbf{r}$ is given by $\mathbf{r} = \sum_{i=1}^{3} x_i \mathbf{a}_i$) so the crystal has period 1 in each axis, the Fourier transform of the electron density ρ is given by

$$F_{\mathbf{h}} = \int_{\mathbf{x} \in \text{unitcell}} \rho(\mathbf{x}) \exp(2\pi i \mathbf{x}.\mathbf{h})\, d^3x \quad \stackrel{\text{def}}{\equiv} \quad \mathcal{F}\rho, \tag{1}$$

and is non-zero only when each component h_i of $\mathbf{h}$ is integral. Measurements of intensities of X-rays diffracted by the crystal give values of $|F_{\mathbf{h}}|^2$, but not of $\phi_{\mathbf{h}}$, the phase of $F_{\mathbf{h}}$. To calculate the density ρ requires that the $\phi_{\mathbf{h}}$ be found, thus posing the 'phase problem'.

No progress can be made without some assumptions about the contents of the unit cell. The obvious one is that of atomicity, which immediately implies a positive electron density. Direct methods exploit this in several ways to give relationships between intensities and phases:- positivity alone gives Karle-Hauptman determinants (Karle & Hauptman, 1950); equal point atoms give the Sayre equation (Sayre, 1952); and the assumption of an independent random distribution of atoms allows calculation of the joint probability density functions of Fourier coefficients (Klug, 1958). These methods have been successful in solving structures of 100 atoms or so, but fail for larger macromolecular problems, which are conventionally (Blundell & Johnson, 1976) solved by the method of multiple isomorphous replacement (MIR). Fourier intensity data are also collected for isomorphous heavy-atom derivatives, the heavy-atom positions deduced, the phases solved

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 213–224.

for explicitly, and hence a unique density map calculated, although a whole range of special solution methods also exist for cases where there is similarity to other solved structures, or where non-crystallographic symmetry provides redundancy in the data. Due to experimental difficulties, the calculated isomorphous phases are often of indifferent quality, in terms of accuracy and resolution, and so building a stereochemically correct molecular model into the density is not always straightforward. The final stage of analysis consists of optimising atomic positions with respect to both stereochemical constraints and the native intensity data. There are often many more degrees of freedom than observations, and the positions of the atoms are not determined by the intensity data alone. So, despite the successes of MIR, there is still a need for improved methods of solving structures using as little data additional to the native intensities as possible. For this reason, maximum entropy is seen as a promising technique in crystallographic analysis.

In this paper, the application of maximum entropy to the phase problem and requirements for numerical algorithms are outlined, followed by a review of some practical calculations on trial data for a small protein in which the quality of the phase information is successively reduced. It will be seen that ordinary maximum entropy is insufficient to compensate for a complete lack of phase information, and that further knowledge of molecular structure must be introduced, which it is proposed to do by means of a multiple-sample prior based on the triple correlation function of known structures.

1.1. APPLICATION OF MAXIMUM ENTROPY.

Maximum entropy may be applied to this problem in a similar way to other image processing problems, although the exact philosophy varies (*e.g.*, Wilkins *et. al.*, 1983, Bricogne, 1984, Livesey & Skilling, 1985, Navaza, 1985, Bryan & Banner, 1987). A statistic measuring the misfit between the experimental data and synthesised data from a trial solution map is defined, and the entropy of the map maximised subject to the statistic indicating a suitable fit to the data, in that the differences between the observed and calculated quantities can be attributed solely to noise. The entropy (Jaynes, 1968) is defined on the suitably discretised density ρ as

$$S(\rho) = -\sum_j \rho_j \log \rho_j / em_j, \tag{2}$$

and $\rho = \mathbf{m}$ has the global unconstrained entropy maximum. It is essential to use this unnormalised form as F_0 cannot be measured. All native and derivative data may be included with the correct weights by using the statistic

$$\chi^2(\rho; I^n, I^{d_i}, F^p) = \sum_{\mathbf{h}} \Big\{ \quad w_{\mathbf{h}}^n \left(|F_{\mathbf{h}}|^2 - I_{\mathbf{h}}^n\right)^2 \tag{3a}$$

$$+ \sum_i w_{\mathbf{h}}^{d_i} \left(|F_{\mathbf{h}} + H_{\mathbf{h}}^{d_i}|^2 - I_{\mathbf{h}}^{d_i}\right)^2 \tag{3b}$$

$$+ w_{\mathbf{h}}^p |F_{\mathbf{h}} - F_{\mathbf{h}}^p|^2 \quad \Big\}, \tag{3c}$$

where $I_{\mathbf{h}}^n$ are the observed native intensities, $I_{\mathbf{h}}^{d_i}$ the observed intensities for the i^{th} derivative, weighted by $w_{\mathbf{h}}^n$ and $w_{\mathbf{h}}^{d_i}$ (usually inverse variances) respectively, $H_{\mathbf{h}}^{d_i}$ the transform of heavy atom contribution to the i^{th} derivative, and the $F_{\mathbf{h}}^p$ phased data, included to take account of Fourier coefficients such as centrics (whose phases are restricted to 0 or π by the space group symmetry) which can be phased reliably by conventional isomorphous replacement. For a large

number M of observations, agreement is achieved when $\chi^2 \leq M$.† If the F_h fit the native data exactly, (3b) is the same as the commonly used expression of Hendrickson & Lattman (1970) for the phase log likelihood distribution at fixed amplitude, and similar to that of Blow & Crick (1959).

1.2. ALGORITHMS.

If phased data only are used (constraint 3c), $\nabla\nabla\chi^2$ is positive semi-definite, surfaces of constant χ^2 are convex, and the entropy maximisation problem has a unique solution (Gull & Daniell, 1978). Taking one term from constraint (3b), $\nabla\nabla\chi^2$ can be expressed in the space of the real and imaginary parts of one Fourier coefficient as

$$4w \begin{pmatrix} 3(F_r + H_r)^2 + (F_i + H_i)^2 - I & 2(F_r + H_r)(F_i + H_i) \\ 2(F_r + H_r)(F_i + H_i) & (F_r + H_r)^2 + 3(F_i + H_i)^2 - I \end{pmatrix}, \tag{4}$$

where r and i stand for real and imaginary part, and other subscripts have been dropped. Diagonalising gives eigenvalues $4w(3|F+H|^2-I)$ in the direction of $F+H$ and $4w(|F+H|^2-I)$ in the orthogonal ('phase') direction. Setting $H = 0$ gives the results for native data. More generally, with MIR data there will be contributions to this matrix from each term in (3), but it is clear that $\nabla\nabla\chi^2$ may have negative eigenvalues, and the problem is no longer convex.

Non-convexity removes uniqueness of the solution and also demands a more careful analysis of the criteria for a *local* optimum. The Kuhn-Tucker conditions for a local constrained maximum are

1. The gradients of S and χ^2 are parallel, $\alpha\nabla S = \nabla\chi^2$, $\alpha \geq 0$. In the convex case this is also sufficient.
2. $\forall$ d such that $\mathbf{d}.\nabla\chi^2 = 0$, $\mathbf{d}^t\nabla\nabla Q\mathbf{d} \leq 0$, where Q is the (rescaled) Lagrangian $\alpha S - \chi^2$, *i.e.*, there are *no* directions *along* the constraint surface that give an improved S.

As there are no conditions on the curvature of Q in the gradient direction, the optimum may be a saddlepoint of Q at fixed α, and methods based on its unconstrained maximisation may not work. This difficulty gave part of the motivation for modelling S and χ^2 as separate quadratic functions, rather than working just with the Lagrangian (Bryan, 1980, Skilling & Bryan, 1984). Such a formulation allows one to remain on a surface of constant χ^2 irrespective of convexity. Continuing the analysis, if $\mathbf{v}_j$ is an eigenvector of $\nabla\nabla\chi^2$ with respect to $-\nabla\nabla S \equiv \mathrm{diag}\{1/\rho\}$, eigenvalue λ_j, normalised to $\mathbf{v}_i\mathrm{diag}\{1/\rho\}\mathbf{v}_j = \delta_{ij}$, then $\nabla\nabla Q$ becomes $-\alpha I - \mathrm{diag}\{\lambda_j\}$ in the $\{\mathbf{v}_j\}$ basis. Assuming ordered eigenvalues, if $\lambda_1 > -\alpha$ the second optimality condition is satisfied, whereas if $\lambda_2 \leq -\alpha$ it certainly will not be. For the intermediate case, $\lambda_1 \leq -\alpha < \lambda_2$, the eigenvalues in the subspace orthogonal to the gradient must be examined, which may be done *via* the characteristic polynomial of $\nabla\nabla Q$ (Skilling, 1986b).

Non-uniqueness is less easily dealt with. The multiplicity of possible solutions must depend on the initial object. Consider, at one extreme, intensity data for a single point source. The only positive solution is with all phases zero (or a trivial translation). Sparse distributions of point sources have indeed been successfully reconstructed by maximum entropy (Gull & Daniell, 1978, Bryan, 1980, Bryan & Skilling, 1986). Alternatively, a large, diffuse object can have almost any phase assigned to the weaker, higher resolution Fourier coefficients, and potentially many local entropy maxima. The practical problems of crystallography lie between these extremes.

1.2.1. Standard algorithm. First, the algorithm for convex problems (Skilling & Bryan, 1984) is briefly summarised. The idea is to follow the locus of S maxima on surfaces of constant χ^2 in

† Skilling and Gull have shown at this meeting that $\chi^2 = M$ is incorrect, and due to a frequentist interpretation of the noise. However, the computational results presented here should not be significantly affected, as the final map is more strongly dependent on the obtained phases.

a series of finite steps, over decreasing values of χ^2 until its correct value is attained. At each iteration the increment $\delta\rho$ is found as a linear combination of a set of a small number of search directions $\mathbf{e}_\mu$, with $\delta\rho = x^\mu \mathbf{e}_\mu$, for some set of coefficients x^μ. Quadratic models of S and χ^2 are constructed in the subspace spanned by the search directions, and to ensure accuracy of this approximation, a limit is put on the step length at each iteration, by imposing $|\delta\rho|^2 \leq l^2$, where distances are calculated using the second derivative of S as a metric, so $|\delta\rho|^2 = \sum_i \delta\rho_i^2/\rho_i$. The problem of selecting the x^μ is one of quadratic optimisation in the subspace. For convex problems, the $\mathbf{e}_\mu$'s are usually constructed as the contravariant gradients of S and χ^2, *i.e.*, $\mathrm{diag}\{\rho\}\nabla S$ and $\mathrm{diag}\{\rho\}\nabla\chi^2$, plus $\mathrm{diag}\{\rho\}\nabla\nabla\chi^2$ acting, perhaps repeatedly, on them. The only computations required in the full space are Fourier transforms and vector operations.

1.2.2. Finding Eigenvectors. If χ^2 surfaces are non-convex, S may have saddle points in the surface, and give rise to bifurcations in the solution path (Bryan, 1980). A multisolution strategy may be used to examine all optima which can be reached in this way (Bricogne, 1984, Gilmore *et. al.*, 1988), although the topology of the space may be sufficiently complicated that other, completely isolated, optima exist (Skilling, private communication). Nevertheless, negative curvature directions must be investigated even if only to establish local optimality. The analysis above shows that the most important directions are the eigenvectors $\nabla\nabla\chi^2$ with respect to $\mathrm{diag}\{1/\rho\}$ with the most negative eigenvalues, and one approach taken (Bryan, 1980, Bryan & Skilling, 1986) has been to select suitable candidate directions, and include them in the search space. The algorithm of §1.2.1 operates as before, but with the distance limit now playing an essential rôle when moving in negative curvature directions. A slightly more sophisticated approach to selecting directions, used for the calculations in Bryan & Banner (1987) and Bryan (1988b, c) is described here.

If $\mathbf{e}$ satisfies $\nabla\nabla\chi^2\mathbf{e} = \lambda\mathrm{diag}\{1/\rho\}\mathbf{e}$ with $\lambda < 0$, then $\mathbf{e}^t\nabla\nabla\chi^2\mathbf{e} < 0$. Although the spaces of eigenvectors with negative eigenvalue with respect to the entropy and Euclidean metrics are not identical, as a first approximation those with respect to a Euclidean metric are identified. They are easily found in the complex space of each Fourier coefficient, as in (4), leading to the Fourier-space Euclidean eigenvectors $\mathbf{A}_k = c_k\delta_{\mathbf{h}\mathbf{h}_k}$, eigenvalues $R_{\mathbf{h}_k}$, where the c_k are complex coefficients of unit magnitude. The $\nabla\nabla\chi^2$ and $\mathrm{diag}\{1/\rho\}$ matrices in the subspace spanned by K such vectors can be formed as follows, which is the Fourier space equivalent of the standard method, and similar to a suggestion of Bricogne (1984). If $\mathbf{a}_k$ are the inverse Fourier transforms of the $\mathbf{A}_k$, $k = 1, \ldots, K$, then

$$S_{kl} = \mathbf{a}_k^t \mathrm{diag}\{1/\rho\}\mathbf{a}_l = c_k c_l \mathcal{F}^*_{\mathbf{h}_k+\mathbf{h}_l}(1/\rho), \tag{5}$$

and

$$C_{kl} = \mathbf{a}_k^t \nabla\nabla\chi^2 \mathbf{a}_l = |c_k|^2 R_{\mathbf{h}_k}\delta_{kl}, \tag{6}$$

so that the subspace C matrix is diagonal, and both matrices may again be computed by Fourier transforms and purely vector operations in the full space. Complex conjugates and space group symmetries must be taken into account in these expressions.

The subspace C matrix can be diagonalised with respect to the S matrix by standard linear algebra methods, yielding the Fourier coefficients of the approximate negative curvature eigenvectors. $K = 50$ has been used regularly, and gives a reasonable compromise between cost of the Fourier transforms used in the construction of the standard search directions and the linear algebra in the eigenvector routine. Using 8-10 search directions in total, the eigenvector routine is called at intervals of a few iterations, giving a library of directions, 2-3 of which can supplement the main search space at each iteration, together with negative curvature eigenvectors retained from the previous iteration. This procedure has proved to be reasonably reliable, in that if a run is stopped and then restarted, so that the current map is unchanged but all accumulated

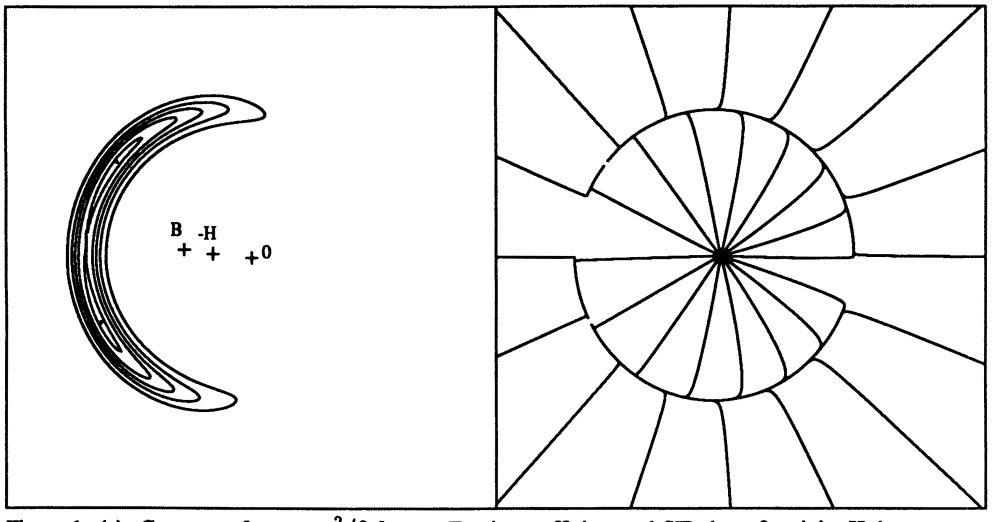


Figure 1. (a). Contours of $\exp -\chi^2/2$ for one Fourier coefficient and SIR data. 0, origin, H, heavy atom vector, B 'best' value. (b). Orthogonal trajectories of (a).

eigenvector information is lost, the wanted eigenvectors are usually found again within a few iterations.

2. Applications.

Some results of applying the above algorithm are reviewed and discussed here. Complete phase information gives the same Fourier problem with a unique solution as that of Gull & Daniell (1978), and maximum entropy provides the expected improvement in quality (*e.g.*, Bricogne, 1984, Wei, 1985, Wilkins & Stuart, 1986). Here, in §2.1 maximum entropy is shown to give useful solutions to some problems, which, through incompleteness of the phase information, are insoluble by classical methods, whereas with amplitude data only (§2.2) its limitations become apparent.

2.1. INCOMPLETE PHASE INFORMATION.

2.1.1. Single isomorphous replacement. Assuming, for the moment, that the heavy atom positions have been established, data for the native structure and one derivative do not uniquely determine the phase. Fig. 1a shows the contours of likelihood, $\exp -\chi^2/2$, for one Fourier coefficient in the argand diagram. There are two regions of high likelihood, each with the correct native amplitude, but with the phases symmetrical positioned about the heavy-atom phase, one of which is the correct phase. Conventionally, (Blow & Rossmann, 1961) the average over this distribution is taken as the so-called 'best' Fourier coefficient, but a map synthesised from these is often uninterpretable. Maximum entropy may be used to make a selection between the two high-likelihood possiblities, consistent across the different Fourier coefficients. However, the shape of the constraint leads to further algorithmic complications. If the orthogonal trajectories of the χ^2 contribution for a single Fourier coefficient, fig. 1b, are examined, they are seen to be more-or-less radial until very close to the correct intensity. Consequently, during the iterative solution, the Fourier coefficients increase in amplitude with virtually no change in phase until they are near the correct amplitude, or are deflected by an entropy gradient. If the starting map is flat, $F = 0$, the direction of departure from the origin is down the gradient of χ^2, proportional to $(|H|^2 - I^d)H$,

hence either in the direction of H, or opposite, and depending on the relative amplitudes of H, I^d and I^n, it could be either towards the 'best' phase, or 180° away, giving aligned phases and a map with a single large peak. This inevitably results in computational problems, either because a large number of iterates will be required eventually to move in phase at constant amplitude, or because the algorithm may stop at an incorrect optimum without having fitted the data.

Previously (Bryan *et. al.*, 1983), this problem was avoided by imposing an upper bound on ρ *via* a Fermi-Dirac entropy, thus suppressing large peaks and avoiding extreme phase alignment. The same calculations can be performed if the phases are biased to start with, by taking origin and enantiomorph defining Fourier coefficients plus those with better-defined 'best' phases (*e.g.*, if the two most likely phases differ by less than 45°), calculating a map with them, and using it temporarily as a prior in a calculation using all the data. Clearly, many variations along these lines are possible. More recently (Bryan, 1988b), a calculation with a centrosymmetric heavy atom has been performed, (which conventionally would give a centrosymmetric map, totally uninterpretable), showing that this ambiguity can also be resolved.

2.1.2. Unknown heavy-atom positions. So far, the heavy-atom Fourier coefficients have been fixed. The step of deducing their positions can be avoided if a second map σ is introduced to represent the heavy-atom density, so now $H_{\mathbf{h}} = \mathcal{F}\sigma$. The total entropy of both maps is maximised, again subject to the χ^2 constraint. It is important that the ρ and σ maps have the correct relative weights in the entropy. The analysis (Bryan, 1988a) may be performed if the numbers and types of atoms each map represents is known. If ρ and σ now represent the number density of atoms, then $F_{\mathbf{h}} = Z_\rho(\mathcal{F}\rho)_{\mathbf{h}}$ and $H_{\mathbf{h}} = Z_\sigma(\mathcal{F}\sigma)_{\mathbf{h}}$, where Z_ρ and Z_σ are the respective atomic numbers, and the priors for ρ and σ, integrated over the respective maps, should be proportional to the numbers of atoms the maps represent. Anomalous scattering effects may also be allowed for by using a complex Z.

This method has been demonstrated (Bryan, 1988b) on a small trial structure (160 light atoms, $Z = 8$, 1 heavy atom, $Z = 32$), using a small set of starting phases, and gives a σ map with a single peak at the correct position, and the correct ρ map. The knowledge that σ usually consists of only a few atoms is also important, and the solution should be completed by building an atomic model into the map. More generally, for the MIR problem, a further σ map should be introduced for each derivative.

2.1.3. Phase Extension. The formation of a heavy-atom derivative may disrupt the structure locally, or, if it causes changes in inter-molecular contacts, alter the unit-cell size. Thus the phases determined by isomorphous replacement are often accurate only to lower resolution than the diffraction data itself, so the 'phase extension' problem then arises. The *ab initio* problem is a limiting case, with only origin and enantiomorph defining phases specified. There have been many approaches suggested for phase extension, one of the most fruitful being solvent flattening, requiring interpretation of the low resolution map into molecule and solvent regions, and the seeking of a solution for the phases which leaves the solvent region uniform. Alternatively, one may try to extend phases directly, without preliminary interpretation of the structure, using constraints on Fourier coefficients at low, and on intensities at high, resolution. There is some scope for varying the exact details; *e.g.*, whether the result of a lower resolution calculation with Fourier coefficients is used as a prior map; whether the intensity data are introduced all at once, or successively, according to resolution, or to some other criterion such as the accuracy of the predicted intensity (Gull *et. al.*, 1987). The final results will inevitably depend heavily on the predictions of the high-resolution phases from the low-resolution data, so the accuracy of these predictions before the intensity constraints are applied will also be investigated.

The results of some calculations performed on synthesised data are presented here. The data were calculated from a solved structure, that of scorpion neurotoxin 3 (Almassy *et. al.*, 1983), which has been refined to high resolution. This small protein has 65 residues and 72 bound

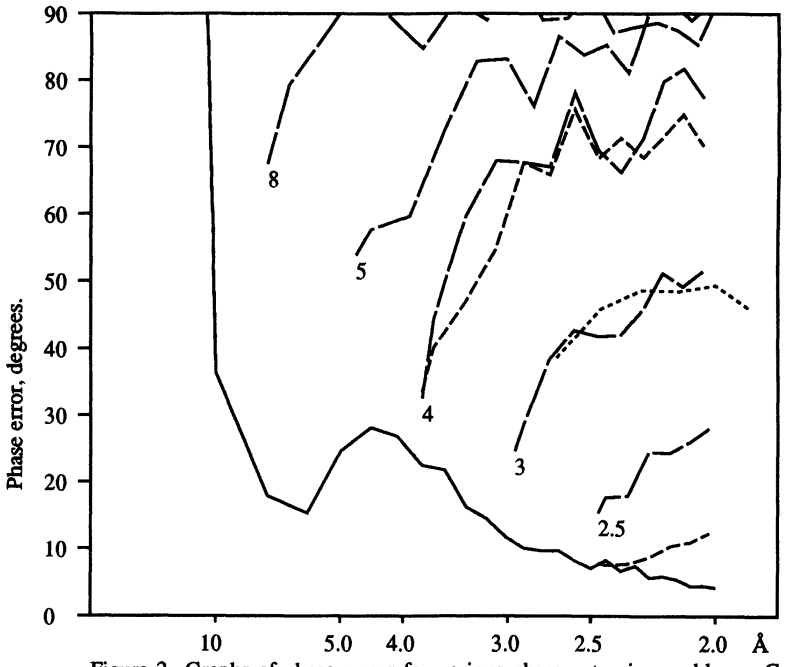


Figure 2. Graphs of phase errors for various phase extension problems. Continuous: mean intensity as a function of resolution (arbitary scale). Long-short dash: mean phase error of extrapolated Fourier coefficients using data with phases to given resolution only. Dotted: for comparison, the results of Bricogne (1984). Short dashes: mean phase error after intensity constraints applied.

solvent molecules, giving a total of around 600 atoms, whose coordinates were obtained from the Brookhaven database. Fig. 2 summarises the results. For comparision, the results of Bricogne (1984) for phases predicted beyond 3Å for the somewhat smaller protein Crambin are also plotted, and are in close agreement. It is seen that the accuracy of predicted phases depends very strongly on the cutoff resolution of the Fourier coefficient data provided. Even medium resolution data of 5Å does not give good predictions. In two cases, intensity constraints at higher resolution were then applied. Extending from 2.5Å gave an excellent result, with the additional constraints pulling the phases even closer to the true values. At 4Å, there is a modest improvement at slightly higher resolution, but essentially none beyond 3Å, where the phases are hopelessly wrong. It seems to be necessary to have correct phases to 3.5Å or so before this procedure gives reasonable good phases. Why should this resolution be critical? As pointed out by several authors, the entropy penalty on introducing new data is least if the phases cause new density to line up with old. Upon examination of correctly-phased maximum entropy maps at 2Å and 4Å resolution, fig. 3, it is seen that the peaks of the 4Å map are centred, not at atomic positions, but more on atomic groups. The predicted phase for higher resolution data causes an alignment with these peaks, and not with the correct atomic positions.

2.2. NO PHASE INFORMATION.

As discussed in §1, the pure phase problem is likely to give many local entropy optima, so an *ab initio* calculation will be avoided here, and instead the question of whether the true structure is indeed near *some* optimum will be adressed. Previously (Bryan & Banner, 1987), a

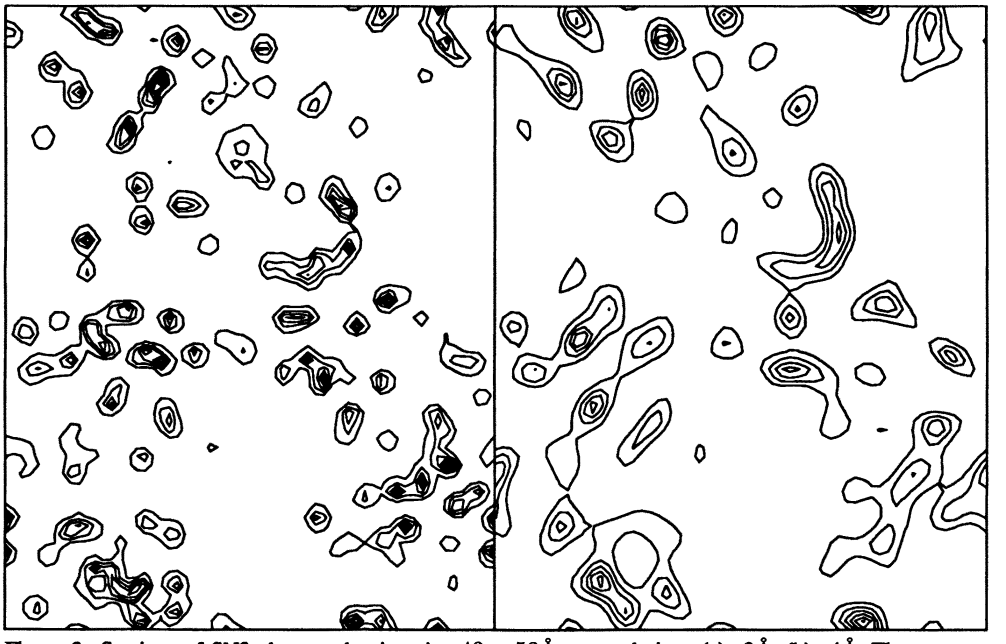


Figure 3. Sections of SN3 electron density, size 42 × 52Å, at resolutions (a). 2Å, (b). 4Å. The contour interval in (a) is twice that in (b).

maximum entropy map was calculated from synthetic 3Å Fourier coefficients, using constraint (3c). The phases of the data were then forgotten, and only the intensity contraint (3a) used. If maximum entropy really compensated for the loss of phase information, then the same map should be a maximum of entropy subject to the new constraint. However, it was not, and a path of monotonically increasing entropy at constant χ^2 could be followed until a maximum was attained, with a considerable phase shift. Repeating the calculation with the neurotoxin structure again gave large phase shifts, the amplitude weighted average being some 50°, and an essentially uninterpretable map. Clearly, maximum entropy does not compensate for the loss of phase information at this resolution. If the same test is performed at better resolution, the phase shift is reduced, to 35° at 2Å, due possibly to the stabilisation effect of a greater effective volume of zero density in the inter-molecular spaces.

2.3. DISCUSSION.

Why is there such a lack of success when the data contain little phase information? Perhaps, because a reconstruction of point sources will have fewer ambiguities than extended objects, it would be better to calculate the distribution of atomic positions, rather than the electron density itself. For atoms at rest, this could be done by multiplying the transform of the number density of atoms by the atomic scattering factor (Fourier transform of atomic electron distribution), and then comparing with the observed data, similar to §2.1.2., but with Z now a function of h. Direct methods indeed work with 'normalised structure factors', representing the estimated intensities of point scatterers, rather than extended atoms. Such normalisation is practicable only if the atoms all have the same effective shape, which is a reasonable assumption for small molecules. In contrast, the thermal motion of atoms in macromolecules is much larger, and moreover varies greatly between different parts of the structure, such that exposed sidechains are often invisible even in correctly-phased maps. Thus no simple normalisation will give intensities representing

the scattering from point atoms. A theoretically appealing solution to this would be to have a series of densities, indexed by temperature, to which the appropriate factors are applied before combining to give the effective density. To compute many 3D maps is, however, unappealing. Perhaps one could go further:- maps of positions and orientations of amino-acids, etc, trading greater information about the structures in the map for a more complicated parameterisation.

The very failure of maximum entropy shows that 'hidden' constraints are acting (Jaynes, 1982). It is well-known that the density is non-uniform, in that atoms are separated by specific bond lengths, that there are well-defined angles between bonds, with a further hierarchy of larger-scale structure, characteristic of proteins, nucleic acids, *etc.* Applying maximum entropy directly to the problem means that *a priori* independence of the density at different points is assumed, and also uniformity since a flat prior must be used, as the position of the molecule is initially unknown.

These assumptions are seen to be broken at two length scales; at large scales, molecules pack together with disordered solvent filling the spaces, and at small lengths, there is the atomic bonding pattern. Taking these effects into account should mean that information about molecular conformation is used at an early stage in the phasing procedure, and not just at the end, when trying to build a model into an already-phased density.

3. Triple Correlations.

Previously (Bryan, 1988c), the use of priors based on second-order correlations of atomic positions in the phase problem has been considered. This function is itself the Fourier transform of the intensities, and therefore introduces no information that is not in principle measurable, except that many protein crystals do not diffract well enough for these data to be collected. Although the one-dimensional examples presented showed the forcing of a density into atomic form, when extended to three-dimensions the spherically symmetric prior density that results is of little use (unpublished results). Inspection of molecular structures shows that bond angles, as well as bond lengths, fall at certain well-defined values. Considerable further information should therefore be contained in the triple correlation,

$$C^{(3)}(\mathbf{p},\mathbf{q}) = \int \tau(\mathbf{x})\tau(\mathbf{x}+\mathbf{p})\tau(\mathbf{x}+\mathbf{q})\,d\mathbf{x}, \tag{7}$$

or, in terms of Fourier transforms, $\tilde{C}^{(3)}_{\mathbf{u},\mathbf{v}} = \tilde{\tau}_{\mathbf{u}}\tilde{\tau}_{\mathbf{v}}\tilde{\tau}_{-(\mathbf{u}+\mathbf{v})}$, where τ represents a typical solved structure. $C^{(3)}$ is a function of two vectors, and computationally quite unmanagable. Moreover, for use in constructing a prior, a rotationally as well as translationally invariant function is required, which can be obtained by averaging $C^{(3)}$ over all directions of $\mathbf{p}$ and over the azimuthal angle of $\mathbf{q}$ about $\mathbf{p}$, leaving a function of the lengths p, q and the included angle ϕ only (or, in a more symmetric form, the lengths of the three sides of the triangle).

The triple correlation has been calculated for several solved structures, and fig. 4 shows a typical plot, calculated from atomic coordinates, rather than density, to show the features more clearly. The first, at 1.3-1.5Å, and second, around 2.4Å, nearest neighbour positions are closely defined, with appropriate bond angles. With p large, beyond 6Å, the function becomes almost independent of ϕ, except near $(p,p,1)$, and as a function of q, closely related to the second-order correlation. With both p and q large, the function is modulated by the effects of molecular size, but is otherwise fairly smooth. There is no apparent long-range order in the atomic positions. Thus there are three regimes of interest:- short distances, up to perhaps three bond lengths (say 4Å), where atomic positions are highly correlated; medium distances, up to a few tens of Å, where the atomic positions are essentially uncorrelated, but have a constant average density; and large distances, greater than the molecular size, where the the triple correlation takes a constant value related to the relative volumes occupied by molecules and solvent.

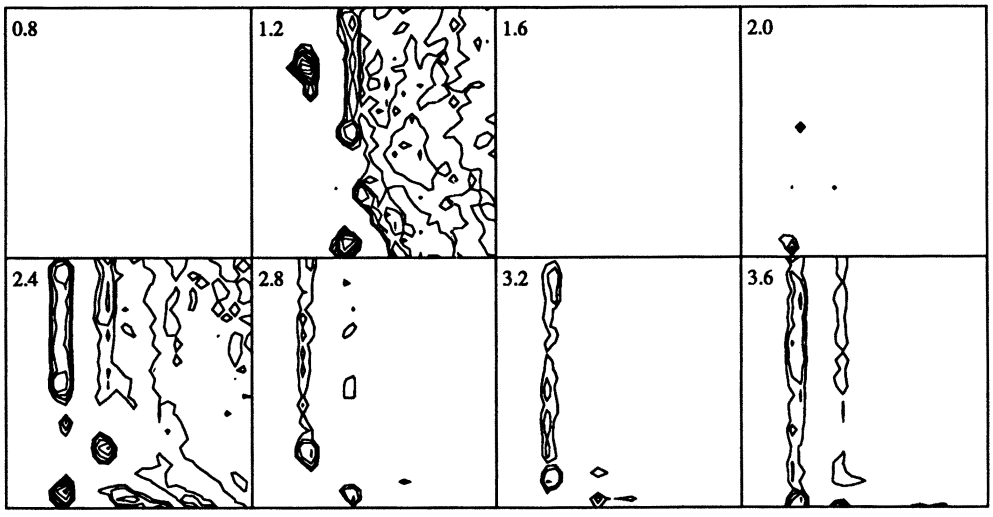


Figure 4. Rotationally averaged triple correlation calculated from atomic coordinates. The q values are given for each section, and in each section p ranges over 0–6Å horizontally, and $\cos\phi$ over -1–1 vertically downwards. The origin spike and $(p, p, 1)$, $(p, 0, \cos\phi)$ and $(0, q, \cos\phi)$ elements (essentially the double correlation) have been suppressed. The contour heights are logarithmic, with the first at twice the average value for large p and q.

3.1. EFFECTIVE PRIOR.

The effective prior $\mathrm{m_{eff}}$ is given (Skilling, 1986a) by the biconvolution

$$\log m_{\mathrm{eff}}(\mathbf{x}) = \int \rho(\mathbf{p})\rho(\mathbf{q}) \log C^{(3)}(\mathbf{x}-\mathbf{p}, \mathbf{x}-\mathbf{q})\, d\mathbf{p}\, d\mathbf{q}, \tag{8}$$

which is most time consuming to compute, requiring either N^9 operations in real-space, or, upon Fourier transformation,

$$(\mathcal{F} \log m_{\mathrm{eff}})_{\mathbf{u}} = \sum_{\mathbf{v}} F_{\mathbf{v}} F_{\mathbf{u}-\mathbf{v}} (\mathcal{F} \log C^{(3)})_{\mathbf{v},\mathbf{u}-\mathbf{v}}, \tag{9}$$

which takes N^6 operations. If the range of integration in the real-space expression (8) is restricted to a volume V, the calculation requires only N^3V^2 operations, which can be performed in a reasonable time for small, but useful, V. To get similar short-distance information into a Fourier space evaluation of the biconvolution, a sum to high resolution is required.

If all of $\mathrm{m_{eff}}$, ρ and C are assumed to be constants plus some variation (using μ, ψ, Γ for the variations), then S may be represented in terms of their Fourier transforms as

$$S = -\tfrac{1}{2} \sum_{\mathbf{k}} |\tilde{\psi}_{\mathbf{k}}|^2 + \tfrac{1}{6} \sum_{\mathbf{kl}} \tilde{\psi}_{\mathbf{k}} \tilde{\psi}_{\mathbf{l}} \tilde{\psi}_{-(\mathbf{k}+\mathbf{l})} + \sum_{\mathbf{k}} \tilde{\psi}_{\mathbf{k}} \tilde{\mu}_{-\mathbf{k}} + \ldots, \tag{10}$$

where

$$\tilde{\mu}_{\mathbf{k}} = 2\tilde{\psi}_{\mathbf{k}} \tilde{\Gamma}_{\mathbf{k},0} + \sum_{\mathbf{l}} \tilde{\psi}_{\mathbf{l}} \tilde{\psi}_{\mathbf{k}-\mathbf{l}} \tilde{\Gamma}_{\mathbf{l},\mathbf{k}-\mathbf{l}}, \tag{11}$$

If Γ is flat, *i.e.*, no triple correlation effect, the second-order term is independent of phase and each term in the third-order expression is a maximum if $\phi_{\mathbf{k}} + \phi_{\mathbf{l}} + \phi_{-\mathbf{k}-\mathbf{l}} = 0$, which is the usual triplet phase relation, well known in direct methods. Including a non-uniform Γ gives

$$S = -\tfrac{1}{2} \sum_{\mathbf{k}} |\tilde{\psi}_{\mathbf{k}}|^2 (1 + 4\tilde{\Gamma}_{\mathbf{k},0}) + \sum_{\mathbf{kl}} \tilde{\psi}_{\mathbf{k}} \tilde{\psi}_{\mathbf{l}} \tilde{\psi}_{-(\mathbf{k}+\mathbf{l})} (\tfrac{1}{6} + \tilde{\Gamma}_{\mathbf{l},-\mathbf{k}-\mathbf{l}}) + \ldots, \qquad (12)$$

so the relative weighting of the terms now depends on the triple correlation, with possibly even a reversal of the expected triplet phase. The rotationally averaged triple correlation is real, but this need not be the case in general, so phase information could be imposed. Higher order correlations will similarly affect the quartet and higher order multiplets. Hauptman (1964) has performed an extremely lengthy Fourier space calculation to derive the distribution of triplets if information about atomic bond lengths is used, and also shows that a modified triplet relationship may be derived. The current approach is very much more flexible; any information to this order may be used directly in maximum entropy with no further calculation of phase distributions.

3.2. A PRELIMINARY APPLICATION.

The 'local' formulation of the problem can be implemented as follows. If we assume the correlation function is uniform when both p and q are large, taking the value C_∞, then $\log C = \log C/C_\infty + \log C_\infty$, whose second term, after the biconvolution (8), gives a constant, and the first is non-zero only within our 'local' volume. The p-small-q-large regime can be similarly taken care of, and introduces a second-order correlation. The correct scaling in this expression is crucial, as $\mathrm{m}_{\mathrm{eff}}$ results from the exponentiation of the biconvolution. This scaling, corresponding to powers of $\mathrm{m}_{\mathrm{eff}}$, was examined in the second-order case in Bryan (1988c), and was found to give a critical phenomena as it was varied.

Even with the local formulation, calculating triple correlations is computationally intensive, and it is not yet really practicable to re-compute the prior each iteration, as one would like to. Updating it every few iterations is more reasonable. Perhaps a suitable criterion would be when the relative entropy has changed by more than a certain amount, although this is still to be investigated.

This method has been applied to a very simple example, consisting of a slightly distorted ring of six gaussian atoms in two dimensions, fig. 5a, with its centrosymmetry ignored. A naïve application of maximum entropy using only two origin-fixing phases, intensity data and a flat prior stops before fitting the data, with an obviously incorrect solution, fig. 5b. The central peak is much larger than the others. Applying the triple correlation (this time calculated for 2D Gaussian atoms) at intervals during the calculation (in fact, only three times in all) to create a new prior gave the results in fig. 5c and d, which speak for themselves.

References.

Almassy, R. J., Fontecilla-Camps, J. C., Suddath, F. L. & Bugg, C. E. (1983). *J. Mol. Biol.*, **170**, 497–522.

Blow, D. M., & Crick, F. H. C. (1959). *Acta Cryst.*, **12**, 794–802.

Blow, D. M., & Rossmann, M. G. (1961). *Acta Cryst.*, **14**, 1195–1202. Correction. *Acta Cryst.*, **15**, 1060.

Blundell, T.L., & Johnson, L.N. (1976). *Protein Crystallography*. New York: Academic Press.

Bricogne, G. (1984). *Acta Cryst.*, **A40**, 410–445.

Bryan, R. K. (1980). *Maximum Entropy Image Processing*. PhD Thesis, University of Cambridge.

Bryan, R. K. (1988a). *Acta Cryst.*, **A44**, 672–677.

Bryan, R. K. (1988b). *Scanning Microscopy Suppl.*, **2**, 99–105.

Bryan, R. K. (1988c). In *Maximum-Entropy and Bayesian Methods in Science and Engineering (Vol. 2)*, ed. Gary J. Erickson & C. Ray Smith, pp. 155–169. Dordrecht: Kluwer Academic.

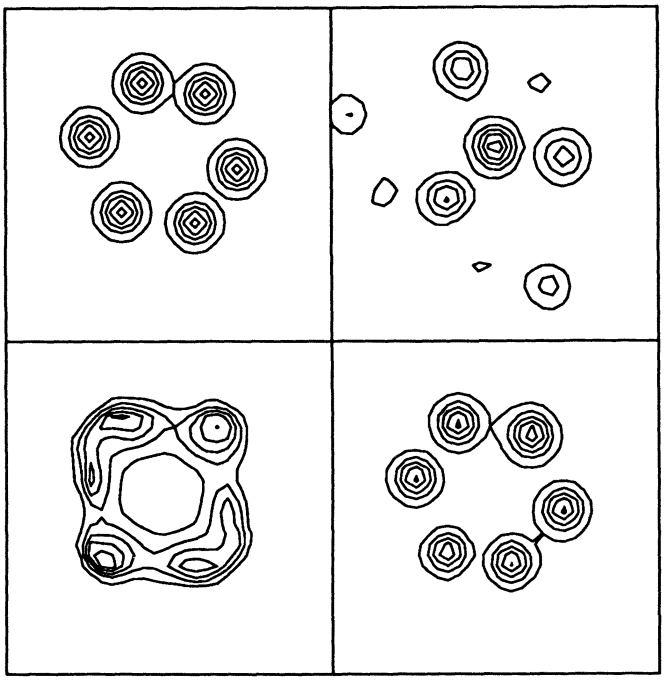


Figure 5. (a). Original map of six Gaussian atoms. (b). Unsuccessful attempt at solution by plain maximum entropy. (c) & (d). Final triple correlation prior and map.

Bryan, R. K. & Banner, D. W. (1987). *Acta Cryst.*, **A43**, 556–564.
Bryan, R. K., Bansal, M., Folkhard, W., Nave, C. & Marvin, D. A. (1983). *Proc. Natl. Acad. Sci. USA*, **80**, 4728–4731.
Bryan, R. K., & Skilling, J. (1986). *Optica Acta*, **33**, 287–299.
Gilmore, C. J., Henderson, K. & Bricogne, G. (1988). This meeting.
Gull, S. F. & Daniell, G. J. (1978). *Nature*, **272**, 686–690.
Gull, S. F., Livesey, A. K. & Sivia, D. S. (1987). *Acta Cryst.*, **A43**, 112–117.
Hauptman, H. (1964). *Acta Cryst.*, **17**, 1421–1433.
Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.*, **B26**, 136–143.
Jaynes, E. T. (1968). *IEEE Trans.*, **SCC–4**, 227–241.
Jaynes, E. T. (1982). *Proc. IEEE*, **70**, 939–952.
Karle, J. & Hauptman, H. (1950). *Acta Cryst.*, **3**, 181–187.
Klug, A. (1958). *Acta Cryst.*, **11**, 515–543.
Livesey, A. K. & Skilling, J. (1985). *Acta Cryst.*, **A41**, 113–122.
Navaza, J. (1985). *Acta Cryst.*, **A41**, 232–244.
Sayre, D. (1952). *Acta Cryst.*, **5**, 60–65.
Skilling, J. (1986a). In *Maximum Entropy and Bayesian Methods in Applied Statistics*, ed. James H. Justice, pp. 156–178. Cambridge: Cambridge University Press.
Skilling, J. (1986b). *ibid.*, pp. 179–193.
Skilling, J. & Bryan, R. K. (1984). *Mon. Not. R. astr. Soc.*, **211**, 111–124.
Wei, W. (1985). *J. Appl. Cryst.*, **18**, 442–445.
Wilkins, S. W. & Stuart, D. (1986). *Acta Cryst.*, **A42**, 197–202.
Wilkins, S. W., Varghese, J. N., & Lehmann, M. S. (1983). *Acta Cryst.*, **A39**, 49–60.

A MULTISOLUTION PHASE DETERMINATION METHOD IN X-RAY CRYSTALLOGRAPHY

Colin Bannister[†], Gerard Bricogne[‡] and Christopher Gilmore[†]

[†]Department of Chemistry, University of Glasgow, Scotland G12 8QQ
[‡]LURE, Bâtiment 209D, 91405 Orsay, France.

Abstract

The maximum entropy method offers several advantages over traditional direct methods in the *a-priori* solution of crystal structures. These include the full use of all invariants at every point without their explicit generation; the use of a non-uniform distribution of atoms, $q^{ME}(\mathbf{x})$, which is constantly updated guaranteeing that the aproximate joint probability distribution of structure factors remains valid even for large deviations from uniformity; the natural incorporation of the variances of the structure factors, a tolerance towards errors in the intensity data, and a stability that is independent of data resolution. In *ab-initio* studies on small organic molecules, starting with origin defining reflections only, the maximum entropy method is used to generate $q^{ME}(\mathbf{x})$ which can then be used as a source of new phase information *via* extrapolation, and this new information is used to update $q^{ME}(\mathbf{x})$ in a cyclic fashion. When the extrapolation process has exhausted the currently assumed phase information, strong unphased structure factors are given permuted phases; this recenters the asymptotic expansion for the joint probability distribution of the phased structure factors, and a likelihood criterion is used to select the most probable phases for these reflections and to carry out phase refinement.

The Crystallographic Phase Problem

A full discussion of the maximum entropy method and its relationship to traditional direct methods in crystallography has been published by Bricogne (Bricogne 1984, 1988a, 1988b). A brief summary of the salient features of this approach is given here, coupled with descriptions of the method in practice.

As an initial step, the experimentally measured structure factor moduli $|F_{\mathbf{h}}|$ are placed on an absolute scale with a correction for overall, isotropic atomic thermal motion to give normalised, unitary structure factors $|U_{\mathbf{h}}|$, scaled such that U_0=1.0. The phase problem is

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 225–232.

concerned with the derivation of the phases ϕ_h of these unitary structure factors given only their moduli. The traditional approach to this problem involves the application of statistical methods.

In this approach N atoms are thrown randomly and independently into the asymmetric unit of the crystal. Using the standard mathematical techniques, we wish to calculate the joint probability distribution (jpd) P(U) of any collection $U = (U_{h1}, U_{h2}, \ldots\ldots U_{hn})$ of n unitary structure factors that we wish. We know by measurement the moduli of the structure factors, and can, therefore, substitute these values into the expression for the jpd P(U), and obtain the conditional distribution of their phases. This distribution will then be used to infer which combinations of phases are most likely since not all combinations will occur with equal probability.

There are however several problems with this approach:

(i) P(U) cannot be calculated exactly. The traditional approach uses the Gram-Charlier or Edgeworth series to generate approximations to P(U), and always assumes a uniform distribution of atoms. This yields approximations to P(U) which are good only near U=0 i.e. for *small* moduli, yet the expressions obtained have to be used with *large* moduli otherwise the conditional distributions are essentially featureless. In fact there is no tractable unique expression for P(U)

(ii) Direct methods fragment P(U) into small marginal distributions, of which the three-phase Cochran distribution is the most used.(Cochran, 1955) These marginals are then pieced back together to give a poor approximation to the full jpd; it would be much better to maintain and use large-base jpd's at the outset but this is difficult to do with traditional methods.

However, better analytical devices can be found to overcome these difficulties. The locus M defined by the large moduli is a high-dimensional torus which can be split up into sub-regions. Each such sub-region consists of a patch of M surrounding a point $U^* \neq 0$ located on M. Such a point U^* is obtained by assigning trial values to the phases of known, large moduli and can thus be compared with the multi-solution techniques used in traditional direct methods (Germain and Woolfson, 1968 and subsequent papers) For each sub-region a local approximation to P(U) can be constructed. With a sufficiently large collection of such constructs we have the means to calculate P(U) anywhere on M, and this leads to the concept of recentering . Recentering the Gram-Charlier or Edgeworth asymptotic expansion for P(U) away from U=0 by making trial phase assignments is equivalent to using a non-uniform prior distribution of atoms $q(\mathbf{x}) \neq 1/V$ (V is the volume of the unit cell) reproducing among its Fourier coefficients the components of U^*. This means, of course, that q(x) is highly indeterminent. There is however a uniquely defined 'best' choice for it, and that is the distribution which has the maximum entropy under these constraints. This can be justified either *via* Shannon's theory of information (Shannon and Weaver, 1949) or by calculating the $P^{SP}(U)$ saddlepoint approximation to P(U) (Daniels, 1954; Bricogne, 1984, 1988a, 1988b).

Significantly, the saddlepoint approximation yields identical expressions for the best choice of a non-uniform prior $q(\mathbf{x})$ without ever mentioning entropy or invoking the associated formalism.

Solving Structures via the ME-Saddlepoint Method

In the phasing process we never start from a position of complete ignorance. In order to define the origin of the unit cell in real space one to three $|U_{\mathbf{h}}|$'s can usually be given arbitary phase angles (subject to constraints arising from space group symmetry). In addition, in the non-centrosymmetric case, the enantiomorph needs to be defined; this is done in conjunction with the origin defining reflections, and usually involves a further phase assignment, so that the phases of 1-4 reflections are assigned. These form the basis-set , and they can be used as constraints in an entropy maximisation procedure to generate a non-uniform prior, $q^{ME}(\mathbf{x})$. The algorithm used for entropy maximisation is based on exponential modelling (Collins 1978; Collins and Mahar; 1983, Bricogne, 1984) with careful attenuation of shifts and a plane search algorithm which controls the rate at which detail is built in $q^{ME}(\mathbf{x})$ by modelling the entropy and the constraint as a bi-cubic function of two shift variables. A general purpose computer program MICE (Maximum entropy In a Crystallographic Environment) has been written for this purpose. It has an interface to the direct methods program MITHRIL (Gilmore 1984, 1988), which is used to calculate U-magnitudes, their standard deviations (Hall and Subramanian, 1982) and to select those reflections that optimally define the origin and the enantiomorph.

The distribution $q^{ME}(\mathbf{x})$ is a computational intermediate in obtaining an optimal approximation $P^{SP}(\mathbf{U})$ to $P(\mathbf{U})$. The approximate jpd $P^{SP}(\mathbf{U})$ extrapolates non-zero structure factors beyond the basis reflections which were used as constraints in the entropy maximisation; i.e. $q^{ME}(\mathbf{x})$ has non-zero coefficients even when no constraint values were specified *via* $\mathbf{U}^*$. Non basis-set reflections which are strongly extrapolated i.e. which have a large $|U^{ME}|$ value and a large $|U_{obs}||U^{ME}|$ product, can be incorporated into the basis-set with observed moduli and phases from the coefficients of $q^{ME}(\mathbf{x})$, and the non-uniform prior $q^{ME}(\mathbf{x})$ updated. Each new increment of the basis set defines a *node* on a phasing tree. When the basis set is small, this process reaches a point, however, where extrapolation becomes very weak, and new phase information is required.

At this point a suitable reflection or set of reflections is chosen with large $|U_{\mathbf{h}}|$ values. The choice is critical, and an algorithm based on an optimal extension of the second neighbourhood of the basis-set reflections is used. (For a discussion of neighbourhoods see Hauptman, 1980) Since the phases of such reflections are unknown, all possible permutations of the phases are tried, and for each new node the prior $q^{ME}(\mathbf{x})$ is updated. As described previously, this has the effect of recentering the distribution $P^{SP}(\mathbf{U})$. For centric reflections, this requires sampling the unknown phases at two points (ϕ and $\pi+\phi$. where ϕ is restricted by space group symmetry) and two new nodes are created,

whereas for acentric reflections the choices are $\pm\pi/4$ and $\pm 3\pi/4$ so that four new nodes are created. This method has parallels in the highly successful multisolution approach to the phase problem in traditional direct methods (Germain and Woolfson, 1968) In the maximum entropy method a tree structure is thus created. Such a tree soon becomes computationally unwieldy and needs to be 'pruned', and for this a likelihood function is used.

Likelihood

The approximate jpd $P^{SP}(\mathbf{U})$ depends on the phases of the constraints used to build $q^{ME}(\mathbf{x})$. For given values of these parameters yielding vector $\mathbf{U}^*$ we may look at the conditional distribution $P^{SP}(\mathbf{U}_\perp|\mathbf{U}^*)$ of any set $\mathbf{U}_\perp$ of structure factors for which no phase assignments have been made, and integrate it with respect to the phases of the $\mathbf{U}_\perp$ to get the conditional marginal distribution of the moduli $P^{SP}(|\mathbf{U}_\perp||\mathbf{U}^*)$. This differs from the Wilson distribution that is usually used in crystallography (Wilson, 1949) in two respects:

(i) P^{SP} is centred around $\mathbf{U}_\perp^{ME}$ not the origin.
(ii) P^{SP} has a covariance matrix $\mathbf{Q}=T(q^{ME})$, the Toeplitz matrix formed from $\mathbf{U}^{ME}$, not the identity matrix.

Thus phase choices for the basis set reflections $\mathbf{U}^*$ induce a deformation of the conditional marginal distribution of the moduli $|\mathbf{U}_\perp|$ away from their usual, Wilson distribution. Hypotheses about the phase values in $\mathbf{U}^*$ may then be tested as hypotheses about the distribution of moduli $|\mathbf{U}_\perp|$. For this purpose we define the likelihood Λ of the parameter values in $\mathbf{U}^*$ as the conditional marginal probability of the observed moduli:

$$\Lambda(\mathbf{U}^*) = P^{SP}(|\mathbf{U}_\perp|_{obs}|\mathbf{U}^*)$$

This likelihood can be used as a figure of merit in a multisolution-tree structure environment. It was shown by Bricogne (Bricogne 1984, 1988a) that this is a generalised quartet figure of merit (De Titta, Edmonds, Langs and Hauptman, 1975). For each node on the tree we can use likelihood to rank the equivalent nodes and so keep the computational aspects of the calculations under control.

In practice, it is convenient to normalise Λ with respect to the null hypothesis (H_0) that the distribution of atoms is uniform (i.e. that $\mathbf{U}^*=0$) and to use the likelihood ratio:

$$\frac{\Lambda(\mathbf{U}^*)}{\Lambda(\mathbf{O})} = \frac{P^{SP}(|\mathbf{U}_\perp|_{obs}|\mathbf{U}^*)}{P^{SP}(|\mathbf{U}_\perp|_{obs}|\mathbf{O})}$$

in its logarithmic form $L(\mathbf{U}^*) - L(\mathbf{O})$ where $L=\log\Lambda$.

For computational expediency, the diagonal approximation to the covaraince matrix of $Q(P^{SP})$ is used, so that explicit expressions for likelihoods are easily obtained (Bricogne, 1984,1988a). In spite of this approximation, the likelihood function still retains a great deal of its discriminating power as is shown in the next section.

Application to Two Small Organic Molecules

(i) Phase Extension

Phase extension is a process in which a set of phased U-magnitudes is used to generate phase angles for previously unphased reflections. It differs from *ab-initio* phasing in the size of the basis-set; for *ab-initio* methods the initial set $\mathbf{U}^*$ is very small and typically comprises 8-10 reflections, whereas for phase extension a set of 50 or more phased reflections is available. As a test of the exponential modelling algorithm, phase extension was carried out on sucrose octa-acetate $C_{28}H_{38}O_{19}$, space group $P2_12_12_1$ with Z=4. The top 300 U-magnitudes calculated from experimental data were given correct phases and the system subjected to 23 cycles of entropy maximisation *via* exponential modelling coupled with a plane search algorithm and shift attenuation. This is a very demanding test of the algorithm. Since they are corrected for thermal motion, U-magnitudes give very sharp maps. In consequence, q^{ME} has a huge dynamic range with even sharper peaks. Aliasing is also a problem. The results are summarised in Table 1.

At χ^2=1.0, 492 non basis-set U_h's having the product $|U_{obs}||U^{ME}| \geq 0.01$ have calculated phase angles with a mean absolute phase error of 9.0 degrees. It is worth noting that, unlike the tangent formula which is traditionally used in phase extension in direct methods (Karle and Karle, 1966), this calculation has been carried out *without* consulting the magnitudes of these reflections.

Cycle	χ^2	cos(∇S,∇C)	Entropy
1	25.0	0.83	-0.01
5	21.3	0.96	-0.25
10	12.8	0.94	-1.55
15	3.5	0.69	-13.41
20	1.4	0.82	-38.00
23	1.0	0.84	-67.80

Table 1: Phase Extension on Sucrose Octa-acetate. Cos(∇S,∇C) is the cosine of the angle between the gradient of the constraint C and the gradient of the entropy S.

(ii) ***Ab-initio* phasing**

As a test of the use of likelihood and the ability of the maximum entropy method to extrapolate phase information even when the basis set is very small, the *ab-initio* phasing of a small organic molecule was performed. The molecule was (-)platynecine, $C_8H_{15}NO_2$, which crystallises in space group $P2_12_12_1$ with Z=4. The basis set was chosen to comprise 3 reflections which simultaneously defined the origin and enantiomorph. From these reflections with the subsequent addition of permuted reflections the structure was readily solved. Figure 1 shows a summary of the first stages of the phasing tree.

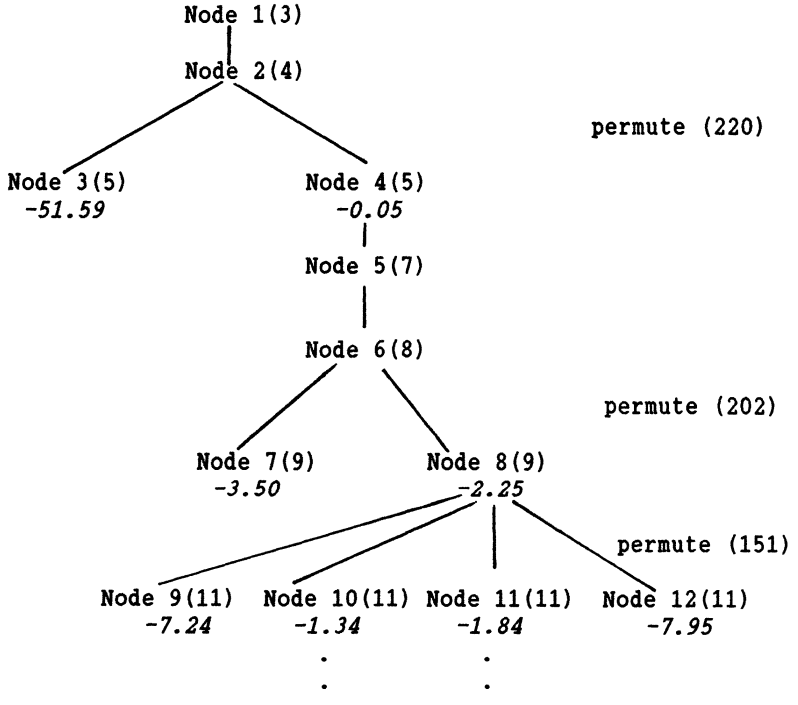


Figure 1: The Initial Stages of the Phasing Tree for (-)Platynecine. The figures in parentheses refer to the number of basis reflections; the figures in *italics* are $L(U^*) - L(O)$. Node 11 is the correct node; Node 10 is shown to be incorrect several levels later.

Whereas the entropy of a node is of little value as a figure of merit, the power of the likelihood function as a discriminator is quite extraordinary: no other figure of merit exists which can discriminate so precisely with such a small basis set of reflections. Direct methods uses figures of merit based on consistency indices, but they are unable

to provide reliable indications until the basis set spans at least 50 reflections and even then the reliabilty can be poor. However, care does need to be taken; normally large differences in likelihood are necessary for complete confidence, and very small differences, such as between Nodes 10 and 11 are such that both nodes must be further explored.

Computer graphics has an important role to play here. The visual examination of appropriate maps as they are generated is very important in following and assessing the phasing process which allows the incorporation of the user's chemical knowledge in a weak way.

Phase Refinement

Likelihood calculations also permit phase refinement of the phases of the basis set reflections *via* optimisation of $L(\mathbf{U}^*)$ with respect to the phases in $\mathbf{U}^*$. The Jacobian matrix $\partial(\mathbf{U}^{ME})/\partial(\mathbf{U}^*)$ must be obtained. This is a relatively simple calculation. Table 2 shows how the mean absolute phase error is reduced from 11.1 degrees to 9.3 degrees in 4 successive cycle of phase angle refinement for node 18 of (-)Platynecine. It may seem a small improvement, but small changes of this magnitude can have a highly significant effect on subsequent phase extrapolations. Note also that the likelihood function $L(H) - L(H_0)$ is substantially increased.

Cycle	Mean Absolute Phase Error	$L(H) - L(H_0)$
1	11.1	-4.50
2	10.1	-0.31
3	9.7	-0.26
4	9.3	-0.20

Table 2: Variation of Mean Absolute Phase Error in Successive Cycles of Phase Angle Refinement using Likelihood Optimisation

Acknowledgements

This work was supported via a grant from the SERC. CJG wishes to acknowledge the Ciba-Geigy Fellowship Trust for a Senior Ciba-Geigy Fellowship which led to the work reported here.

References

G. Bricogne (1984) **Acta Cryst.**, **A40**, 410-445.
G. Bricogne (1988a) in *Crystallographic Computing 4* ed. N.W. Isaacs and M.R. Taylor, O.U.P pp 60-79.
G. Bricogne (1988b) **Acta Cryst.**, **A44**, 517-545.
W. Cochran (1955) **Acta Cryst.**, **8**, 473-478.
D.M. Collins (1978) **Nature (London)**, **298**, 49-51.
D.M. Collins and M.C.Mahar (1983) **Acta Cryst.**, **A39**, 252-256.
H.E. Daniels (1954) **Ann. Math. Stat.**, **25**, 631-650.
G.T. DeTitta, J.W. Edmonds, D.A. Langs and H.A. Hauptman (1975) **Acta Cryst. A31**, 474-479.
G. Germain and M.M.Woolfson, (1968) **Acta Cryst**, **B24**, 91-96.
C.J. Gilmore (1984) **J. Appl. Cryst.**, **17**, 42-46.
C.J. Gilmore (1984) **J. Appl. Cryst.**, **21**, In the press.
S.R. Hall and V. Subramanian (1982) **Acta Cryst.**, **A38**, 598-608.
H.A. Hauptman in *Theory and Practice of Direct Methods in Crystallography* ed. M.C.F. Ladd and R.A. Palmer, Plenum Press, 1980, pp 151-197.
J. Karle and I.L. Karle (1966) **Acta Cryst. 21**, 849-859.
C.E. Shannon and W.Weaver (1949) *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Il.
A.J.C. Wilson (1949) **Acta Cryst.**, **2**, 318-321.

THE CHALLENGE OF X-RAY AND NEUTRON POWDER DIFFRACTION

Keith Henderson & Christopher Gilmore,
Department of Chemistry,
The University,
Glasgow G12 8QQ

ABSTRACT. There are many structures of considerable interest for which crystals of sufficient size for single crystal diffraction experiments cannot be grown. In these cases powder diffraction patterns can be taken. Ab initio structure solution from such patterns is difficult due to the limited amount of data which can be extracted. We are investigating the application of ME and related techniques to the solution of this problem. We present a method for computing the probability that a trial combination of phases is correct. Extensions to the ME formalism which are necessary to deal with unequal scatterers, including negative neutron scattering lengths, are then described.

1. INTRODUCTION

Virtually all crystal structures are solved from single crystal data, where each of the diffracted intensities can be observed separately. Frequently the task of obtaining a sufficiently large crystal for these methods is very difficult, and may prove impossible. This is the case with many crystalline substances of interest, such as for example zeolites. A fine powder of microcrystallites, from which a powder diffraction pattern may be taken, is often easily obtained. In such a pattern the diffracted intensities are sorted purely on their Bragg angle, with inevitable overlapping of data which rapidly becomes worse as the Bragg angle increases. As a result, the number of reflexions for which intensities can be unambiguously assigned is smaller than with single crystal diffraction, and these reflexions tend to be of low resolution.

This is not to say that structures cannot be solved from their powder patterns. If there are a small number of parameters to be determined this can be done. Many examples of such determinations are to be found in the early volumes of *Acta Crystallographica*, particularly by Zachariasen (1948). In more recent years interest has returned, but with the need to solve structures of greater complexity.

The width of the diffraction peaks comes from two distinct sources. The first is instrumental, and can be greatly reduced by the use of

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 233–239.

devices such as focussing cameras and, more recently, the synchrotron (Parrish & Hart,1987)and intense neutron sources (Clearfield & Rudolf,1987). With such high resolution equipment the problem of overlap can be reduced to the level where sufficient data for input to conventional direct method packages can be obtained for structures of a few atoms. A recent paper by McCusker (1988) reports the solution of a zeolite with eleven atoms in the asymmetric unit. The paper also references some other *ab initio* determinations which have been made in the last decade. The second source of peak width is due to the microcrystalline composition of the specimen, and this sets an absolute limit on the observable peak width.

As the structures become larger, the amount of available data will fall short of the amount needed by conventional techniques to solve the phase problem. There is thus a need for procedures which make better use of the available information than do the conventional approaches.

2. EVALUATING PHASE PERMUTATIONS

The major problem which confronts us in *ab initio* structure determination is the extension of structure factor phases from a small 'basis set' to a number sufficient to reveal the atomic positions through a Fourier synthesis.

Entropy maximisation based on the reflexions we can choose arbitrarily yields predictions for the rest of the data set. We might hope to phase extend by taking the 'extrapolated' phase and associating this with the known magnitude of the reflexion. In practise this procedure is unlikely to succeed. In the early stages of the solution attempt extrapolations are small and unreliable - unsurprisingly since many structures would satisfy the constrained reflexions. We are then forced into the growing of trees, each node of which involves an entropy maximisation. This is rather expensive in terms of computer time, so a question which arises is: Can we reduce the number of times we perform an entropy maximisation by assessing the plausibility of a set of permuted phases being correct?

2.1. Derivation of the Figure of Merit

To deal with a concrete example consider a centrosymmetric structure, where all structure factor phases have values of either 0 or π. We can say that the signs of the structure factors are either 'plus' or 'minus' respectively. Choosing such an example simplifies the problem in two ways. Firstly we do not have the problem of deciding how finely to sample the trial phase values. Secondly we need not worry about the definition of 'enantiomorph' which occurs with non centrosymmetric structures.

Suppose that we wish to add m unitary structure factors U_i, with associated signs s_i. We would like to estimate the probability that the particular choice of signs is correct. That is, we would like to be able to estimate

$$p(s_1,..,s_m \mid |U_1|...|U_m|,C)$$

where C stands for our state of knowledge at the current node.

We can obtain such an approximation by recourse to methods which play an important part in traditional direct methods. If a quantity is formed as the sum of independently and identically distributed random variables, in this case a collection of structure factors is formed as the sum of contributions from each atom, an asymptotic expansion of the result exists having the general form

$$p(\mathbf{U}|C) = A \exp\{-(\mathbf{U}-\mathbf{P})^T\mathbf{M}^{-1}(\mathbf{U}-\mathbf{P})/2\} \{ 1 + \text{series in } \mathbf{U} \}$$

Here the exponential term is a multivariate Gaussian, **P** is the vector of mean values, **M** is the variance-covariance matrix and A is a normalising constant. This is the factor which corresponds to the central limit theorem of statistics. The terms in the perturbing series compensate for deviations from normality due to summing over a finite number of random variables.

The form of **M** depends upon the prior distribution of atomic positions. If all positions are equally probable **M** becomes diagonal, decoupling the reflexions from each other. With such a prior, connections between different structure factors must be sought in the perturbing series, leading to conventional direct method formulae. However, if the prior distribution is non-uniform, linkage between reflexions occurs at the Gaussian level of approximation. We can then usefully approximate the joint distribution for a set of structure factors by the Gaussian part alone. The prior may be constructed by maximising the entropy of the positional distribution under the constraint that certain structure factors take prescribed values.

A trial value of **U** may be constructed by combining the observed magnitudes with a trial vector of signs **s**, and we may then determine its probability of being correct by the Gaussian approximation. In practice, we compute the quadratic form

$$Q = (\mathbf{U}-\mathbf{P})^T\mathbf{M}^{-1}(\mathbf{U}-\mathbf{P})$$

The smaller this quantity is, the more likely the corresponding **U** is deemed to be correct. It is useful to look at this quantity in real space. It can be shown (Bricogne,1984) that Q is proportional to the following quantity

$$Q' = \int \frac{\delta^2(\mathbf{x})}{q^{ME}(\mathbf{x})}\, d^3\mathbf{x}$$

where $q^{ME}(\mathbf{x})$ is the prior distribution for atomic positions and $\delta(\mathbf{x})$ is a fourier synthesis constructed from **U-P**. For this to be small, fluctuations in $\delta(\mathbf{x})$ must coincide with large values of $q^{ME}(\mathbf{x})$.

2.2 An Application

We have applied this method to a trial structure, Mercury Phosphate, with the formula $Hg_3(PO_4)_2$(Aurivillius & Nilsson, 1975). The data is for X-ray diffraction so that the scattering factor for each atom is closely proportional to the atomic number. The Hg atoms may then be expected to dominate the maps. The three reflexions permitted to fix the origin were chosen and a prior distribution generated by entropy maximisation. The three Hg atoms all lie in regions of high density, the lighter P atoms do not. Four more structure factors were chosen to add to this set. Only one of the four had a substantial extrapolated value and this proved to be correct. The remaining three extrapolations were small, and were in fact all wrong. The 16 possible combinations of the 4 signs were evaluated by the procedure described above. Table 1 gives the results. Permutations are with respect to the correct set of signs, so that the ++++ permutation which occurs first is the correct solution, which would be expected to have the smallest value of Q'. Also shown as Q" is the value Q' which would occur if we took $q^{ME}(\mathbf{x})$ to be uniform. The value Q" thus corresponds to assuming the ME extrapolations are independent. The units are arbitrary since Q' and Q" were evaluated by summation over a grid. Evidently the value of Q' is a good discriminator against almost all the other sign permutations. Only one gives a smaller value for Q', and this corresponds to taking all signs equal to those of the extrapolated structure factors. Clearly the procedure is not perfect, but it could have substantially reduced the amount of exploration of

TABLE I

Permutation	Q'	Q"
+ + + +	20.964	5.619
- + + +	67.679	5.455
+ - + +	25.260	6.762
- - + +	68.314	5.565
+ + - +	35.065	6.597
- + - +	35.110	5.400
+ - - +	43.133	6.707
- - - +	40.643	6.542
+ + + -	30.278	5.120
- + + -	35.220	4.955
+ - + -	39.302	6.262
- - + -	40.584	6.097
+ + - -	57.990	5.065
- + - -	16.264	4.900
+ - - -	71.912	6.207
- - - -	26.525	6.042

nodes. After entropy maximisation with the four reflexions (with correct

signs) added to the origin defining set, the Hg atoms were already well localised, and one would anticipate phase extension proceeding easily from this point.

The problem which now arises is the determination of the positions of the lighter atoms, here the phosphorous and oxygen. If we construct a measure in which the features we already know about, the Hg atoms, are present, we can again maximise the entropy. The quotient of the resulting probability and the measure, which we may call the 'modulating function', shows the additional modification of probability needed to fit the constraints. If we put the known atoms in the measure, we would expect the positions of any remaining atoms to become apparent in the modulating function. This approach was checked by generating a measure from an ME map formed by 28 reflexions to a resolution of 2.8A. This is low by direct methods standards. All density below a threshold value was levelled, leaving only the features due to the Hg atoms. Maximising the entropy with this measure yielded a modulating function in which the P atoms appeared clearly.

This all appears very satisfactory, but there are aspects of applying a non-uniform measure in such a way which are ill founded.

3. POSITIONAL PROBABILITIES IN MAXIMUM ENTROPY

What do the density maps represent? From the point of view of Bayesian methods there is little doubt that they should be a measure of the probability that a particular point in the unit cell is occupied by an atom. Here we run into a problem that is apparent in the structure described earlier. In the map using origin defining reflexions alone, the 'heavy' Hg atoms were concentrated in high density. On the other hand the lighter P atoms were not. This seemed natural since the Hg atoms are by far the stronger scatterers. However, it means that the 'density' generated by the entropy maximisation cannot be a real estimate for positional probability for all the atoms because such a density should depend on the strength of scattering of the atom. The situation becomes even more confusing if we want to consider neutron scattering, where the scattering length for H is negative. This would seem to require 'holes' in the density map to correspond to high probability. We really need to write the equations for entropy maximisation in a way in which it is clear that we are working with a probability of atomic position.

The application of ME principles to crystallography, as described by Bricogne(1984), Wilkins et al.(1983) and others, deal with entropy maximisations in a single 3 dimensional space. Extensions to this work have been made recently by Bricogne(1988). In Bricogne's papers the maximisation of entropy plays an important but incidental part in the approximation of joint conditional distributions of structure factors through a 'saddlepoint' approximation. All his results are demonstrated without appealing to a maximisation principle for entropy. Nevertheless, if we start from such a principle some of the extensions easily emerge.

Following Bertaut(1958), we can say that the most general form for the joint distribution of t atomic positions is defined within the

configuration space of the structure. In principle we would like to maximise the entropy defined as

$$S = -\int\!\!\int p(\mathbf{x}_1,..,\mathbf{x}_t)\{\log\, p(\mathbf{x}_1,..,\mathbf{x}_t)-\log\, m(\mathbf{x}_1,..,\mathbf{x}_t)\}d^3\mathbf{x}_1..d^3\mathbf{x}_t$$

In practise it is not feasible to perform computations in this many dimensions. If we assume each of the atoms follows an independent distribution function, then the joint distribution of all the atoms is a product of distributions for the separate atoms. The entropy can then be written as

$$S = -\sum_{j=1}^{t}\int p_j(\mathbf{x}_j)\{\log\, p_j(\mathbf{x}_j)-\log\, m_j(\mathbf{x}_j)\}d^3\mathbf{x}_j = \sum_{j=1}^{t} S_j$$

Introducing constraints on m structure factors and following the usual methods we find the following solutions for the positional probabilities

$$p_j(x_j) = m_j(x_j)\, \exp\{\, f_j \sum_{k=1}^{m} \lambda_k \xi_j(\mathbf{h}_k) + \nu_j \,\}, \; j=1,..,t$$

with ν_j a multiplier for normalisation of the probability density. λ_k is the multiplier associated with the constraint on the kth reflexion. This appears in the distribution for each of the atoms, so relating them, but each also has the scattering factor for the atom scaling the effect of λ_k. It is apparent that the depth of modulation of the positional probability depends on the scattering factor for the atom. For a 'heavy' atom (large f_j) the positional probability is going to be more concentrated than for a lighter atom. This corresponds to our intuitions about atoms of different weights in the earlier maps, and results of this kind are well known in direct method theory. It is also clear that negative scattering factors do not cause any problem, so neutron scattering can be treated.

4. ACKNOWLEDGEMENT

We wish to acknowledge BP Petroleum PLC for the provision of an Extra-mural Research Award which funded this research.

5. REFERENCES

Aurivillius, K. & Nilsson, B. A. (1975) *Z. Kristallogr.* **141**,1-10
Bertaut, E.F. (1958) *Acta Cryst.* **11**,405-412
Bricogne, G. (1984) *Acta Cryst.* **A40**,410-445
Bricogne, G. (1988) *Acta Cryst.* **A44**,517-545
Clearfield, A. & Rudolf, P.R. (1987) *Transactions ACA* Vol **23**, 35-49

McCusker, L. (1988) *J. Appl. Crystallogr.* **21**,305-310
Parrish, W. & Hart, M. (1987) *Z. Kristallogr.* **179**,161-173
Wilkins, S.W., Varghese, J.N. & Lehmann, M.S. (1983)
Acta Cryst. **A39**,47-60
Zachariasen, W.H. (1948) *Acta Cryst.* **1**,263-265

A Statistical Potential for Modelling X-ray Electron Density Maps with Known Phases

Andrew D. McLachlan,
Medical Research Council Laboratory of Molecular Biology,
Hills Road Cambridge, CB2 2QH

Abstract

We describe methods for constructing a Maximum Entropy X-ray electron density map in crystallography that fits certain amplitudes and phases uniquely subject to a squared residual constraint. The calculations use a free energy analogue, or *statistical potential* that derives from the grand partition function of the maximum entropy problem in Fourier space. It is a function of the *statistical forces*, the Lagrangian multipliers of the entropy. Three new functions Y, G and Ψ allow us to fit the data with predetermined accuracy, and to avoid divergences which would otherwise occur. The method is able to handle physically realistic and elaborate models: cells with fixed density regions, several types of scattering atom, anomalous dispersion. The control algorithm, and the relation between the domains of the force and probability variables are outlined.

1. INTRODUCTION

Any application of maximum entropy principles to crystallography requires trial maximum entropy electron density maps that aim to fit certain target sets of amplitudes and phases. The best current methods for calculating these maps fall into two classes. Those of the first class [1-7] deal directly with the entropy and the constraints, using an atomic probability distribution as the basic working medium. Those of the second [8-14] minimise a type of free energy function or 'statistical potential' and use the entropy gradients or 'statistical forces' as working variables. The indirect nature of the free energy methods is offset by certain advantages. All the generated maps are physically feasible maximum entropy distributions; constraints can often be transformed away; mathematical concepts drawn from statistical mechanics [13] and optimisation theory [15,16] can be deployed.

Here we develop some new kinds of statistical potential [14] that are stable when the constraints are barely feasible, and that allow precise control of the quality of the fit to the target. We also outline some example applications that extend the scope of the theory to treat more realistic density map models.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 241–249.

2. STATISTICAL FORCES AND POTENTIALS

2.1 Entropy and Constraint Functions

We take a crystallographic unit cell in real space, divided into a discrete set of grid points dimensioned (L_1, L_2, L_3) with a total of $L = L_1L_2L_3$ points. Each point, indexed $j = (j_1, j_2, j_3)$, has coordinates $x_j = (x_1, x_2, x_3) = (j_1/L_1, j_2/L_2, j_3/L_3)$.
Suppose that there is a random distribution of N atoms within the cell. Then we associate a *probability density* p_j and an *entropy* $s_j(p_j)$ with each grid point. The entropy normally takes the classical form $s = -p \log p$. We also introduce, for later use, two dual variables. The *statistical force* q_j conjugate to p_j, is defined as $-ds_j/dp_j$. The *statistical potential* $\omega_j = s_j + p_jq_j$ is a function of q_j and is defined so that $p_j = d\omega_j/dq_j$. Thus we have

$$s_j = -p_j \log p_j, \quad \omega_j = \exp(q_j - 1). \tag{2.1}$$

$$q_j = -\frac{ds_j}{dp_j} = 1 + \log p_j, \quad p_j = \frac{d\omega_j}{dq_j} = \exp(q_j - 1). \tag{2.2}$$

The gradients and second derivatives of s and ω are related by a dual Legendre transformation [15-18]

$$ds = -qdp, \quad d\omega = pdq, \quad H(s) = -\frac{1}{p}, \quad H(\omega) = \exp(q - 1), \quad H(s)H(\omega) = -1. \tag{2.3}$$

In reciprocal space the probabilities and statistical forces have complex Fourier components t_h and T_h respectively

$$p_j = \frac{1}{\sqrt{L}} \sum_h t_h \exp(-2\pi i h x_j), \quad q_j = \frac{1}{\sqrt{L}} \sum_h T_h \exp(-2\pi i h x_j). \tag{2.4}$$

Here $h = (h_1, h_2, h_3)$ is a reciprocal lattice vector and hx stands for the scalar product $(h_1x_1 + h_2x_2 + h_3x_3)$. We use total S and Ω functions scaled by a factor $1/L$, so that the functions and their differentials are

$$S = \frac{1}{L} \sum_j s_j(p_j), \quad \Omega = \frac{1}{L} \sum_j \omega_j(q_j), \tag{2.5}$$

$$dS = -\frac{1}{L} \sum_h T_{-h} dt_h, \quad d\Omega = \frac{1}{L} \sum_h t_{-h} dT_h. \tag{2.6}$$

Note that in Fourier space the differentials involve conjugate pairs h and $-h$, and that we define the gradient operators by equations such as $\nabla_h = L\partial/\partial T_{-h}$, scaled by a factor L.

Suppose that we wish to match a 'constraint subset' $[h]$ of the probability Fourier components t_h to fit some target amplitudes and phases t_h^0. We can either seek an exact fit $t_h = t_h^0$, or else aim for a suitably low value $E = \epsilon_m$ for a quadratic error function [2,3,5] such as

$$E^2 = \frac{1}{L} \sum_{[h]} \frac{|t_h - t_h^0|^2}{\alpha_h}. \tag{2.7}$$

Here the sum is over the constraint subset $[h]$ and the weighting factors α_h are used to control the relative mean square deviations of the different Fourier components. For an exact fit the well-known normalised solution [1,8,11], with the mean value of p_j fixed as 1, is

$$p_j = \frac{1}{Z} \exp\left(\frac{1}{\sqrt{L}} \sum_{[h]} \exp(-2\pi i h x_j) T_h \right), \tag{2.8}$$

where Z is the usual partition function (scaled by a factor L). The statistical forces T_h in the constraint subset $[h]$ must be chosen to match the desired targets t_h^0. The other free forces vanish, $T_k = 0$. On the other hand the inexact constraint $E = \epsilon_m$ requires the solution [5] of the well-known equation $(\frac{1}{2}\nabla E^2 - \lambda \nabla S) = 0$ or

$$\frac{(t_h - t_h^0)}{\alpha_h} + \lambda T_h = 0, \tag{2.9}$$

within the subset $[h]$, and $T_k = 0$ for the other terms, which corresponds to a unique maximum of the function $W = S - E^2/2\lambda$.

2.2 Exact Target Potentials

The maximum entropy distribution (2.8) requires that we find correct forces T_h that will generate the exact target Fourier components t_h^0. This is usually done by setting up a special statistical function of the forces, which is minimised without constraint. One such *exact target potential*, $J(T)$, derives from the cumulant of Z, $K = \log Z(T)$. This is a function only of the forces in set $[h]$:

$$J(T) = K(T) - \frac{1}{L} \sum_{[h]} t_{-h}^0 T_h, \quad \nabla_h J = (t_h - t_h^0). \tag{2.10}$$

Thus J has a unique minimum wherever $t_h = t_h^0$ for all the constraints. This type of function is used by Wilkins [8], Navaza [11] and Bricogne[13], and is analogous to Levine's available work function [17,18]. One drawback of the J-potential is that T_0 appears as a function of the other T_h. In real space the result is that the grid forces q_j are not independent variables. A rather better potential, Q, is derived [14] from the grand partition function Ω

$$Q(T) = \Omega(T) - \frac{1}{L}\left[t_0^0 T_0 + \sum_{[h]} t_{-h}^0 T_h \right], \quad \nabla_h Q = (t_h - t_h^0), \tag{2.11}$$

which allows t_0^0 to have any target value. Unfortunately both the Q and the J potentials share a serious intrinsic weakness, which has often gone unrecognised. If the target amplitudes and phases t_h^0 are physically unattainable, as for example when they demand negative probabilities p_j, the search for a minimum of the potential diverges, and some forces T_h tend to infinity! Obviously it is desirable to limit the magnitudes of the T_h in some way, for example by adding a force penalty function [12]. A systematic approach is better.

2.3 Approximate Target Potentials

We first introduce two force averages F and θ

$$F^2 = \frac{1}{L}\sum_h |T_h|^2 = \frac{1}{L}\sum_j q_j^2, \quad \theta^2 = \frac{1}{L}\sum_{[h]} \alpha_h |T_h|^2. \tag{2.12}$$

Here F is simply the root mean square average of ∇S, including *all* h values,while θ^2 is a weighted mean square force chosen to complement the error function E in (2.2). We can now replace the Q-potential by a more robust function $Y(T)$, which includes a new quadratic control term C

$$Y = Q + \lambda C, \quad C = \frac{1}{2}\theta^2. \tag{2.13}$$

This possesses a unique minimum for any finite positive value of λ. Furthermore, the stationary point of Y in the $[h]$ space satisfies the equation

$$\nabla_h Y = (t_h - t_h^0) + \lambda \alpha_h T_h = 0. \tag{2.14}$$

Our equation (2.14) is identical with (2.9), and thus demonstrates that $Y(T)$ is the counterpart in the domain of force variables to Bryan and Skilling's [2] function $W(p) = (S - E^2/2\lambda)$ in the domain of probability variables. In fact $Y_{min}(T, \lambda) \equiv W_{max}(t, \lambda)$ at the corresponding solution points.

A maximum entropy solution with the Y function can be completed when we know the appropriate value of λ. An exact fit to the target uses $\lambda = 0$, but generally we wish to fit $E = \epsilon_m$, where ϵ_m is prescribed, and find a corresponding value for $\lambda = \lambda_m$. We now deduce from (2.7),(2.12) and (2.14) that at any solution point for E

$$E^2 = \frac{1}{L}\sum_{[h]} |t_h - t_h^0|^2/\alpha_h = \frac{1}{L}\sum_{[h]} \lambda^2 \alpha_h |T_h|^2 = \lambda^2 \theta^2. \tag{2.15}$$

The result, that $\lambda = E/\theta$, can be used to prove important relations [14] which hold at the balance of Q and λC. Here θ behaves like a statistical temperature parameter for E.

$$\frac{dS}{dE} = \theta, \quad \frac{dQ}{d\theta} = -E; \quad \left(\frac{d^2S}{dE^2}\right)\left(\frac{d^2Q}{d\theta^2}\right) = -1. \tag{2.16}$$

More useful for our immediate purpose is the fact that (2.15) allows the construction of a new 'specified ϵ_m accuracy' target potential

$$G(T) = Q + \epsilon_m |\theta| \tag{2.17}$$

which has its unique minimum at the required point where $E = \epsilon_m$. Since the gradient of G is $\nabla G = \nabla Q + (\epsilon_m/\theta)\nabla C$ the factor $\epsilon_m/\theta = \lambda^*$ behaves as a 'self-adjusting' Lagrangian multiplier that achieves the correct value $\lambda_m = \epsilon_m/\theta_m$ at the solution point where $\theta = \theta_m$.

Unfortunately there remains the possibility that owing to a misjudgement the desired targets t_h^0 cannot be fitted as accurately as expected, with $E = \epsilon_m$, because ϵ_m is too small. This causes a divergence, as the minimum of $G(T)$ recedes to infinity. Therefore

it is necessary to provide further controls. One approach is to set a fixed upper limit θ_m for the temperature parameter and accept the corresponding best value of E (there is a corresponding 'θ_m target potential' in probability space that achieves this: $U(p) = S(p) - \theta_m E(p)$). However, because of the various weights α_h in the different terms it is not easy to choose θ_m suitably in advance. We prefer to choose a preset limit F_m^2 for the mean $\langle q_j^2 \rangle$ or equivalently $\langle (\log p_j + 1)^2 \rangle$ by working with a control function of the form

$$\Psi(T) = G + \frac{1}{2}\mu F^2 = Q + \epsilon_m|\theta| + \frac{1}{2}\mu F^2 \tag{2.18}$$

Here the Lagrange multiplier μ is increased from zero, if necessary, to ensure that $F \leq F_m$. The Ψ potential control has the useful property that it first tries to fit the targets t_h^0 with the desired accuracy if this can be done in the region $F \leq F_m$. If not, then a best compromise fit $E > \epsilon_m$ is selected. The stationary point satisfies the equation

$$\nabla_h \Psi = (t_h - t_h^0) + (\lambda^* \alpha_h + \mu) T_h = 0 \tag{2.19}$$

in which $\lambda^* = \epsilon_m / \theta$. The value of μ required is a unique decreasing function of the chosen F_m. We see that the F^2 term in Ψ acts as if each deviation weight α_h was loaded with a uniform increase to $(\alpha_h + \beta)$, where $\beta = \mu / \lambda^*$.

2.4 Summary

The search for a reasonable maximum entropy procedure in the statistical force domain has led us through several stages. We began with the well-known partition functions Z or Ω. Their 'available work' exact target potentials J and Q may, in favourable cases, possess unconstrained minima that yield a perfect fit to any feasible amplitude-phase target. We have seen, however, that the forces diverge to infinity if the target is not physically feasible, but that then the Y function (in the force domain) or the W function (probability domain) give a unique range of finite solutions for the typical quadratic constraint $E^2 = \epsilon_m^2$. When ϵ_m is feasible the Lagrange multiplier λ_m can be obtained uniquely by using the self-adjusting G function. If ϵ_m is not feasible we control any divergence of G by adding a term proportional to $\frac{1}{2}F^2$ and thus keep F within a reasonable bound F_m.

The choice of the F_m limit is a reasonable and simple practical device to ensure convergence of the maximum entropy calculation. It has no deeper statistical significance. All these methods require of course that ϵ_m must be less than the global maximum 'flat map' value $E = \epsilon_{max}$. The control of non-quadratic constraints in the force domain (e.g. fitting Fourier intensities) [5,19,20,25] still requires further development.

3. APPLICATIONS

In many simple maximum entropy calculations it is equally easy to use either the probability variables (p_j, t_h) or force variables (q_j, T_h), but for more realistic and complicated scientific applications it is most natural to work with the forces. We outline three examples below. These are analysed from a different viewpoint by Bricogne [21].

3.1 Crystal Regions of Known Density

Suppose that we have known tied probability densities p_f^0 on a certain set of grid points $[j = f]$. The constrained maximum entropy map can be calculated very easily by using

the forces q_j and T_h as before, but with a modified 'fictitious entropy' function at each tied grid point in $[f]$.

$$s_f^\bullet = -\frac{1}{2}k_f(p_f - p_f^0)^2, \quad \omega_f^\bullet = p_f^0 q_f + \frac{q_f^2}{2k_f}, \tag{3.1}$$

where $q_f = k_f(p_f - p_f^0)$. The factor k_f has a large value chosen to hold the density close to p_f^0. The ficititious force q_f increases rapidly with any deviations from the local grid target density. This method can be used for solvent flattening [22-24], or for crystals that contain known structural fragments. Note that this method is not equivalent to the use of a prior model density $m_f = p_f^0$, which is a part of the true entropy function [3] and allows no independent accuracy control .

3.2 Several Atom Types with Spatial Exclusion

Our cell grid is supposed to contain a population of different atoms, with N_ν atoms of type ν, each having an atomic charge z_ν. Furthermore, no more than one atom will be allowed to occupy each grid point. The maximum entropy distribution over the grid will have a partial probability density $p_{j\nu}$ for each type ν, which contributes a partial charge $p_{j\nu}z_\nu$ to the charge ρ_j. The constraints are the given Fourier components ρ_h for the total charge, and the given total atom populations N_ν. The required statistical potential Ω is a generalised form of Fermi-Dirac function

$$\Omega(\mu_\nu, T_h) = \sum_j \log\left(1 + \sum_\nu exp(\mu_\nu + z_\nu q_j)\right). \tag{3.2}$$

in which μ_ν is a chemical potential for each atom type, in place of T_0. The forces q_j are constructed from the remaining statistical forces T_h in Fourier space. The partial atomic probability distributions come out as

$$p_{j\nu} = \frac{exp(\mu_\nu + z_\nu q_j)}{1 + \sum_\kappa exp(\mu_\kappa + z_\kappa q_j)}. \tag{3.3}$$

The whole multi-atom density map is represented in terms of a *single* force map q_j with its associated μ_ν parameters, rather than separate maps for the individual atom types ν. The heavy atoms tend to congregate at the high peaks of ρ_j (driven by the factor $exp[z_\nu q_j]$) while the light atoms are displaced to the low peaks. A classical entropy multi-atom Ω function can easily be generated instead. Notice that these methods allow for negative scattering powers in z_ν (e.g. neutron diffraction).

3.3 Shaped Atoms with Anomalous Scattering

Each atom has a shaped density contributing to real and imaginary scattering densities $(\rho_j' + i\rho_j'')$ in real space or $(\rho_h' + i\rho_h'')$ in Fourier space. The atomic centres have an unknown real probability distribution p_j with Fourier components t_h, and every atom has a complex scattering factor $z_h = (z_h' + iz_h'')$ such that

$$\rho'_h + i\rho''_h = t_h(z'_h + iz''_h). \tag{3.4}$$

To construct a maximum entropy distribution that matches ρ'_h and ρ''_h, using Fermi-Dirac statistics for the atoms, we introduce partial force fields γ'_h, γ''_h and statistical forces T_h into a potential $\Omega(\mu, \gamma'_h, \gamma''_h)$ as follows:

$$\Omega = \sum_j \log[1 + exp(\mu + q_j)] \tag{3.5}$$

$$T_h = \gamma'_h z'_{-h} + \gamma''_h z''_{-h}. \tag{3.6}$$

The charge densities are expressed in the equations

$$d\Omega = \frac{1}{L}\sum_h (\rho'_{-h} d\gamma'_h + \rho''_{-h} d\gamma''_h) + \frac{N}{L} d\mu. \tag{3.7}$$

The targets are then satisfied by minimising the appropriate target potential Q, Y or Ψ. These equations allow phase information from anomalous scattering to be incorporated into the maximum entropy analysis.

4. CONTROL ALGORITHMS

The control process is the part of the maximum entropy algorithm that estimates the moves needed to reach a solution in force space and arrives at trial values for the Lagrange multiplier λ. It is not trivial [2,3,5], and the sketch below only indicates a few key points in our approach.

4.1 Probability and Force Domains

We set our first objective as finding minima of the statistical potential $Y = Q + \lambda C$ in the force domain, and estimating λ. The analogous calculation in the probability domain is to maximise $W = (S - E^2/2\lambda)$. Bryan and Skilling's algorithm [2] solves the latter problem by maximising S on a descending sequence of error contours $E = \epsilon$ where E decreases towards ϵ_m. The moves are over search ranges $p_j + \sum_a u_{ja} x_a$ in the probability domain. An entropy distance metric $D(s)^2 = -H(p)(\Delta p)^2 = (\Delta p^2)/p$ is used to limit the moves and to precondition the search directions.

The statistical potentials are well-behaved functions in the force domain, and we use search ranges $q_j + \sum_a v_{ja} x_a$. Straight paths in q correspond to curved paths in p, and vice versa, but over short segments there is a linear correspondence between directions and distances in the two domains

$$u_{ja} = H(\omega_j) v_{ja}, \quad D(s)^2 = D(\omega)^2 = H(\omega)(\Delta q)^2 = p(\Delta q^2). \tag{4.1}$$

Since $\omega(q)$ is normally an exponential function of q the curvature only becomes important

over force shifts of order $\Delta q = 1$. The ω-metric inhibits the largest force shifts at grid points with large densities, which often have the greatest effect on S and E. Bryan and Skilling used two basic search directions $u_a = -p\nabla S$ and $u_b = p\nabla(\frac{1}{2}E^2)$. The counterpart directions in the force domain, (now converted to Fourier space) are

$$v_{ha} = T_h = \frac{1}{\alpha_h}\nabla_h C, \quad v_{hb} = \frac{(t_h - t_h^0)}{\alpha_h} = \frac{1}{\alpha_h}\nabla_h Q. \tag{4.2}$$

These are just the gradients of Q and C preconditioned [15,16] by $H(C)_{hh} = \alpha_h$. They give high weight to the stringent constraints with small values of α_h. If λ is known the Y potential can be minimised easily by moving along these paths, or more efficiently by using a conjugate gradient method [15].

4.2 The Balance Curve

By a suitable choice [3] of basic path coordinates x_r with transformed search directions v_{hr} we reduce Q and C locally to positive definite diagonal quadratic forms

$$Q + \lambda C = \frac{1}{2}\sum_r \left[q_r''(x_r - q_r')^2 + \lambda c_r''(x_r - c_r')^2\right] \tag{4.3}$$

whose stationary points lie on the balance curve

$$x_r(\lambda) = \frac{q_r' + \lambda c_r'}{q_r'' + \lambda c_r''}. \tag{4.4}$$

The extreme points of this curve are the Q-point ($x_r = q_r'$), which is the best fit point in the search space for the constraints, and the C-point ($x_r = c_r'$) which has the smallest weighted force θ. The choice of λ_m to match $E = \epsilon_m$ is now made by searching along the $x(\lambda)$ curve for the unique position where $\lambda = \epsilon_m/\theta(\lambda)$. In difficult cases, where ϵ_m is not attainable, we may have to solve a more elaborate balanced model problem in search space of the type

$$\nabla_x(Q + \lambda C + \frac{1}{2}\mu F^2 + \frac{1}{2}\nu D^2) = 0 \tag{4.5}$$

with prescribed limits F_m and D_m on the mean force and the distance moved. The control algorithm starts at the current position x_{orig}, on the current force contour $F = F_{orig}$, and tries for a feasible move in order of preference: (1) towards the minimum point of G, at $x_G = x(\lambda_m)$; (2) down the G-F^2 balance curve towards the minimum of Ψ, which lies at $F = F_m$; (3) along the $F = F_{orig}$ contour towards its intersection with the G-F^2 balance curve. The move must satisfy both the D_m and F_m limits.

Acknowledgements

It is a pleasure to thank my colleagues Alastair Livesey, John Skilling and Steve Gull for introducing me to the art of maximum entropy analysis, for their many helpful suggestions, their encouragement and for critical guidance to avoid pitfalls. I also thank Gérard Bricogne for some expositions of his fundamental theories of direct methods and Bayesian statistics.

References

1. Gull, S.F. & Daniell, G.J. (1978). Nature **272**, 686-690.
2. Skilling, J. & Bryan, R.K. (1984). Mon. Not. R. Astron. Soc. **211**, 111-124.
3. Skilling, J. & Gull, S.F. (1985). In *Maximum-Entropy and Bayesian Methods in Inverse Problems.* Ed. C. Ray Smith & W.T. Grandy. pp. 83-132. Dordrecht Holland: Reidel Publishing Company.
4. Livesey, A.J. & Skilling, J. (1985). Acta Cryst. **A41**,113-122.
5. Skilling, J. (1986) In *Maximum Entropy and Bayesian Methods in Applied Statistics.* Ed. James H. Justice. Cambridge: University Press. pp.156-193.
6. Collins, D.M. (1982). Nature **298**, 49-51.
7. Gull, S.F., Livesey, A.K. & Sivia, D.S. (1987). Acta Cryst. **A43**, 112-117.
8. Wilkins, S.W., Varghese, J.N. & Lehmann, M.S. (1983). Acta Cryst. **A39**, 47-60.
9. Wilkins, S.W. (1983). Acta Cryst. **A39**, 892-896.
10. Wilkins, S.W. (1983). Acta Cryst. **A39**,896-898.
11. Navaza, J. (1985). Acta Cryst. **A41**, 232-244.
12. Navaza, J. (1986). Acta Cryst. **A42**, 212-223.
13. Bricogne, G. (1984). Acta Cryst.**A40**, 410-445.
14. McLachlan, A.D. (1987). Gazzetta Chimica Italiana **117**, 11-15.
15. Luenberger, D.G. (1984). *Linear and Non-linear Programming.* 2nd Edition. Reading Mass: Addison-Wesley Publishing Co.
16. Gill, P.E., Murray, W. & Wright, M.M. (1981). *Practical Optimisation* . New York: Academic Press.
17. Levine, R.D. (1976). J. Chem. Phys. **65**, 3302-3315.
18. Levine, R.D. (1986) In *Maximum Entropy and Bayesian Methods in Applied Statistics.* Ed. James H. Justice. Cambridge: University Press. pp.59-84.
19. Bryan, R.K., Bansal, M., Folkhard, W., Nave, C. & Marvin,D.A.(1983). Proc. Nat. Acad. Sci. USA **80**, 4728-4736.
20. Bryan, R.K. & Banner, D.W. (1987). Acta Cryst.**A43**, 556-564.
21. Bricogne, G. (1988). Acta Cryst.**A44**, 517-544.
22. Wang, B.C. (1985). In *Methods in Enzymology*, Vol 114. *Diffraction Methods in Biological Macromolecules.* Ed. H.W. Wyckoff, C.H.W. Hirs & S.N. Timasheff, pp. 114-167. New York: Academic Press.
23. Prince, E., Sjölin, L. & Alenljung, R. (1988). Acta Cryst. **A44**, 216-222.
24. Podjarny, A.D., Moras, D., Navaza, J. & Alzari, P.M. (1988). Acta Cryst. **A44**, 545-550.
25. Marvin, D.A., Bryan, R.K. & Nave, C. (1987). J. Mol. Biol. **193**, 315-343.

ENHANCED INFORMATION RECOVERY FROM SPECTROSCOPIC DATA USING MAXENT

A I GRANT AND K J PACKER
Spectroscopy Branch, BP Research International,
Research Centre, Chertsey Road,
Sunbury-on-Thames, Middlesex TW16 7LN

ABSTRACT

An outline is given of the application of MAXENT to the reconstruction/deconvolution of data from spectroscopic experiments. Examples are shown of its use in Raman and Nuclear Magnetic Resonance spectroscopies. In particular, quantitation and choice of lineshape function for deconvolution are addressed, with examples for the latter including the use of both analytic and experimentally determined functions

1. INTRODUCTION

All data generated by spectroscopic experiments are inadequate in some way. Data will be incomplete because of digitisation, finite bandwidth, truncation etc. and contain noise. The spectroscopist's aim is to interpret the data to yield information and this process will always draw on theoretical models of the relevant spectroscopy as well as, usually, some data manipulation.

This, latter, can take many forms but, typically, may involve filtering the data in some way. For example, it is standard practice in high-resolution nuclear magnetic resonance (NMR) spectroscopy to multiply the time domain interferogram, the free-induction decay (FID) by some arbitrary weighting function which can be chosen either to enhance resolution or signal-to-noise (S/N) ratio (1) in the derived spectrum. In infrared spectroscopy this process is referred to as Fourier self-deconvolution (2). Whilst these methods are valuable, quick, and indeed, often readily available through computing facilities provided with modern spectroscopic equipment, they are limited in scope because, for example, of the trade-off noted between S/N and resolution. In addition, they usually lack any degree of objectivity or figure of merit. Recently there has been increasing interest in other, more elegant, data processing methods of which that employing the Maximum Entropy (MAXENT) criterion is an example.

In this paper we present some results of investigations of the use of MAXENT in the processing of data from Raman and NMR spectroscopies. Our object in undertaking this work has been to evaluate the method as practising spectroscopists and to assess what it might offer in the way of improved information recovery from data.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 251–259.

The work is still in progress and this constitutes a preliminary report of our user's view of the method. We make no attempt to review or comment on other such investigations already published.

2. THE MAXENT METHOD

The immediately attractive aspect of the MAXENT method is that it utilises a well defined and rigorous criterion, that of maximum entropy (minimum information), to guide the spectroscopist. In addition, it immediately acknowledges the fact that, statistically, noise will have different properties from signal, a fact which is not made use of in the types of data processing referred to in the Introduction. Typically, an experimental data set $\underline{D}$ will be a representation of the ideal spectroscopic response $\underline{f}$ and noise $\underline{n}$. Thus, one model of this could be

$$\underline{D} = \underline{O}.\underline{f} + \underline{n} \qquad 1$$

in which $\underline{f}$ is blurred by some natural or instrumentally imposed function $\underline{O}$. The noise may have different qualities depending on the nature of the experiment but, most simply, can be modelled as being uncorrelated with $\underline{f}$ and having a Gaussian distribution with standard deviation, σ.

The blurring function, $\underline{O}$, may be a lineshape function and may also involve a mathematical relationship such as Fourier transformation etc. The form of equation 1 is not unique. Indeed, it is a model for the expected response of the experiment and draws on the spectroscopist's knowledge of that experiment. To that extent it is no different from the spectroscopist's choice of, such things as, amplifier bandwidth, slit width, sample configuration etc, all of which may influence the experiment, and constitute part of the prior knowledge the spectroscopist should use in interpreting his data.

The essence of the MAXENT procedure is to generate "guesses" or "reconstructions" of the function $\underline{f}$ and select that which, with given information on $\underline{O}$ and σ or, what ever other knowledge is available, are minimally consistent with the data $\underline{D}$. This involves maximising the entropy content of the reconstruction subject to some appropriate constraint on consistency with the data $\underline{D}$. All of the examples given below have employed the MEMSYS programme (3) which uses a χ^2 constraint. This programme has been integrated with software written by us for applications to particular forms of spectroscopic data.

3. RESULTS

3.1 The Raman Spectrum of CCl_4

The symmetric stretching mode of CCl_4 is a resolution standard for Raman spectroscopy. It comprises five bands arising from the five combinations of the two isotopes ^{35}Cl and ^{37}Cl. Figure 1(a) shows the Raman spectrum recorded with a slit width of 5 cm^{-1} which is too large to allow resolution of the underlying five components. Figure 1(b) is a MAXENT reconstruction obtained by assuming that $\underline{O}$ is a Lorentz lineshape with a FWHM of 4.5 cm^{-1} and giving the programme an experimental value for σ.

Figure 1(c) is an experimental spectrum recorded with a spectrometer slit width of 0.9 cm^{-1}. It can be seen that the reconstruction 1(b) is very similar to 1(c). The experimental process of narrowing the slit in order to increase resolution is eventually limited by deteriorating S/N, already apparent in 1(c). Figure 1(d) is a MAXENT reconstruction of the data in Figure 1(c), again using a Lorentz function for $\underline{0}$ but with a FWHM of 0.8 cm^{-1} and, as before, an experimental value for σ. Also shown in Figure 1(d) are the integrals for each line in the reconstruction

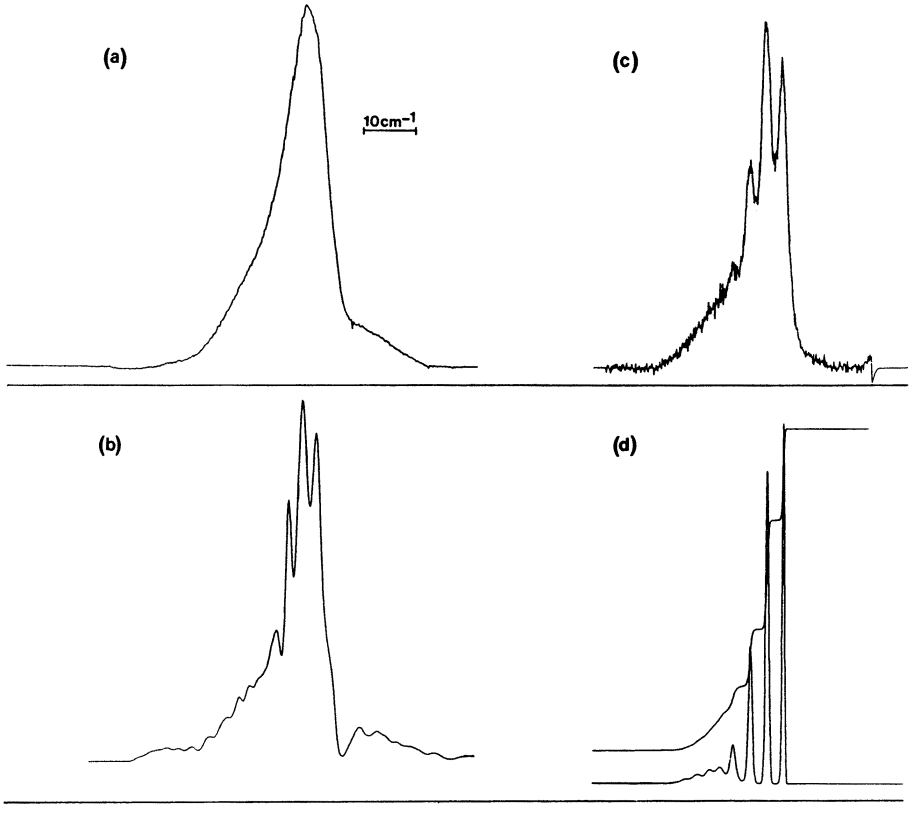


Figure 1 : Experimental (E) and MAXENT reconstruction (R) of the 459 cm^{-1} Raman band of CCl_4 a) E: slit width 5 cm^{-1}; b) R of a), Lorentzian profile FWHM 4.5 cm^{-1}; c) E: slit width 0.9 cm^{-1}; d) R of c), Lorentzian profile FWHM 0.8 cm^{-1}, plus integral.

and these are given in Table 1 together with the line intensities expected from the natural abundances of the two chlorine isotopes. The reconstructed spectrum of Figure 1(d) is quantitative in terms of intensity, shows remarkable resolution and an apparent S/N which is very large. Apparent is used to qualify the term S/N in the context of the reconstruction because what the process has done is to separate signal from noise and in this case has been very successful because our prior knowledge or information was itself of good quality. For the spectroscopist it is perhaps important to emphasise that all that was assumed was that the lines were Lorentzian (FWHM = $0.8cm^{-1}$) and the noise Gaussian (σ measured and used as input). Nothing was assumed concerning the number or value of the frequencies present in the required spectrum, $\underline{f}$.

TABLE 1

Band	ν_{CCl4}/cm^{-1}	MEM Image* Integral Intensity/%	Theoretical Intensity/%
$C^{37}Cl_4$	Not observed	N/A	0.4
$C^{37}Cl_3{}^{35}Cl_1$	452.0	6	4.7
$C^{37}Cl_2{}^{35}Cl_2$	456.4	20	21.1
$C^{37}Cl_1{}^{35}Cl_3$	459.4	41	42.2
$C^{35}Cl_4$	462.4	31	31.6

*Estimated error 0.5%: all figures rounded down to nearest integer value

3.2 The Raman Spectra of Graphite-$FeCl_3$ Intercalates

Figures 2(a) and (b) shows two Raman Spectra of a graphite-$FeCl_3$ intercalate which have very poor S/N. Figures 2(c) and (d) show MAXENT reconstructions carried out as for the CCl_4 case ie. assuming an Lorentzian lineshape (FWHM = 20 cm^{-1}) and using a measured σ. The results by any criteria are striking. The question which has to be addressed is whether they can be substantiated.

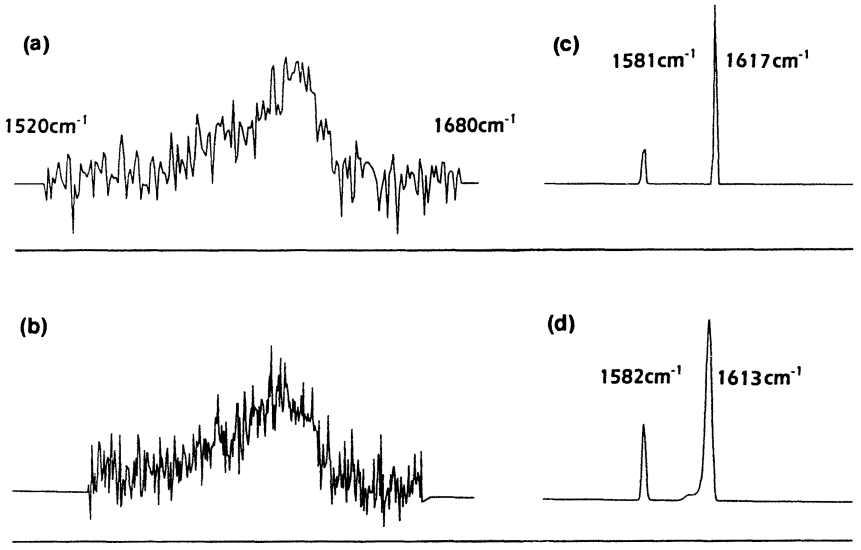


Figure 2 : The experimental (a and b) Raman spectra of an $FeCl_3$-graphite intercalate taken at different times and their MAXENT reconstructions (c and d) obtained as described in the text.

Literature data predict bands at the positions observed in Figures 2(c) and (d) for intercalates with staging between 2-3 and, perhaps, somewhat higher. This is consistent with the nature of the samples which gave rise to the spectra of Figure 3 and, as we show in more detail elsewhere (4), leads to confidence that the features observed are real and reliable. Again, for the spectroscopist, once this is established the use of MAXENT makes possible experiments which were not feasible before. However, as always, the spectroscopist must be critical in the use of this tool, as with any other.

3.3 The Raman Spectrum of the Sulphate ion

Both of the above examples have used MAXENT to reconstruct deconvoluted spectra using an assumed, analytical lineshape. Whilst the chosen function, Lorentzian in these examples, may often be expected, there may be many situations in which the appropriate function is not known. However, in some cases there may be ways of determining the relevant line profile from experimental observations. The sulphate ion has been used to test this approach. Figure 3(a) is a composite spectrum obtained by recording the Raman band of aqueous sulphate ion around 980 cm^{-1} and adding to it a band derived by shifting the same band by 0.4 of the FWHM and multiplying its intensity by 0.75.

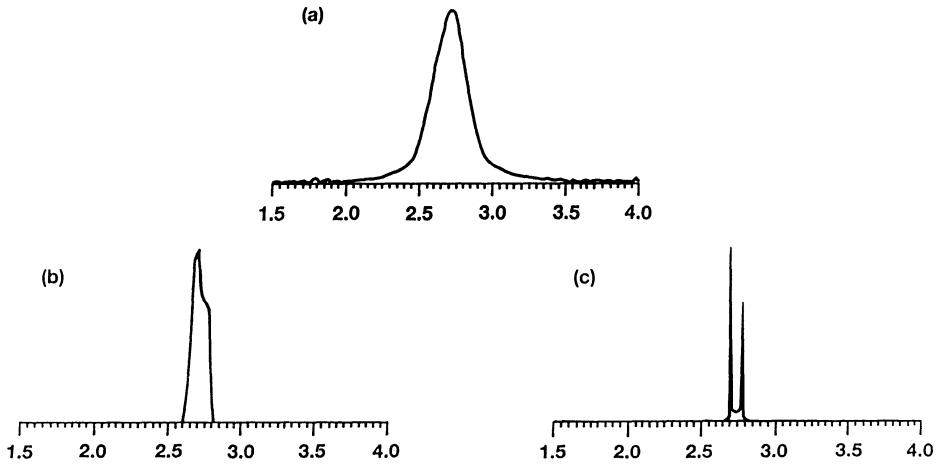


Figure 3 : The effects of the use of a non-analytic, experimentally determined lineshape in MAXENT reconstruction/deconvolution. The spectrum in (a) was obtained by taking an experimental Raman spectrum of aqueous sulphate ion and, adding to it a shifted (0.4FWHM) and attenuated (0.75) version of itself. (b) is an optimised MAXENT reconstruction/deconvolution using an assumed Lorentz lineshape whilst (c) is a MAXENT reconstruction/deconvolution using the learnt experimental lineshape at the pattern match function.

Two points are worth noting. First, the 980cm^{-1} sulphate band is symmetric but non-Lorentzian and, secondly, the composite spectrum in Figure 3(a) is not readily discerned as comprising two bands. Figures 3(b) and (c) are two MAXENT reconstructions of the data in Figure 3(a). Figure 3(b) was obtained assuming a Lorentzian lineshape and has been optimised for resolution by adjusting the FWHM. Figure 3(c) was obtained by using the digitised experimental lineshape of the 980 cm^{-1} band as the pattern matching profile in the MAXENT reconstruction. It is clear that in Figure 3(c) remarkable and quantitative resolution of the two features is achieved. A (properly) critical spectroscopist might suggest that it seems as if fore-knowledge of the answer (ie. the correct lineshape) is required to obtain good results. All this example illustrates, we would suggest, is that the better the quality of your model or prior knowledge the better the results achievable. The MAXENT process serves only to make the reconstructions minimally consistent with the data. The noise residuals provide a clear indication of the validity of the model employed.

3.4 ^{29}Si High-resolution NMR of a high-silica zeolite

The last example involves the use of an experimentally determined lineshape for the "deconvolution" or "pattern-matching". Figure 4(a) shows the ^{29}Si high-resolution NMR spectrum of a highly crystalline sample of the zeolite silicalite (5). The multiplicity of lines is interpreted as arising from the existence of 24 non-equivalent crystallographic sites for silicon in the unit cell. The peak at highest frequency corresponds to a single site and the lineshape of this peak has been used in the deconvolution process. Figure 4(b) is a MAXENT reconstruction of the spectrum in Figure 4(a). The raw data used was the experimental FID (5), the lineshape derived from Figure 4(a) and a measure of the noise σ. Also shown in Figure 4(b) is an integral trace.

A number of interesting points emerge from a comparison of Figures 4(a) and (b). In Figure 4(a), the spectrum, derived by Fourier transformation of the raw FID data, has eighteen distinguishable maxima and a shoulder discernible on the low frequency side of the largest spectral feature. Figure 4(b) has twenty maxima with most lines resolved from neighbours down to baseline level. In particular, the group of features labelled B in Figure 4(a), comprising a broad line flanked by a shoulder to low frequency and a barely-resolved peak to high frequency, are revealed as a group of four lines with relative intensities 1:3:1:1 (increasing frequency).

Both the shoulder feature and the partly resolved feature to high frequency in Figure 4(a) are both resolved to baseline level. It is also of interest to note that the integrals accurately reflect the 1:3:1:1 ratios and the total intensity of this feature ie. 6, relative to those lines of unit intensity.

A second interesting point is revealed by examination of the effect of MAXENT reconstruction on the group of three lines marked A in Figure 4(a). This group are maintained as three lines in Figure 4(b) with resolution dramatically enhanced. The overall intensity of this group of three lines is, as required, 5. However, the integrals for the individual lines in the group are close to the ratios $1^1/_3:1^2/_3:2$ instead of the expected 1:2:2 based on a 24 site structure. At present, we have no explanation for this feature which requires further investigation. It seems unlikely that it arises from the MAXENT procedure but this requires a more thorough examination before it can be completely ruled out.

A further feature to note is that the lines representing single sites, and which are well resolved, do not all have the same height. This is particularly noticeable for the second peak from the low frequency end. There are also a number of small features eg. to low frequency of the peak just referred to and between the two peaks at highest frequency. The variation in height implies that the lineshapes/widths for these lines are not identical to the pattern match profile used ie. the lineshape for the highest frequency line. This can be seen clearly in the FT spectrum in Figure 4(a). If the lineshape deviates significantly from the chosen pattern-match then it is possible for false peaks to appear (4) and this should always be born in mind. However, this is no more of a problem than is always encountered in the type of data treatments referred to in the introduction where, as resolution is increased, S/N decreases and the distinguishing of spurious side-lobes from real information becomes increasingly difficult and does not appear to be a problem in this example. The integral traces show that even the faster relaxing features still produce unit integral values for peak area.

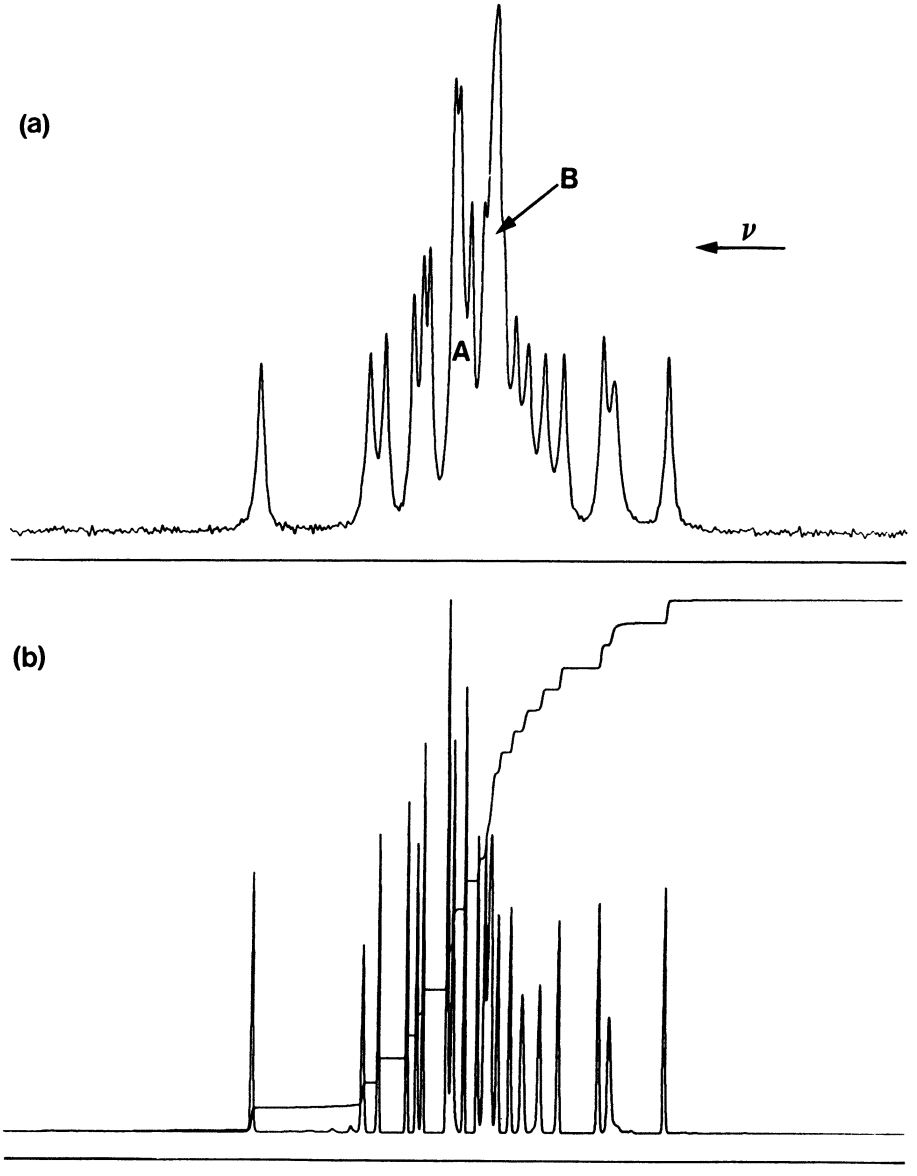


Figure 4 : (a) the experimental ^{29}Si MAS NMR spectrum of a highly crystalline sample of the zeolite, silicalite, obtained by direct Fourie transformation of the free induction decay. (b) The MAXENT reconstruction/deconvolution of the data used to derive spectrum (a). The pattern match profile used was the lineshape of the resonance at higher frequency.

4. SUMMARY AND CONCLUSIONS

In each of the examples presented above, MAXENT reconstructions have revealed enhanced and, in some cases, new information recovery from the raw data. No attempt has been made here to present a systematic evaluation of the scope of this approach but, as practising spectroscopists in an industrial context, the results obtained so far show considerable promise for enhancing information recovery from all manner of spectroscopic and similar data records. The enhancement may result in reduced time for acquiring experimental data on expensive instrumentation and new information not accessible in a reliable way using alternative approaches.

5. ACKNOWLEDGEMENTS

We thank Mr C J Dyos for introducing one of us (AIG) to the details of MEMSYS and Professor C A Fyfe for providing both the raw FID data for the ^{29}Si NMR of silicalite and valuable critical discussions of the results of the MAXENT treatment of this data. BP Research International are thanked for permission to publish this work.

6. REFERENCES

1. Ernst, R R, Adv. In Mag. Res. (AP) (1966), 2, pp1-135.

2. Hawkes, S, Maddams, W F, Mead W C and Southon, M J, Spectr. Chim Acta A, (1982) 38, 337.

3. Bryan R K and Skilling J, J Mon. Not. R. Ast Soc. 191, 69, (1980).

4. A I Grant and K J Packer, to be published.

5. ^{29}Si FID of silicalite provided by Professor C A Fyfe, Department of Chemistry, University of British Columbia, Vancouver, Canada.

BAYESIAN SPECTRUM ANALYSIS ON QUADRATURE NMR DATA WITH NOISE CORRELATIONS

G. LARRY BRETTHORST
Department of Chemistry
Campus Box 1134
Washington University
1 Brookings Drive
St. Louis, MO 63130

Abstract. In NMR data analysis a great deal of prior information is available. We know, in general terms, what characteristic signal will be received, that for quadrature measurements it will be the same in both channels and that the noise is potentially correlated. We have shown in previous work [1], [2] that when prior information is incorporated into the analysis of data, the frequencies, decay rates, and amplitudes may be estimated much more precisely than by using the discrete Fourier transform directly. Here we extend the Bayesian analysis to include the quadrature nature of the data and noise correlations. We then show that in typical NMR data the frequencies and decay rates may be estimated with a precision several orders of magnitude better than directly from the discrete Fourier transform.

Introduction

In NMR, theory tells us that the free induction decay time series must be sinusoidal with exponential or Gaussian decay. When this information is incorporated into the spectral estimation problem, one may estimate the frequencies and decay rates much more accurately than directly from a discrete Fourier transform of the data [1], [2]. More importantly this information allows one to separate frequencies and decay rates that are too close for one to resolve using a discrete Fourier transform. The initial work [1] did not incorporate all of the information we possessed about NMR signals. We used the functional form of the signal, but we utilized the data as if two distinct measurements were available having the same frequencies and decay rates, but different amplitudes and phases. This gave $\sqrt{2}$ improvement in the parameter estimates. However, we have more information; in particular, we know that the signal in the second channel is 90° out of phase with that in the first channel. Also, we know that the noise is potentially correlated, and that the phases of all the sinusoids are typically the same. When more information is incorporated into a probability

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 261–273.

calculation, we expect that information to improve the estimates of the parameters. In this paper we specialize the Bayesian calculation to include quadrature, noise correlations, and phase coherences.

The General Model Equation

The basic model we are considering is: given a quadrature detected data set (*i.e.*, two data sets collected with a 90° phase difference), then the real data may be modeled as

$$d_R(t_i) = f_R(t_i) + n(\sigma, 0)$$

where $n(\sigma, 0)$ is a Gaussian noise component of mean zero and standard deviation σ, $f_R(t)$ is a model of the real signal, and the quadrature or imaginary data may be modeled as

$$d_I(t_i) = f_I(t_i) + n(\sigma, 0).$$

The basic problem we would like to solve is: "what are the best estimates of the parameters (frequencies and decay rates) hidden in f_R and f_I that one can make from the data and the prior information?" We will solve this problem using Bayesian probability theory and apply the calculation to several examples.

We write the model equations $f_R(t)$ and $f_I(t)$ as a sum over functions G_j and F_j such that

$$f_R(t) = \sum_{j=1}^{m} B_j G_j(\{\omega\}, t) \quad \text{and} \quad f_I(t) = \sum_{j=1}^{m} B_j F_j(\{\omega\}, t) \tag{1}$$

where m is the total number of model functions, B_j is the amplitude of the jth model function, and $G_j(\{\omega\}, t)$ and $F_j(\{\omega\}, t)$ are typically sinusoids with either exponential or Gaussian decay. The model functions F_j and G_j are functions of a continuous variable time t; however, the data have been sampled at discrete times $\{t_1, \cdots, t_N\}$. Additionally, the models are functions of other continuous parameters, which we collectively label $\{\omega\}$. These parameters are frequencies, decay rates or any other parameters which could be needed to model the data, for example the phase if it is the same on all of the sinusoids. Although the amplitudes $\{B\}$ are of substantial interest, for the purposes of analyzing the data, we wish to formulate the problem independently of these parameters to see what probability theory can tell us about the frequencies and decay rates. The quadrature information has been incorporated by assuming the amplitudes B_j are the same in both channel.

We would like to compute the posterior probability of the frequencies and decay rates, given the data D and the prior information I. This requires us to obtain two terms: the direct probability of the data and the prior probability of the parameters. We will compute the direct probability of the data first. Making the standard

assumptions about the noise, the direct probability of the data is:

$$\begin{aligned} P(D|\{B\},\{\omega\},\sigma,\rho,I) &= (2\pi\sigma^2)^{-N}(1-\rho^2)^{-\frac{N}{2}} \\ &\times \exp\left\{-\sum_{i=1}^{N}\frac{[d_R(t_i)-f_R(t_i)]^2+[d_I(t-i)-f_I(t_i)]^2}{2\sigma^2(1-\rho^2)}\right\} \\ &\times \exp\left\{-2\rho\sum_{i=1}^{N}\frac{[d_R(t_i)-f_R(t_i)][d_I(t_i)-f_I(t_i)]}{2\sigma^2(1-\rho^2)}\right\} \end{aligned}$$

where ρ is the correlation coefficient – see Jeffreys [3] for a discussion of correlation, and [2], [4] for a discussion of when a Gaussian should be used to represent the noise. The symbol I in $P(D|\{B\},\{\omega\},\sigma,\rho,I)$ is there as a reminder that all probability distributions are computed based on our prior information I. Now substituting model Eq. (1) we have the direct probability of the data given the parameters:

$$P(D|\{\omega\},\{B\},\sigma,\rho,I) = (2\pi\sigma^2)^{-N}(1-\rho^2)^{-\frac{N}{2}}\exp\left\{-\frac{Q}{2\sigma^2(1-\rho^2)}\right\},$$

where

$$\begin{aligned} Q &\equiv d_R\cdot d_R - 2\rho d_R\cdot d_I + d_I\cdot d_I \\ &- 2\sum_{j=1}^{m}B_j[d_R\cdot G_j - \rho(d_R\cdot F_j + d_I\cdot G_j) + d_I\cdot F_j] \\ &+ \sum_{j=1}^{m}\sum_{k=1}^{m}B_jB_k[G_j\cdot G_k - \rho(G_j\cdot F_k + G_k\cdot F_j) + F_j\cdot F_k] \end{aligned}$$

and $(\cdot)$ means the sum over the discrete times: $d_I\cdot F_j \equiv \sum_{i=1}^{N} d_I(t_i)F_j(t_i)$.

Bayes' theorem tells us that the posterior probability of the nonlinear $\{\omega\}$ parameters, independently of the amplitudes, given the data and our prior information is

$$P(\{\omega\},\sigma,\rho|D,I) \propto \int d\{B\}P(\{B\},\{\omega\},\sigma,\rho|I)P(D|\{B\},\{\omega\},\sigma,\rho,I),$$

where $P(\{B\},\{\omega\},\sigma,\rho|D,I)$ is the posterior probability of the parameters, the direct probability of the data is $P(D|\{B\},\{\omega\},\sigma,\rho,I)$, and $P(\{B\},\{\omega\},\sigma,\rho|I)$ represents what was known about these parameters before we took the data and is called a prior probability. In this problem we assume that the data determine the parameters much more accurately than our prior information. Therefore, we assign a broad uninformative prior to the parameters: we use a uniform prior for the amplitudes and a Jeffreys prior for the variance.

Introducing the transformation

$$B_k = \sum_{j=1}^{m}\frac{A_j e_{jk}}{\sqrt{\lambda_j}}, \quad R_k = \sum_{j=1}^{m}\frac{G_j e_{kj}}{\sqrt{\lambda_k}}, \quad I_k = \sum_{j=1}^{m}\frac{F_j e_{kj}}{\sqrt{\lambda_k}},$$

and

$$dB_1 \cdots dB_m = \lambda_1^{-\frac{1}{2}} \cdots \lambda_m^{-\frac{1}{2}} dA_1 \cdots dA_m$$

where e_{jk} is the kth component of the jth eigenvector of the interaction matrix

$$g_{jk} \equiv \sum_{i=1}^{N} G_j(t_i)G_k(t_i) - \rho(G_jF_k + G_kF_j) + F_j(t_i)F_k(t_i) \tag{2}$$

and λ_j is the jth eigenvalue, then the posterior probability of the parameters becomes

$$P(\{\omega\}, \sigma, \rho | D, I) \propto \sigma^{-2N}(1-\rho^2)^{-\frac{N}{2}} \lambda_1^{-\frac{1}{2}} \cdots \lambda_m^{-\frac{1}{2}} \int_{-\infty}^{\infty} dA_1 \cdots dA_m \exp\left\{-\frac{Q'}{2\sigma^2(1-\rho^2)}\right\}$$

where

$$Q' = d_R \cdot d_R - 2\rho d_R \cdot d_I + d_I \cdot d_I - 2\sum_{j=1}^{m} A_j h_j + \sum_{j=1}^{m} A_j^2$$

and

$$h_j(\{\omega\}, \rho) \equiv d_R \cdot R_j - \rho(d_R \cdot I_j + d_I \cdot R_j) + d_I \cdot I_j.$$

After completing the square in Q' and performing the m integrals, we have

$$\begin{aligned} P(\{\omega\}, \sigma, \rho | D, I) &\propto \sigma^{m-2N}(1-\rho^2)^{-\frac{N-m}{2}} \lambda_1^{-\frac{1}{2}} \cdots \lambda_m^{-\frac{1}{2}} \\ &\times \exp\left\{-\frac{d_R \cdot d_R - 2\rho d_R \cdot d_I + d_I \cdot d_I - m\overline{h^2}}{2\sigma^2(1-\rho^2)}\right\} \end{aligned} \tag{3}$$

where

$$\overline{h^2} \equiv \frac{1}{m}\sum_{j=1}^{m} h_j^2.$$

If the variance of the noise σ^2 and the correlation coefficient ρ are known, then the problem is completed. The posterior probability of the frequencies and decay rates conditional on the data and our assumed knowledge of σ and ρ is

$$P(\{\omega\} | \sigma, \rho, D, I) \propto \lambda_1^{-\frac{1}{2}} \cdots \lambda_m^{-\frac{1}{2}} \exp\left\{\frac{m\overline{h^2}}{2\sigma^2(1-\rho^2)}\right\}. \tag{4}$$

But if σ is not known, then it too becomes a nuisance parameter to be removed by integration. Multiplying Eq. (3) by a Jeffreys prior and integrating with respect to σ, we obtain the posterior probability of the frequencies, decay rates, and the correlation coefficient ρ

$$P(\{\omega\}, \rho | D, I) \propto \lambda_1^{-\frac{1}{2}} \cdots \lambda_m^{-\frac{1}{2}} (1-\rho^2)^{\frac{N-2m}{2}} \left[1 - \frac{2\rho d_R \cdot d_I + m\overline{h^2}}{d_R \cdot d_R + d_I \cdot d_I}\right]^{\frac{m-2N}{2}} \tag{5}$$

where $\overline{h^2}$ is a sufficient statistic for inferences about the $\{\omega\}$ parameters. Equation (5) is an exact result and does not depend on uniform sampling nor does it depend on the models being sinusoidal. Any quadrature data set that can be modeled by Eq. (1) can be used in these equations.

The Single Stationary Harmonic Frequency

What is to be gained from the use of Eq. (4) or (5) compared to a discrete Fourier transform of the data? The answer to this question is easily demonstrated by investigating one of the simplest quadrature spectral estimation problems: the single stationary harmonic frequency. Suppose we take

$$f_R(t) = B_1 \cos \omega t + B_2 \sin \omega t$$

as the model for the signal in the real channel and

$$f_I(t) = B_1 \sin \omega t - B_2 \cos \omega t$$

as the model for the signal in the imaginary channel. If the noise is uncorrelated, *i.e.*, $\rho = 0$, the interaction matrix, Eq. (2), becomes

$$g_{jk} = \begin{pmatrix} N & 0 \\ 0 & N \end{pmatrix}.$$

The R_j and I_j functions are given by

$$R_1 = \frac{\cos \omega t}{\sqrt{N}}, \quad R_2 = \frac{\sin \omega t}{\sqrt{N}}, \quad I_1 = \frac{\sin \omega t}{\sqrt{N}}, \quad I_2 = -\frac{\cos \omega t}{\sqrt{N}}.$$

The sufficient statistic $\overline{h^2}$ is given by

$$\begin{aligned} \overline{h^2} &= \frac{1}{2}\left[(R_1 \cdot d_R + I_1 \cdot d_I)^2 + (R_2 \cdot d_R + I_2 \cdot d_I)^2\right] \\ &= \frac{1}{2N}\left\{[C_R(\omega) + S_I(\omega)]^2 + [S_R(\omega) - C_I(\omega)]^2\right\} \end{aligned}$$

where

$$C_R(\omega) \equiv R_1 \cdot d_R = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_R(t_i) \cos \omega t_i$$

and

$$S_R(\omega) \equiv R_2 \cdot d_R = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} d_R(t_i) \sin \omega t_i$$

are the cosine and sine transforms of the real data, and $C_I(\omega)$ and $S_I(\omega)$ are the transforms for the imaginary data. The posterior probability of a stationary harmonic frequency ω, given the variance of the noise σ^2 and assuming the noise is uncorrelated, is

$$P(\omega|\sigma, D, I) \propto \exp\left\{\frac{[C_R(\omega) + S_I(\omega)]^2 + [S_R(\omega) - C_I(\omega)]^2}{2N\sigma^2}\right\}.$$

How does this compare to a discrete Fourier transform of the data? If we assume the data are the real and imaginary parts of a complex data set, then

$$d(t_i) = d_R(t_i) + id_I(t_i).$$

Because

$$e^{-i\omega t} = \cos\omega t - i\sin\omega t,$$

the squared magnitude of the discrete Fourier transform may be written

$$\left|\sum_{k=1}^{N}[d_R(t_k) + id_I(t_k)]\,e^{-i\omega t_k}\right|^2 = [C_R(\omega) + S_I(\omega)]^2 + [S_R(\omega) - C_I(\omega)]^2\,.$$

Up to the constant factor $1/2N$ the sufficient statistic $\overline{h^2}$ is the squared magnitude of a discrete Fourier transform of the complex data. Therefore, the discrete Fourier transform is essentially the natural logarithm of the posterior probability of a stationary harmonic frequency, given the variance of the noise σ^2, assuming the noise is uncorrelated, and assuming the channels are exactly 90° out of phase.

The implications of this are quite profound, because it means that only the highest peak in a discrete Fourier transform is of any importance for the estimation of a single stationary frequency, and then it is only the region around the maximum that is of importance. Moreover, the discrete Fourier transform will always interpret the data in terms of a single stationary harmonic frequency. If the data does not contain a single stationary harmonic frequency, or even if the data contain more than one stationary frequency, the discrete Fourier transform may give misleading or even incorrect results when compared to other more complex models. This is not because the discrete Fourier transform is wrong, but because it is answering what we should regard as the wrong question.

If we know that the signal consists of a single stationary harmonic frequency, how accurately can a frequency be estimated? We will assume that the data contain a single stationary sinusoid with no noise. Thus the accuracy estimates we derive will be optimistic in the sense that in real data, with a given noise variance σ^2, one would always make slightly worse frequency estimates than the ones we will derive. We take

$$d_R(t_i) = \hat{B}\cos\hat{\omega}t_i \quad \text{and} \quad d_I(t_i) = \hat{B}\sin\hat{\omega}t_i$$

as the signal in the real and imaginary channels, where $\hat{B}$ is the true amplitude of the sinusoid and $\hat{\omega}$ is the true frequency. We have set the phase of this sinusoid to zero. It will be obvious at the end of the calculation that the result for an arbitrarily phased sinusoid may be obtained by the replacement $\hat{B}^2 \to \hat{B}_1^2 + \hat{B}_2^2$. For uniformly sampled data we may take t_i to be integer or half integer, *i.e.*, $t_i = \{-T, -T+1, \cdots, T\}$ and

$2T + 1 = N$. The sufficient statistic $\overline{h^2}$ is

$$\begin{aligned} \overline{h^2} &= \frac{1}{2N}\left[\sum_{i=1}^{N}\hat{B}(\cos\hat{\omega}t_i\cos\omega t_i + \sin\hat{\omega}t_i\sin\omega t_i)\right]^2 \\ &\approx \frac{\hat{B}^2}{2N}\left[\frac{\sin\frac{N}{2}(\hat{\omega}-\omega)}{\sin\frac{1}{2}(\hat{\omega}-\omega)}\right]^2 \end{aligned} \tag{6}$$

where we have explicitly performed the sum and have ignored terms of order one compared to N.

To estimate the accuracy of the frequency, we Taylor expand $\overline{h^2}$ in posterior probability

$$P(\omega|\sigma, D, I) \propto \exp\left\{\frac{\overline{h^2}}{\sigma^2}\right\}$$

around the maximum, and then make the (mean) $\pm$ (standard deviation) approximation. Around the maximum, the first derivative of $\overline{h^2}$ is zero, and the second is given by

$$\frac{\partial^2\overline{h^2}}{\partial\omega^2} \approx -\frac{\hat{B}^2N^3}{12}.$$

The Gaussian approximation to the posterior probability density is

$$P(\omega|\sigma, D, I) \approx \left(\frac{\hat{B}^2N^3}{24\pi\sigma^2}\right)^{\frac{1}{2}}\exp\left\{-\frac{\hat{B}^2N^3}{24\sigma^2}(\hat{\omega}-\omega)^2\right\},$$

from which we estimate the frequency to be

$$(\omega)_{\text{est}} = \hat{\omega} \pm \frac{\sigma}{|\hat{B}|}\sqrt{\frac{12}{N^3}},$$

or in Hertz

$$(f)_{\text{est}} = \hat{f} \pm \frac{\sigma}{2\pi|\hat{B}|T}\sqrt{\frac{12}{N}} \quad \text{Hz},$$

where T is now the total sampling time in seconds. The accuracy of the frequency estimate depends on the signal-to-noise ratio of the data, on the $\sqrt{N}$, and on the total sampling time T. The better the data, the better the estimate. If we double the number of data in the given sampling time we obtain the standard $\sqrt{2}$ improvement. However, if we sample two times longer, we pick up a factor of 2 from sampling longer and a factor of $\sqrt{2}$ from taking two times more data. Clearly for stationary frequencies taking data for a long time is the preferred way to sample the data.

In many NMR applications the discrete Fourier transform is taken directly as a frequency estimator. The accuracy is estimated from the full-width-at-half-maximum of the peak in the discrete Fourier transform. For the case just given, the squared

magnitude of the discrete Fourier transform of the data (up to a constant) is given by Eq. (6). This has dropped to half its maximum value when the argument of the sine function has dropped to $\pi/4$:

$$\frac{N}{2}|\hat{\omega} - \omega| = \frac{\pi}{4}.$$

Thus for the discrete Fourier transform we find that the frequency estimate, in Hertz, is

$$(f)_{\text{est-dft}} = \hat{f} \pm \frac{1}{4T} \quad \text{Hz}$$

which neither depends on the magnitude of the signal nor the variance of noise σ^2.

Suppose we collect data for 1 second, with $\Delta T = 0.001$ seconds, collecting $N = 1000$ data values, and suppose we have RMS signal-to-noise ratio of $\hat{B}/\sqrt{2}\sigma = 1$. From the discrete Fourier transform we estimate the frequency to be

$$(f)_{\text{est-dft}} = \hat{f} \pm 0.25 \text{ Hertz},$$

and the Bayesian estimate is

$$(f)_{\text{est}} = \hat{f} \pm 0.012 \text{ Hertz}.$$

With signal-to-noise ratio of one, the Bayesian result is about 20 times better than the result from the discrete Fourier transform. If the signal-to-noise ratio were more typical of an NMR experiment, for example 100, then the Bayesian estimate would be more than three orders of magnitude better! Thus the probability analysis can estimate the frequency several orders of magnitude more precisely than a discrete Fourier transform directly. But this was in noiseless data. In practice, for frequency estimation, these procedures work at their theoretical best. However, the same cannot be said for other types of model functions. The reason frequency estimation is so accurate has to do with an interaction between the noise and the model functions. The oscillatory model functions and the noise tend to average to zero. When one computes the sufficient statistic, there is a sum of the model function times the data. Since the model and the noise are summing to zero separately, the sum of the product between the model and the noise tends to zero. This insures the projection of the model onto the noise is small compared to the projection of the model onto the signal, and the accuracy of the estimates are near the theoretical best. If the noise or the model did not average to zero, the accuracy estimates would be much worse.

The Single Frequency with Exponential Decay

In NMR the time series is typically the result of a complex chain of events: a sample is placed in a high magnetic field, and the nuclear spins are "excited" using a radio transmitter. These spins are then detected as they relax back to equilibrium. Using

an RF antenna, the signal is amplified, split, mixed with a reference oscillator (a sine or cosine) oscillating with a frequency near the natural resonance of the sample, and low-pass filtered. The beats between the reference oscillator and the sample resonance are what is digitized and recorded. Because the signal in the two channels originated in the same physical event there is reason to expect the noise to be correlated. To give an understanding of what noise correlation can do for estimating the parameters we give a second example. We will use simulated data with noise correlations.

The data used in this example were generated from the following equations:

$$f_R(t_i) = 100\cos(0.3t_i + 1)\exp\{-0.01t_i\},$$

$$f_I(t_i) = 100\sin(0.3t_i + 1)\exp\{-0.01t_i\}.$$

To generate the data we first generated the signal from the above equations and then generated the noise. We generated the noise for the real channel from a Gaussian distributed random number generator with unit variance. To generate the noise for the imaginary channel we generated a second random number with unit variance and then added the noise from the real channel to this second random number. This was divided by $\sqrt{2}$ and then used as the noise in the imaginary channel. The noise in the two channels is, thus, slightly correlated.

The data and the discrete Fourier transform are displayed in Fig. 1. The data resemble an NMR signal which rapidly decays. There are $N = 512$ data values, and the signal-to-noise ratio is approximately 50. The discrete Fourier transform has a peak in the correct vicinity of the frequency. However, the width of the discrete Fourier transform is indicative of the decay rate, not the accuracy of the frequency estimate.

We now apply the results of this calculation to the data. The model we use is

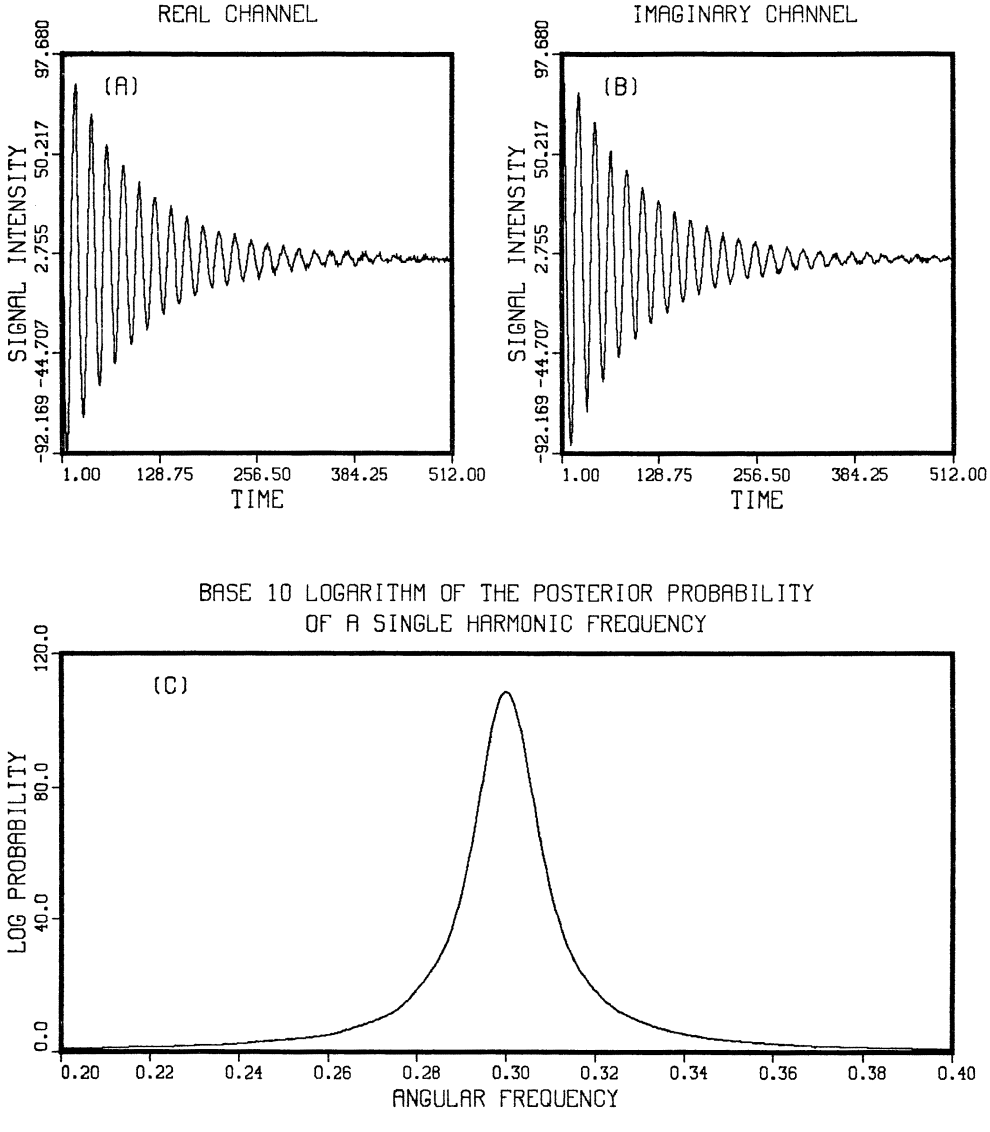
$$f_R(t) = B_1\cos\omega t\exp\{-\alpha t\} + B_2\sin\omega t\exp\{-\alpha t\}$$

for the real channel and

$$f_R(t) = B_1\sin\omega t\exp\{-\alpha t\} - B_2\cos\omega t\exp\{-\alpha t\}$$

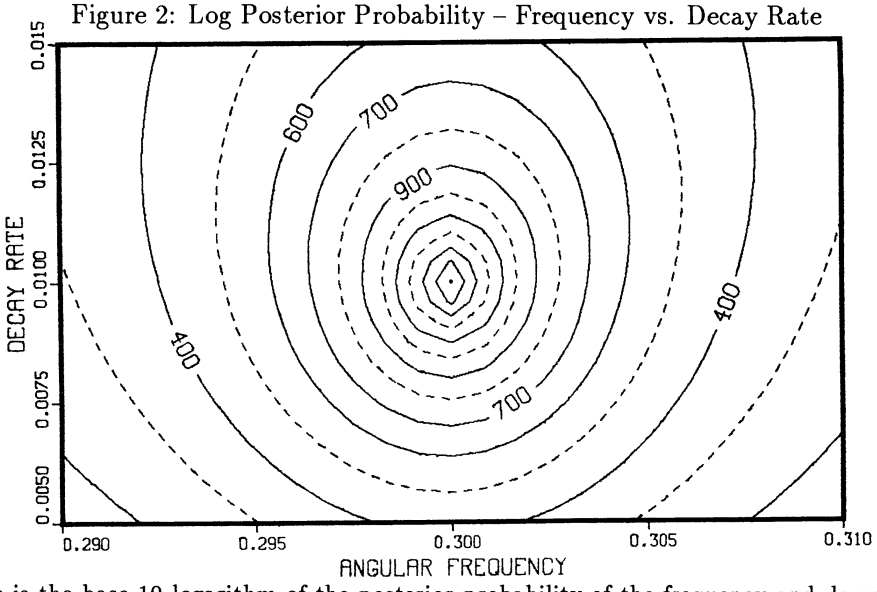
for the imaginary channel. After integrating out the amplitudes and variance of the noise, there are three remaining parameters to be estimated from the data: the frequency ω, the decay rate α, and the correlation coefficient ρ. We present the result of the calculation as three contour plots. First we plot the base 10 logarithm of the posterior probability of the frequency and decay rate while holding the correlation coefficient at its correct value. This is displayed in Fig. 2. We can see from this plot that there is a very sharp peak in the parameter space around the true value of the parameters. The normalization on this figure is irrelevant because of an interesting result, first noted by Jaynes [5]. If the contour lines are in increments of 1 (for example if the maximum posterior probability density were 100 and the contours be

Figure 1: The Computer Simulated Data and the Discrete Fourier Transform



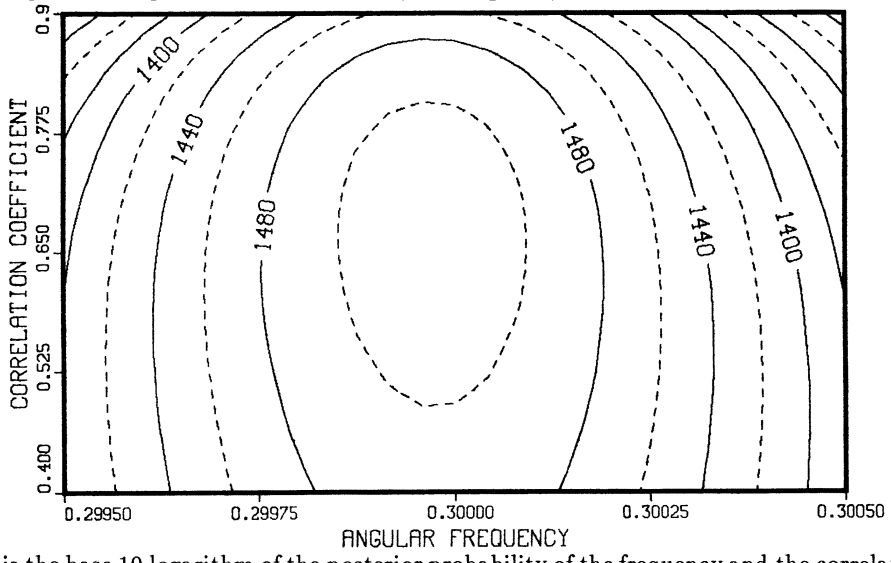
This computer simulated data (A) contain a single frequency which rapidly decays. The signal-to-noise ratio in these data is approximately 50. Now the discrete Fourier transform indicates the presence of a frequency in the right location. However, the width is indicative of the decay rate, not the accuracy of the estimate. Additionally, the discrete Fourier transform knows nothing of the noise correlations.

Figure 2: Log Posterior Probability – Frequency vs. Decay Rate



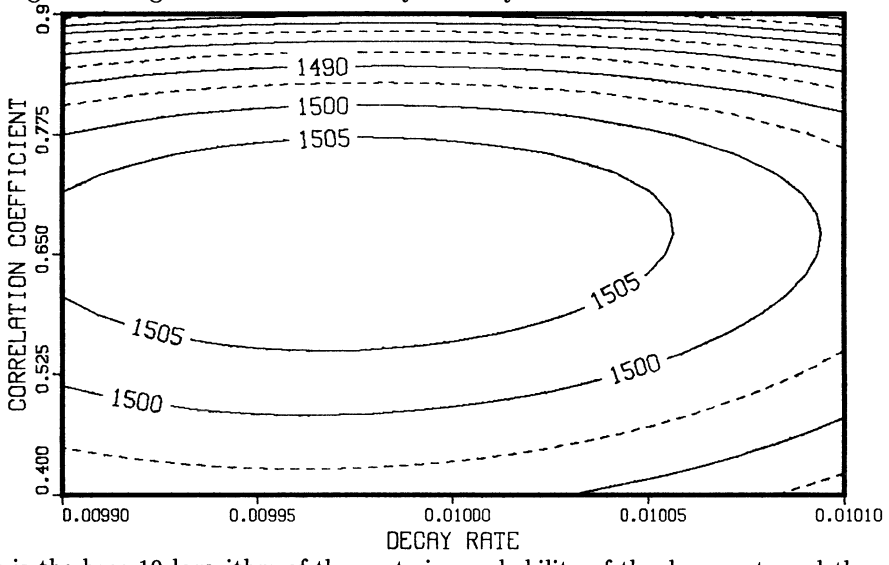
This is the base 10 logarithm of the posterior probability of the frequency and decay rate given the correlation coefficient. The total probability inside the highest contour is nearly 1. To write it out would require a decimal point followed by a string of approximately 200 nines.

Figure 3: Log Posterior Probability – Frequency vs. Correlation Coefficient



This is the base 10 logarithm of the posterior probability of the frequency and the correlation coefficient given the decay rate. The total probability inside the highest contour is nearly 1. Here it would require only a string of 20 nines to write it out.

Figure 4: Log Posterior Probability – Decay Rate vs. Correlation Coefficient



This is the base 10 logarithm of the posterior probability of the decay rate and the correlation coefficient given the frequency. The total probability inside the highest contour is approximately 0.99999.

labeled 99, 98, etc.), then for a 2D contour plot the first contour line contains 90% of the posterior probability, the second contour line contains 99% of the posterior probability, etc. Therefore, Fig. 2 represents an incredibly sharply peaked posterior probability density. The region inside the first contour line contains essentially all of the posterior probability. To write out the total probability enclosed by this contour would require a decimal point followed by a string of 200 nines. The second contour plot, Fig. 3, is the base 10 logarithm of the posterior probability of a frequency and the correlation coefficient given the true decay rate. Again there is a very sharp peak. The probability enclosed by the highest contour is approximately one; however, it would require only 20 nines to write it out. The third contour plot, Fig. 4, is of the base 10 logarithm of the posterior probability of the decay rate and the correlation coefficient given the true frequency. The probability enclosed by the highest contour here is only 0.99999.

Conclusions

In NMR a great deal of prior information is available about the time series. When this information is incorporated into the analysis of the data, the frequencies, decay rates, and amplitudes may be estimated several orders of magnitude better than by

direct use of the discrete Fourier transform. Additionally, if the noise is correlated, substantial improvement in the estimation of the amplitudes, frequencies, and decay rates is possible.

Acknowledgments

This work supported by NIH grant GM-30331, J. J. H. Ackerman principal investigator. The encouragement of Dr. J. J. H. Ackerman and Professor E. T. Jaynes is greatly appreciated.

References

[1] Bretthorst G. L., (1987), Bayesian Spectrum Analysis and Parameter Estimation, Ph.D. thesis, Washington University, St. Louis, MO.; available from University Microfilms Inc., Ann Arbor, Mich.

[2] Bretthorst, G. L., (1988), Bayesian Spectrum Analysis and Parameter Estimation, in *Lecture Notes in Statistics,* Vol. 48, Springer-Verlag, New York, New York

[3] Jeffreys, H., (1939), Theory of Probability, Oxford University Press, London, (Later editions, 1948, 1961).

[4] Jaynes, E. T., (1987), "Bayesian Spectrum and Chirp Analysis," in Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems, C. Ray Smith and G. J. Erickson, eds., D. Reidel, Dordrecht-Holland, pp. 1-37.

[5] Jaynes, E. T., (1988), private communications.

SELECTIVE DATA-SAMPLING AND RECONSTRUCTION OF PHASE SENSITIVE 2D NMR SPECTRA USING MAXIMUM ENTROPY

E.D. LAUE
Department of Biochemistry,
Tennis Court Road,
Cambridge, CB2 1QW,
U.K.

ABSTRACT. Whilst one can now determine the three-dimensional structure of small proteins and nucleic acid fragments using two-dimensional (2D) NMR methods, the technique will always suffer from its inherently low sensitivity. This limits the resolution obtainable, in complex 2D NMR spectra of biomolecules. In this paper we review the methods of selective data-sampling and maximum entropy (MEM) data-processing that we have been developing. The results show that they can increase the resolution in 2D NMR spectra. It is hoped that these methods will help extend the use of 2D NMR to cases where otherwise it would be impractical.

1. Introduction

In recent years the development of two-dimensional nuclear magnetic resonance (2D NMR) spectroscopic techniques has enormously increased the potential of the method. It is now possible to determine the structure of small proteins and nucleic acids (M_r <10,000) by 2D NMR methods alone. Currently many improvements are being made, with which it is hoped it will eventually be possible to study larger proteins up to M_r~40,000.

NMR is, however, a very insensitive technique, because the energy levels involved in the absorption at radiofrequencies are so close. Over the years the development of superconducting magnets of increasingly higher field strength (now up to 14.1 Tesla or 600 MHz for protons) has improved the situation, but nevertheless the problem will always remain. Large amounts of sample, by biochemical standards, are required and often proteins are insufficiently soluble or they may be difficult to obtain in sufficient quantity. The attainment of adequate resolution and sensitivity is a major problem and in this paper I will review our attempts to develop methodology for data-sampling and data-processing in 2D NMR, which we hope will help alleviate some of these problems.

2. 2D NMR Spectroscopy

Since its proposal in 1971 by Jeener *(1)*, many 2D NMR methods have been developed, notably by Ernst and his co-workers *(2)*. Structural determination of proteins, pioneered by Wüthrich and his co-workers *(3)*, uses in the main one 2D NMR method. In essence the idea is to determine whether a given two protons are close in space; if they are, that provides a constraint on the structure. When repeated for all possible pairs of protons, the structure of the protein can be determined.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 275–284.

In one-dimensional NMR, the free induction decay or FID (the signal) is recorded with time, following excitation of the nuclei by a radiofrequency pulse. This is then Fourier transformed to give a spectrum. In 2D NMR experiments a whole series of FID's are recorded; for each FID a variable delay is incrementally increased. This series in time is the second dimension (t_1) and after Fourier transforming each FID, a Fourier transform in this second dimension leads to a 2D NMR spectrum. By convention, the first dimension is called t_2. If this were a NOESY spectrum (a particular type of 2D NMR spectrum), the off diagonal or cross peaks would show that a given two protons were close in space (<5A° apart), thus providing a constraint on the structure.

When recording a 2D NMR spectrum, one is usually compelled to use shorter acquisition times than are required for complete decay of the signal in the second dimension (t_1). This 'truncation' improves the signal to noise ratio (S/N) because the signal decays whilst the noise remains constant *(4)*. However the high resolution information is lost. This is very disadvantageous since obtaining high resolution is crucial for the very complex spectra of proteins. Truncation of the signal also leads to artefacts called 'sinc wiggles' in the spectrum following Fourier transformation. These can be avoided by multiplying the FID by a function such as a sinebell; unfortunately this further reduces the resolution *(5)*. We have been developing methods of selective data-sampling and the maximum entropy method (MEM) with a view to avoiding these problems so that we can obtain the required resolution with adequate S/N and no artefacts.

3. Maximum Entropy and 2D NMR

3.1 AVOIDING LOSS OF RESOLUTION AND TRUNCATION ARTEFACTS

The advantage of methods such as MEM stem from the fact that they involve 'finding' (computing) a spectrum to fit the data; this is the reverse of conventional NMR dataprocessing where the FID is filtered and Fourier transformed to give the spectrum. The principles are illustrated in Figure 1. First a well digitised trial spectrum is inverse Fourier transformed to give a mock dataset (FID). The mock FID data can be calculated for an arbitrarily longer time domain than that of the real FID when the trial spectrum is correspondingly well digitised in frequency space. In effect, this predicts unrecorded data, which would be forced to be zero in a conventional zero filled Fourier transform. In this way truncation artefacts are reduced without degrading the resolution. The resulting mock FID data are then checked for consistency with the real FID, though of course only the points corresponding to the measured points can be used. In successive iterations the trial spectrum is modified until its entropy is maximised, subject to the constraint that the corresponding mock FID data agree with the experimental data to within the noise *(6,7)*.

When produced by conventional Fourier transformation, after zero filling four times in each dimension, the 2D NMR spectrum shows characteristic truncation artefacts (not shown). These artefacts were suppressed by pre-multiplication of the data by a sinebell *(5)* in both t_2 and t_1 prior to transformation (Figure 2(a)). Comparison of this spectrum with the one reconstructed using MEM (Figure 2(b)) shows that in the former the resolution is substantially degraded. The expansions of one of the multiplets show this particularly clearly. This result demonstrates that it should be possible to obtain a given resolution in a 2D NMR spectrum with a shorter measuring time when the FID data are reconstructed using MEM.

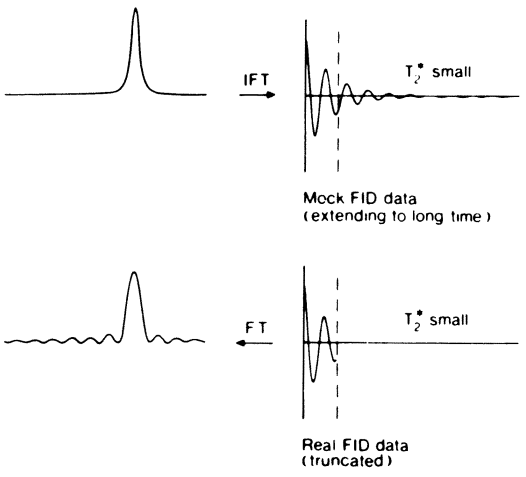


Figure 1 . Flowchart illustrating a MEM calculation; the spectrum to data transform (for a spectrum containing positive peaks only) is shown, as is the result of a conventional zero filled Fourier transform of the deal data. (Reproduced with permission from the *Journal of Magnetic Resonance*, **68,** 14 (1986)).

3.2 AVOIDING ARTEFACTS CAUSED BY BASELINE DISTORTION

If the first few points of an FID are corrupted for any reason, e.g. instrumental imperfections, then low frequency structure such as gentle rolls in the baseline are seen in the spectrum. Conventionally, these may be removed by attempting to fit polynomial or other functions to the shape of the curve and subtracting the function from the spectrum. A much more general approach is provided by MEM in which we can leave out the first few recorded data points when making the comparison with the mock data. This has the effect that MEM now extrapolates back to the time origin from the later uncorrupted data, predicting uncorrupted initial points giving no associated baseline distortions in the final spectrum.

3.3 SUBSPECTRAL EDITING AND PATTERN RECOGNITION

Many 2D NMR spectra contain dispersion peaks in addition to absorption peaks. Often it is not possible to phase such spectra so that all signals are in pure absorption phase. However, because it increases resolution it would be of interest to produce spectra that can be so phased. For spectra that contain positive and/or negative absorption peaks we have both positive and negative trial spectra (f_ω^+ and f_ω^-). We usually display the difference ($f_\omega^+ - f_\omega^-$) (see Figure 2 (b)). This gives a double summation (*8*),

$$S = - \Sigma f_\omega^+ \log f_\omega^+ - \Sigma f_\omega^- \log f_\omega^- \qquad [1]$$

in suitable units.

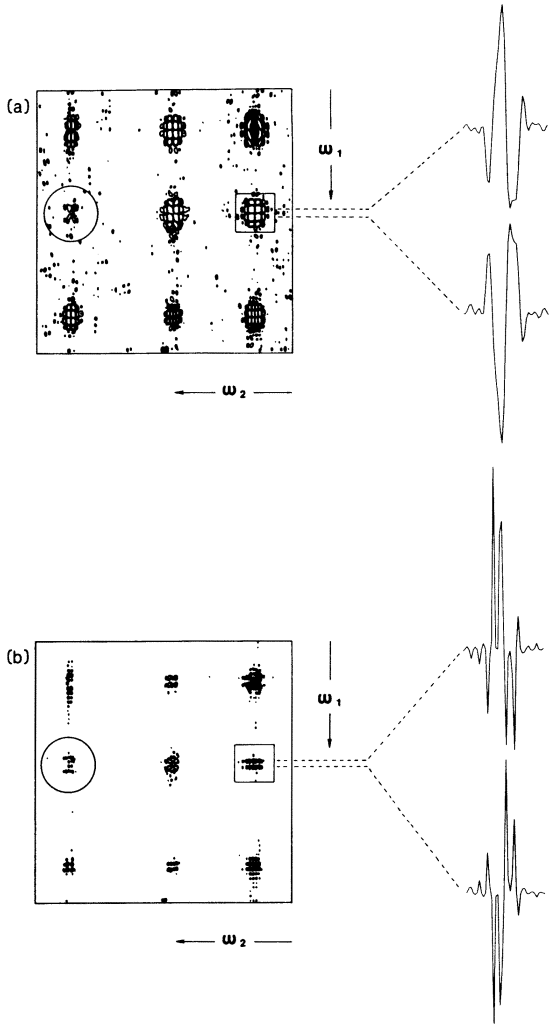


Figure 2. Phase sensitive 2D NMR spectra produced by (a) a conventional Fourier transform, after multiplication by a sinebell in both dimensions, and (b) MEM. All the spectra have equal digital resolution and are contoured at levels of 2% and 5% of the highest peak. Positive and negative levels are plotted without distinction. The expansions are of the multiplet shown. (Reproduced with permission from the *Journal of Magnetic Resonance*, **68**, 14 (1986)).

Prior to inverse Fourier transformation (Figure 1) the two trial spectra f_ω^+ and f_ω^- are combined $f_\omega^+ - f_\omega^-$. With spectra containing both pure absorption and pure dispersion peaks we simple expand this to four separate trial spectra (f_ω^+, f_ω^-, f_ω^i and f_ω^{-i}) and equation [1] becomes,

$$S = - \Sigma f_\omega^+ \log f_\omega^+ - \Sigma f_\omega^- \log f_\omega^- - \Sigma f_\omega^i \log f_\omega^i - \Sigma f_\omega^{-i} \log f_\omega^{-i} \quad [2]$$

Here prior to inverse Fourier transformation (Figure 1) the four trial spectra are combined $((f_{\omega}^{+} - f_{\omega}^{-}) + (f_{\omega}^{i} - f_{\omega}^{-i}))$.

Comparison of the 2D NMR spectra obtained after a conventional Fourier transform (zero filled once in each dimension) (Figure 3(a)) with the absorption difference spectrum $(f_{\omega}^{+} - f_{\omega}^{-})$ and the dispersion difference spectrum $(f_{\omega}^{i} - f_{\omega}^{-i})$ (Figure 3 (b)) obtained after reconstruction using MEM shows that we can considerably enhance resolution. This is because firstly, MEM separates the absorption and dispersion peaks into two subspectra. Secondly, peaks in the dispersion subspectrum now also have pure absorption phase.

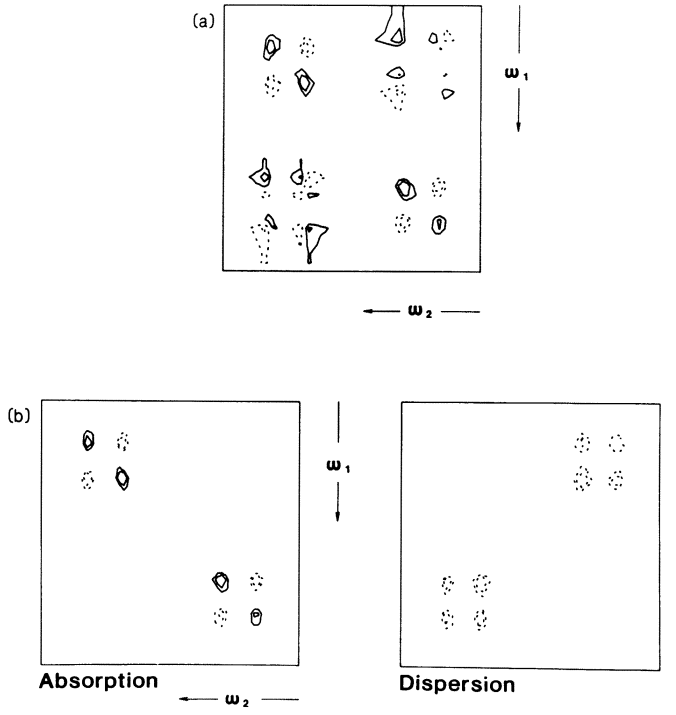


Figure 3. (a) A phase sensitive 2D NMR spectrum produced by a conventional Fourier transform. (b) The absorption and dispersion subspectra produced from the same dataset using MEM. All the spectra are contoured at levels of 20% and 40% of the highest peak. Positive levels are plotted with a solid line whilst negative levels are plotted with a dotted line. (Reproduced with permission from the *Journal of Magnetic Resonance*, **68**, 14 (1986)).

This method is expected to also be useful when applied to the reconstruction and analysis of 2D NMR spectra where the multiplets can have mixed absorption/dispersion lineshapes. For these spectra instead of having a dispersion subspectrum which is 90° out of phase in both dimensions we would reconstruct over subspectra that were 90° out of phase in one dimension only. In more complex situations one can envisage a reconstruction over all the four possible subspectra, (i.e. pure absorption, pure dispersion and those 90° out of phase in either t_1 or t_2). Further details of this work can be found in our published papers *(8,9,10).*

A simple extension of this approach to subspectral editing allows one to recognise patterns in 2D NMR spectra. We have used this approach to recognise a multiplet structure characteristic of a particular type of cross peak in the spectrum *(11)*. The spectrum to data transform used in the MEM calculation now involves a series of pattern channels in which trial spectra are first convolved with a series of patterns of different separation (J_{min} to J_{max}) corresponding to the expected range. These spectra are then combined and added to a background channel prior to inverse Fourier transformation to give the mock data (Figure 4).

Using this approach we have sought to test whether pattern recognition might aid resolution of overlapping multiplets. In Figure 5(a) the spectrum is a conventional Fourier transform of a simulated noisy FID. The spectrum consists of multiplets, with the structure shown, overlapping in various ways. The lower spectrum (Figure 5(b)) is the result of a MEM reconstruction (incorporating pattern recognition), where each negative peak now represents a complete pattern. The resolution of overlapping multiplets is clear, indicating that the incorporation of pattern recognition may well be useful in this context.

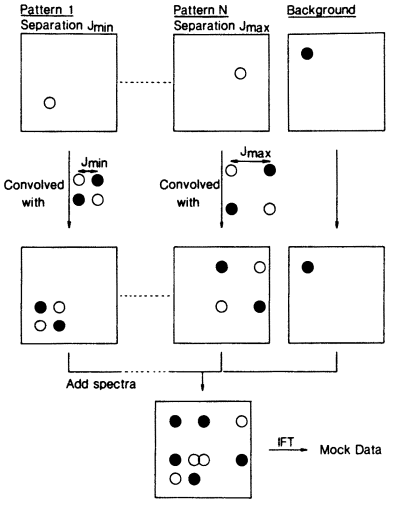


Figure 4. Flowchart showing the spectrum to data transform used for pattern recognition using MEM. In the pattern, the open circles represent negative peaks whilst the filled circles represent positive peaks. A series of pattern channels are used; the number depends on the expected range of separation (J_{min} to J_{max}).

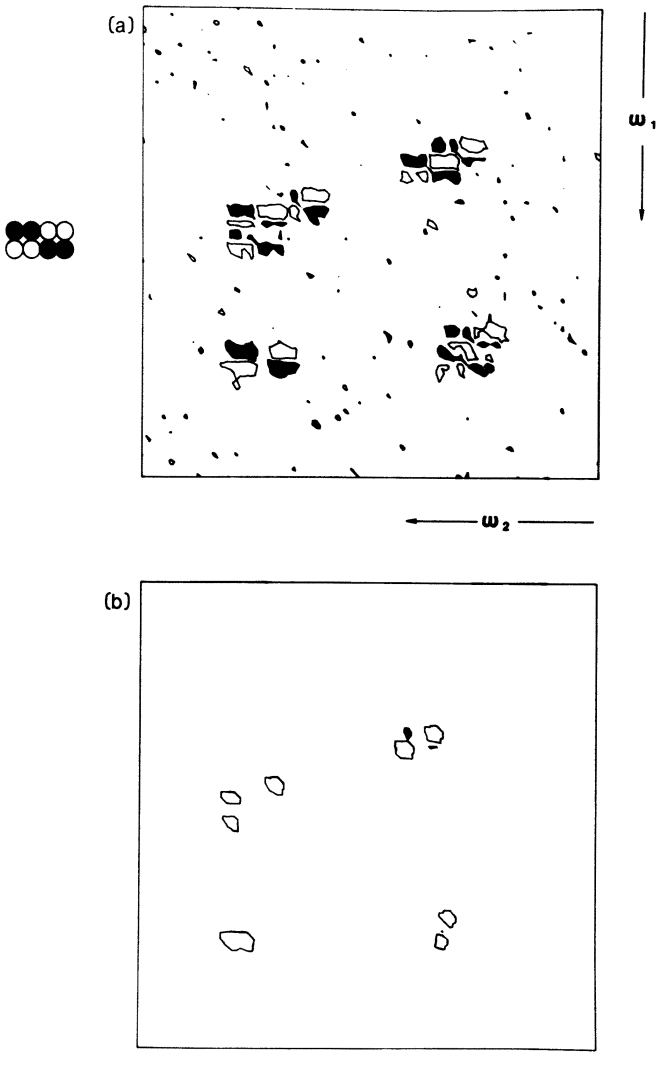


Figure 5. (a) A simulated, noisy, phase sensitive spectrum produced by a conventional Fourier transform. The spectrum consists of the multiplet whose structure is shown in the upper left corner overlapping in various ways. (b) A MEM reconstruction (incorporating pattern recognition) of the same dataset. Open circles/contours represent negative peaks and filled circles/contours represent positive peaks.

4. Selective Data-Sampling

In conventional Fourier transform NMR one always samples the signal at fixed intervals in both t_2 and t_1. One also samples for equal amounts of time at points where the S/N is high and where it is lower. We have seen that MEM can 'reconstruct' some of the lost resolution, effectively by extrapolating from a given data set but there are limits to

this. It seemed that a more logical way to sample an exponentially decaying signal would be to do so in an exponential manner, thus giving a better compromise between S/N and resolution. Many points would be sampled where S/N is high but a few would be sampled where S/N is very low, to aid the MEM reconstruction of high-resoluton information. We have tested our proposal on various one-dimensional spectra which can act as models for the second dimension of 2D NMR experiments.

These tests have demonstrated that the use of selective (exponential) data-sampling allied with MEM data-processing should enable one to increase the available resolution in 2D NMR experiments *(12,13,14)*. This combination gives better results than conventional data-sampling whether allied with either conventional or MEM data-processing methods.

We illustrate here the sort of results one can obtain using a small section of a 2D NMR spectrum. Figure 6(a) shows the result obtained after a conventional Fourier transform of the first 128 peaks in t_1, zero-filled to 1024 points, which represents the case where the data have been truncated in t_1. Multiplication by a sinebell shifted by 45°, was used in t_1 and no filter function was used in t_2. Figure 6(b) shows the MEM result using the same 128 points; it is a slight improvement. Figure 6(c) however, shows the MEM result obtained by sampling 128 points selectively (exponentially) from the first 256 points in t_1. It is dramatically better, as it is very similar to the result obtained by Fourier transformation of 512 points in t_1, shown in Figure 6(d). The last spectrum was obtained using the same filter function as for Figure 6(a). Thus the result in Figure 6(c) is remarkably close to the assumed 'right answer', that is, the conventional Fourier transform of 512 points, Figure 6(d). This represents a quarter of the recording time. Alternatively, in a situation where much less sample was available, exponential sampling could render feasible the recording of useful spectra not otherwise obtainable.

We expect that selective data-sampling should be generally useful in 2D NMR experiments. Exponential sampling is suitable for many 2D NMR experiments but alternative sampling methods have been developed for other experiments where this is necessary *(15)*.

5. Conclusion

We have shown that MEM can be used to increase the resolution in 2D NMR spectra. A combination of selective data-sampling and MEM dataprocessing has been shown to give the best results in general. Where appropriate, however, subspectral editing and pattern recognition methods should offer further improvements. It is hoped that these methods will help extend the use of NMR to cases where otherwise it would be impractical.

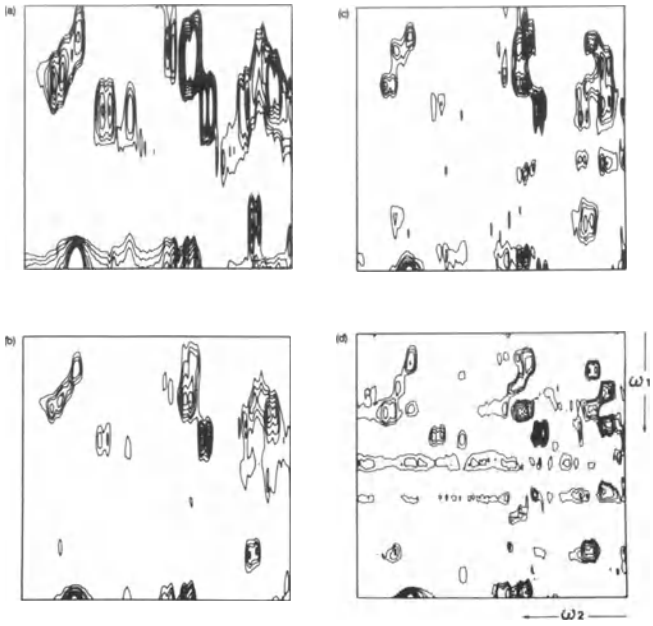


Figure 6. A region of a 2D NMR spectrum produced by (a) a conventional Fourier transform of the first 128 points in t_1 multiplied by a sinebell shifted by 45°, zero-filled to 1024 points; (b) a MEM reconstruction using the same 128 points; (c) a MEM reconstruction using 128 points exponentially sampled out of the first 256 points in t_1; and (d) a conventional Fourier transform of 512 points in t_1, using the same filter function as for (a). All spectra have the same digital resolution and are contoured at the same levels (0.1 to 1.3% of the maximum diagonal peak, not shown here).

6. References

1. J. Jeener, Ampere International Summer School, Basko Polje, Yugoslavia (1971).
2. R.R. Ernst, G. Bodenhausen and A. Wokaun, '*Principles of Nuclear Magnetic Resonance in One and Two Dimensions*', Oxford, (1987).
3. K. Wüthrich, '*NMR of proteins and Nucleic Acids*', Wiley (1986).
4. M.H. Levitt, G. Bodenhausen and R.R. Ernst, *J.Magn.Reson.* **58**, 462 (1984).
5. A. de Marco and K. Wüthrich, *J. Magn.Reson*, **24**, 201 (1976).
6. E.D. Laue, J. Skilling, J. Staunton, S. Sibisi and R.G. Brereton, *J.Magn.Reson.* **62**, 437 (1985).
7. J. Skilling and R.K. Bryan, *Mon.Not.R.Astr.Soc.*, **211**, 111 (1984).
8. E.D. Laue, J. Skilling and J. Staunton, *J.Magn.Reson.* **63**, 418 (1985).
9. E.D. Laue, M.R. Mayger, J. Skilling and J. Staunton, *J.Magn.Reson.* **68**,

14 (1986).
10. E.D. Laue, K.O.B. Pollard, J. Skilling, J. Staunton and A.C. Sutkowski, *J.Magn.Reson.* **72**, 493 (1987).
11. M.R. Mayger and E.D. Laue, unpublished results.
12. J.C.J. Barna, E.D. Laue, M.R. Mayger, J. Skilling and S.J.P. Worrall, *Biochem.Soc.Trans.* **14**, 1262 (1986).
13. J.C.J. Barna, E.D. Laue, M.R. Mayger, J. Skilling and S.J.P. Worrall, *J.Magn.Reson.* **73**, 69 (1987).
14. J.C.J. Barna and E.D. Laue, *J.Magn.Reson.* **75**, 384 (1987).
15. J.C.J. Barna, M.R. Mayger and E.D. Laue, unpublished results.

A NEW MAXIMUM ENTROPY PROCESSING ALGORITHM, WITH APPLICATIONS TO NUCLEAR MAGNETIC RESONANCE EXPERIMENTS.

M.A.Delsuc
Laboratoire de R.M.N. ICSN-CNRS
91190 GIF-SUR-YVETTE France.

Since the introduction of maximum entropy in astronomical images processing(1), this technique has been successfully applied in fields as diverse as astronomy, X-ray cristallography, RAMAN spectroscopy, NMR spectroscopy, medical imaging, etc (1-5). Unfortunately, the difficulty to develop an efficient and simple enough algorithm, seriously hampered the generalization of this very promising technique.

Here we describe a new algorithm, based on a generalization of the Gull & Daniell approach which presents the same ease of implementation associated to a very stable behaviour which permits to handle efficiently a large spectrum of processing conditions. Applications of this algorithm are shown for NMR spectroscopy.

Let consider an object O under study, and the data D resulting from a measurement of this object. Let consider that a linear tranform T can be expressed such as:

$$D_i = \sum_{j=1}^{M} T_{ij}\, O_j + Noise_i \qquad \underline{1}$$

Then the purpose of processing is to reconstruct an image F of the object such that the transform R of this image (R=T(F)) matches as closely as possible the experimental data. The estimator used to evaluate the likelihood of this reconstruction is the classical chi squared statistics C:

$$C = \sum_{i=1}^{N} (D_i - R_i)^2.\sigma_i^{-1} \qquad \underline{2}$$

where σ_i is the standard deviation of the measurement of the i^{th} data point.

The principle of maximum entropy processing is to choose, among all possible images, the one which maximizes the entropy S, being defined as:

$$S = -\sum_{j=1}^{M} P_j \log(P_j) \quad \text{with} \quad P_j = \frac{F_j}{A} \quad \text{and} \quad A = \sum_{j=1}^{M} F_j \qquad \underline{3}$$

Where M is the number of points in the image. Different expressions for the entropy have been used in the literature, the one proposed here has the additional advantage of being insensitive to data scaling.

In order to find the image which maximizes the value of S for a given value of C, a function Q is constructed with the lagrange multiplier λ:

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 285–290.

$$Q = S - \lambda C \qquad 4$$

The derivative of Q is given by:

$$\nabla Q = \nabla S - \lambda \nabla C = 0.$$

with

$$\frac{\delta S}{\delta F_i} = -\frac{1}{A}\left(S + \log(P_i) \right) \quad \text{and} \quad \frac{\delta C}{\delta F_i} = -2 \sum_{i=1}^{N} T_{ij}\,(D_i - R_i).\sigma_i^{-1} \qquad 5$$

At the solution point, $\nabla Q = 0$, which implies:

$$F_j = A \exp\left(-S + \lambda A \sum_{i=1}^{N} T_{ij}\,(D_i - R_i).\sigma_i^{-1} \right) \qquad 6$$

Note that at the solution point, the term within the exponential is always negative, so a linear extention of the exponential can be used for positive values positive values of the argument. This important property comes from the expression of the entropy which has been used here.

Equation 6 can be applied iteratively, starting with any flat image and using the current image F^j (for the j^{th} iteration) in the right hand side of equation 6, the result F'^j in the left hand side used as the next image F^{j+1}. This is the scheme proposed by Gull & Daniell(1). Convergence can be monitored with the value of the angle between the two vectors ∇S and ∇C, these two vectors being antiparallel at the convergence point. However, equation 6 is exact only at the solution point, and is used in this scheme far from this point, Wu(6) proposed a second order correction of the iterative equation which has proved to be very efficient in speeding up the convergence of the algorithm.

The merits of the Gull & Daniell approach are the positiveness of successive iterates and rapid development of large peaks values insured by the exponential function in equation 6. However, this same exponential tends to make the algorithm very unstable. In order to stabilize the iteration process, only a small step is usually taken in the direction F'^j :

$$F^{j+1} = (1-\alpha)\,F^j + \alpha\,F'^j \qquad 7$$

The step α being typicaly as small as 1 to 10 percent to insure convergence (or even lower values if Wu correction is not used).

The main improvement proposed here over the scheme presented above, is to find the optimum α for each new computed image F'^j. In order to do so, the value of Q is maximized over the value of α. This process is a simple one-dimensional maximization process, which has the effect of efficiently stabilizing the convergence, since each iteration insures the improvement of the function Q (this is not guaranted by equations 6 and 7 only).

Another critical point with such an algorithm is to find the optimum λ for which the maximum of Q corresponds to a value C equal to the number of data points N.

What is proposed here is to evaluate the current value for λ such that the current step changes as less as possible the value of the entropy. In other words:

$$\lambda_{opt} \text{ such that } \nabla Q.\nabla S = 0 \quad \lambda_{opt} = \frac{\nabla S^2}{\nabla S.\nabla C} \qquad 8$$

Such a λ_{opt} would lead rapidly the process to the closer convergence point (∇S and ∇C antiparallel, ∇Q is null, but C greater than N), so in order not to stop the convergence toward the point where C=N, the λ actually used λ_{cur} is evaluated from λ_{opt} :

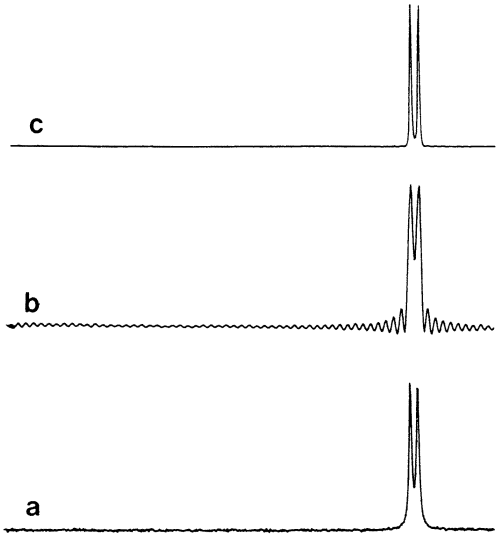


figure 1: *a)* Fourier transform of a simulated NMR spectrum, consisting of two overlapping lines with caracteristic line-width and 40dB signal/noise, 1024 points spectrum. *b)* Fourier transform of the same data-set, using only the first 128 points and padding with zero to 1024 points; *c)* GIFA processing of the first 128 first points of the same data-set, with deconvolution of the line-width, and reconstruction on 1024 points. 10 iterations, 60 transforms were used.

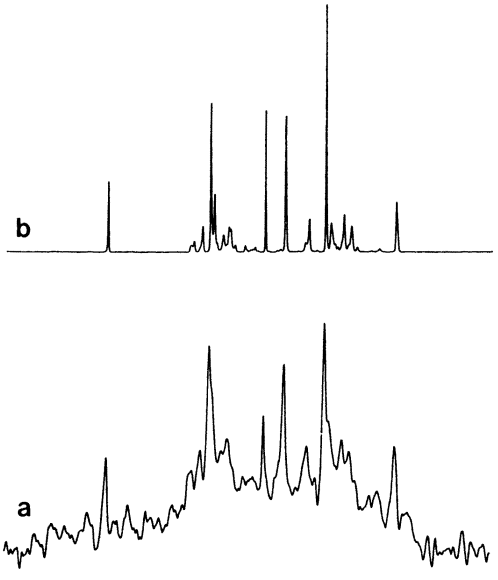


figure 2: The ^{31}P spectrum of an ex-vivo sample from newborn rat brain cells. Experimental data was acquired on a Bruker AM400-WB spectrometer, operating at 163MHz. *a)* classical processing of this data with a 20Hz exponential line-broadening prior to Fourier Transform; note the large base-line distortion due to the phospholipid background. *b)*, result of a GIFA processing applied on the first 1500 points of the same data. A 8Hz exponential line-broadening was applied before processing, and a 20Hz lorentzian deconvolution was used. The very first 4 points were discarded in order to reduce the phospholipid signal. 10 iterations 60 transforms were performed.

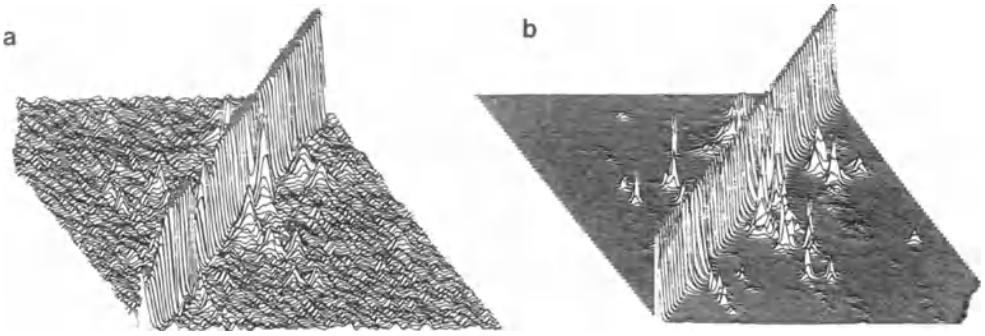


figure 3: Aromatic part of the phase-sensitive 2D HOHAHA spectrum of angioginine protein (123 A.A.), the sample is 3mM in D_2O. This 2D spectrum was acquired on a Bruker WM400 spectrometer operating at 400MHz. on a data matrix 512x1024 with a spectral-width of 4000Hz in both dimensions *a)* classical processing of the data with a cosine filter and zero-filling prior to Fourier transform. *b)* GIFA processing of the same data set. The data set has been Fourier transformed in order to be phase corrected, a 128x256 window corresponding to the aromatic part has been extracted and then inverse Fourier transformed, the result of this pre-processing was then used for image reconstruction. The reconstructed image is 256x256 points. A 5Hz lorentzien deconvolution was used during reconstruction. 40 iterations 240 transforms were performed.

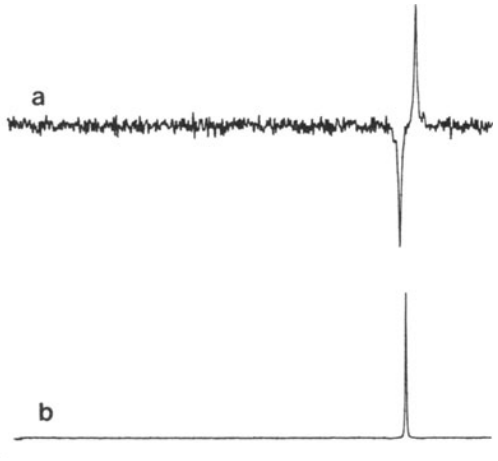


figure 4: *a)* Fourier transform of a simulated anti-phase doublet, with a 20dB signal/noise. *b)* J-Deconvolution of the same data-set, 10 GIFA iterations.

$$\lambda_{cur} = \gamma . \lambda_{opt} \qquad 9$$

γ being typically in the range 3 to 10.

To sumerize the proposed procedure is the following: (i) generate a starting image (usually flat), (ii) evalutate the gradients ∇S and ∇C, and the values of λ_{opt}, λ_{cur} and of the angle between ∇S and ∇C (iii) using equation 6 compute the image F'^j from the current image F^j, (iv) apply if needed the Wu correction on F'^j, (v) with equation 7, find the value of α corresponding to the maximum of Q, and compute the next image F^{j+1}. (vi) loop back to step (ii) as long as C is larger than N or as long as C improves substantially. This procedure has been implemented in FORTRAN 77, on a microVax computer, along with a comprehensive NMR package. This programme has been nicked-named GIFA, standing for General Iterative Fixe-point Algorithm.

This procedure has the advantage of being very robust: the maximization of the function Q for the value α of the step, insures a perfect stability of the process. Computation of the value of λ from the current image permits an optimum driving of the programme. The implemention is straightforward: one only needs a one dimensional maximizing algorithm, and the computation of the transform T and its transpose ^tT. In the present work, the maximization was implemented using a simple parabolic fit (Brent method 7), which needs not any expression of the derivative. Experience has proved that only 4 to 6 iterations for the parabolic fit are usually enough to find a not-so-bad maximium, each iteration needing the computation of one T. This process leads to a total of 6 to 8 evaluations of T or ^tT for each maximum entropy iteration. For the present application: NMR spectroscopy, the transforms T and ^tT used are mainly the Fourier transform which benefits of the very efficient FFT implementation.

As examples, this scheme was used to process NMR spectra. In figure 1 a simulated NMR spectrum with 2 overlaping lines and truncature artifacts is processed. In figure 2 the classical processing of a ^{31}P in-vivo spectrum of living cerebral cells is compared to the maximum entropy processing. The scheme was then applied to the aromatic part of the 2D HOHAHA spectrum of the 123 amino-acids protein Angioginine, homologous to Ribonuclease. The programme was then used to implement the J-Deconvolution technique (8). This technique is used in NMR to reduce the in-phase or anti-phase structures which appears in NMR; this implies deconvolution with the Fourier-transform of sine or cosine functions. Figure 4 shows the anti-phase J-Deconvolution of a simulated NMR spectrum, figure 5 demonstrates the enhancement of resolution obtained with this technique on a 2D-COSY spectrum of a decanucleotide.

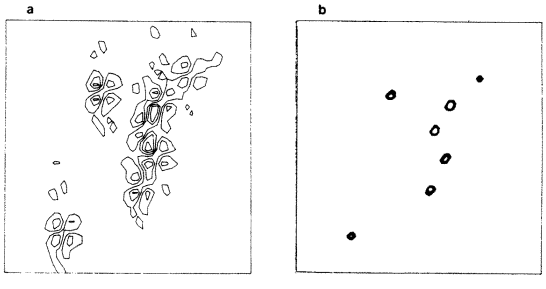


figure 5: *a)* H5-H6 region of a DQF-COSY spectrum performed on the platinated oligonucleotide: (CGCTAGGCCG)-(CGGCCTAGCG) displaying 2D anti-phase patterns; *b)* J-Deconvolution of the same region. Same processing technique as in figure 3, 10 GIFA iterations, 60 transforms.

These examples show that the maximum entropy algorithm proposed here can be successfully used to improve the quality of NMR spectra. More generally, this technique should be very usefull in a wider range of applications.

Acknowledgments:
The author gratefully ackowledges Pf.J.Y.Lallemand, Dr.E.Guittet, V.Stoven, C.Pasquier and M.Robin for their help all along the process of this work.

References:

1) S.F.Gull & G.J.Daniell *Nature* **272** p686 (1978).
2)J.Skilling & R.K.Bryan *Mon.Not.R.Astr.Soc.* **211** p111 (1984).
3) F.Ni & H.A.Scheraga *J.Raman.Spectrosc.* **16** p337 (1985).
4) E.D.Laue, J.Skilling, J.Staunton, S.Sibisi & R.G.Brereton *J.Magn.Reson.* **62** p437 (1985).
5) S.F.Gull & J.Skilling *I.E.E.E. Proceedings* **131F**, p646 (1984).
6) N.L.Wu *Astron.Astrophys.* **139** p 555 (1984).
7) W.H.Press, B.P.Flannery, S.A.Teukolsky & W.T.Vetterling *Numerical Recipes, The Art of Scientific Computing.* (Cambridge Univ. Press) p251-254 (1986).
8) M.A.Delsuc & G.C.Levy J-Deconv *J.Magn.Reson.* 76 p306 (1988).

SAMPLING STRATEGIES FOR MAGNETIC RESONANCE EXPERIMENTS

R. de Beer, D. van Ormondt, W.W.F. Pijnappel,
and J.W.C. van der Veen
Applied Physics Department, Delft University of Technology,
P.O. Box 5046
2600 GA Delft
The Netherlands

ABSTRACT. This contribution is concerned with quantitative analysis of the efficiency of sampling strategies that are intended for resolution enhancement. The criterion for efficiency is based on the supposition that the ultimate aim of a measurement is the quantification of physical model parameters. Thus, the Cramér-Rao lower bounds are introduced as a measure of the efficiency.

1. INTRODUCTION

In most magnetic resonance (MR) experiments, the signal is recorded in the *time* domain, samples being taken at uniformly distributed instants of time t_0, t_1, ..., t_{N-1}. Essentially, a MR signal consists of a number of damped sinusoids (plus white Gaussian noise), the frequencies, damping factors, amplitudes, and phases of which are to be quantified.

In order to carry out the desired quantification, one must have a means to first identify each sinusoid. The latter task, in turn, can be achieved by Fourier transformation of the data to the frequency domain and perusing the resulting spectrum. At this point a problem may be encountered: For economy reasons of some sort, it may have been necessary to limit the *duration* of the measurement. Should such a measure have been effected by prematurely halting the acquisition, while the instants of time of those samples recorded were kept intact, then an unacceptable loss of resolution may be incurred. This would seriously impair the process of identification.

The problem noted in the preceding paragraph has recently been addressed by Barna *et al.* [1]. These authors have pointed out that the loss of resolution attendant on reduction of the duration of a measurement can be significantly alleviated by resorting to MEM (see also [2]). They proposed to restore the time spanned by the reduced number, N', of samples by *exponentially* redistributing the instants of time t'_0, ... ,$t'_{N'-1}$, in such a manner that $t_{N-1}-t_0=t'_{N'-1}-t'_0$, $N'<N$. Fig.1 serves to illustrate the principle of the sampling method, for $N=32$, $N'=8$. For practical reasons, the non-uniformly sampled data are constrained to lie on the original time grid. Also, the notion of the time span, $t_{N-1}-t_0$, of a sampling strategy should not be confused with the actual duration of the associated measurement.

Understandably, Fourier transformation of exponentially sampled data

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 291–296.

yields a severely distorted spectrum, but it is this very aspect that can be effectively remedied by MEM. See [1] for details.

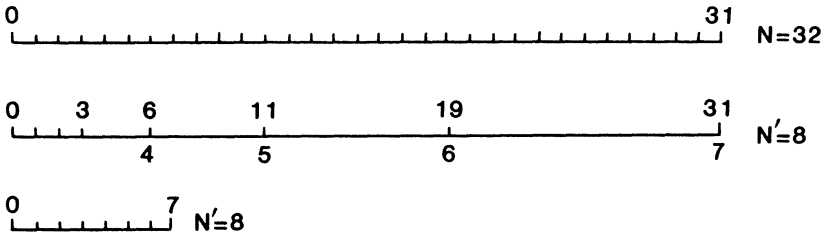


Figure 1. Two alternative ways to reduce the duration of a measurement comprising N=32 samples by a factor of four. Upper: Original uniform sampling scheme, of duration T. Middle: Exponential sampling, such that $t'_7-t'_0=t_{31}-t_0$, and T'=T/4. Bottom: Truncated uniform sampling scheme, i.e., $t'_7-t'_0=(t_{31}-t_0)/4$, and T'=T/4. See also text.

An alternative to reducing the number of the samples and redistributing the remaining ones, as shown in Fig.1, is to reduce the *signal-to-noise ratios* (SNR's) of the samples. Better still, one could devise a *combination* of the two. In summary, there seems to be a variety of choices to comply with the need to contain the duration of a measurement without unnecessary loss of resolution.

In this contribution we address the problem of how to choose between alternative sampling strategies without carrying out trial experiments for each case, so as to save precious time and manpower. We propose to treat the problem from the standpoint of *parameter estimation*, assuming that quantification of the model parameters pertaining to the signal under investigation is the ultimate goal. Our method implies that one evaluate the so-called Cramér-Rao lower bounds on the model parameters [3,4,5] associated with the quantification experiment at hand.

In the sequel, we indicate how to effect the proposed procedure, for the case that the noise that corrupts the signal is white and Gaussian. Subsequently, the procedure is applied to three exponentially damped sinusoids plus white Gaussian noise.

2. EVALUATION OF THE CRAMER-RAO LOWER BOUNDS

The theory and application of the Cramér-Rao (CR) lower bounds is well described in [3,4,5]. Here we confine ourselves to giving a recipe for numerical calculation of the CR bounds. Once numerical results have been obtained, they should be interpreted as follows. Suppose, one has simulated data, consecutively corrupted with, say, 100 realizations of noise. These realizations are all different from each other, but they satisfy one and the same distribution function. Fitting a model function to each of the 100 noise-corrupted versions of the same signal, one obtains 100 noise-corrupted values for each model parameter. If the model function used is correct, the fit procedure efficient, and the number of data points not too small, then the standard deviation of each parameter approaches the CR lower bounds. Thus, the CR lower bounds pertain to *average* results, obtained by repeating an experiment many times. (It should

be realized that it is hazardous to draw conclusions from a single trial.)

The model function of the signal, $\hat{x}_n$, sampled at times t_n, n=0,...,N-1, is

$$\hat{x}_n = \sum_{k=1}^{K} c_k \exp[(\alpha_k + i2\pi\nu_k)t_n] \quad , \quad n=0,...,N-1, \tag{1}$$

where c_k, α_k, ν_k, k=1,...,K, are the amplitudes, damping factors, and frequencies of the K sinusoids. The noise that corrupts the data is assumed to be Gaussian and white, which is reasonable. This assumption enables one to simplify the calculation considerably [3,4,5]. The recipe then amounts to the following. First one forms a matrix, F, comprising the first derivatives of the model function with respect to all 3×K model parameters that have to quantified, for all t_n. In addition, each derivative is divided by the noise standard deviation, σ_n, associated with the real and imaginary parts of sample x_n. Thus, indicating the model parameters by the symbols p_ℓ, ℓ=1,...,L, where L=3×K, we write

$$F = \begin{bmatrix} \frac{1}{\sigma_1}\frac{\partial \hat{x}_1}{\partial p_1} & \frac{1}{\sigma_1}\frac{\partial \hat{x}_1}{\partial p_2} & \cdots & \frac{1}{\sigma_1}\frac{\partial \hat{x}_1}{\partial p_L} \\ \frac{1}{\sigma_2}\frac{\partial \hat{x}_2}{\partial p_1} & \frac{1}{\sigma_2}\frac{\partial \hat{x}_2}{\partial p_2} & \cdots & \frac{1}{\sigma_2}\frac{\partial \hat{x}_2}{\partial p_L} \\ \vdots & \vdots & & \vdots \\ \frac{1}{\sigma_{N-1}}\frac{\partial \hat{x}_{N-1}}{\partial p_1} & \frac{1}{\sigma_{N-1}}\frac{\partial \hat{x}_{N-1}}{\partial p_2} & \cdots & \frac{1}{\sigma_{N-1}}\frac{\partial \hat{x}_{N-1}}{\partial p_L} \end{bmatrix} . \tag{2}$$

The derivatives in F can easily be worked out analytically. Subsequently, the values of all model parameters, noise standard deviations, and sampling times, pertaining to the case at hand are substituted. Finally, the Cramér-Rao lower bounds, σ_{p_ℓ}, follow from taking the square roots of the diagonal elements of the inverse of the real part of the matrix product $F^\dagger F$, i.e.,

$$\sigma_{p_\ell} = \sqrt{\left[\left(\mathrm{Re}(F^\dagger F)\right)^{-1}\right]_{\ell\ell}} \quad , \quad \ell=1,...,L, \tag{3}$$

where † denotes Hermitian conjugation. The interpretation of the numbers obtained from Eq.(3) is given in the first paragraph of this Section.

3. NUMERICAL EXAMPLE

In this Section, we apply the formulae of Sec.2 to a simulated signal comprising three damped sinusoids satisfying the model function of Eq.(1). The values of the nine model parameters are listed in the heading of Table 1. Fig.2 shows the real part of the signal for n=0,...,511. Each of the 512 samples can be

thought of as the result of averaging an equal number of scans, N_{scan}. Thus, the duration of the simulated measurement was proportional to $512 \times N_{scan}$. Fig.3a shows the spectrum, as obtained by FFT of all 512 (complex-valued) data points. The three spectral components can clearly be distinguished from each other. The CR lower bounds for this case are listed in the first three rows of Table 1.

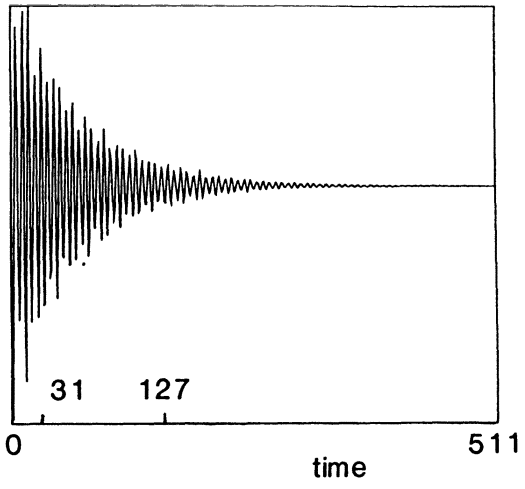


Figure 2. Real part of a time domain signal comprising three sinusoids. The model parameters are given in the heading of Table 1 (noise is omitted here).

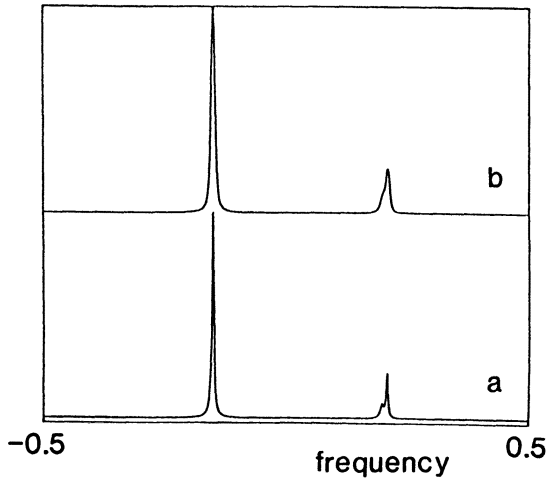


Figure 3. a) FFT of all data points of the simulated signal. b) FFT of the first 128 data points of the signal.

We now turn to the task of reducing the duration of the measurement by a factor of four. In this work four methods are applied:

1) *Uniform* reduction of the number of scans by a factor of four, while keeping the number of samples constant. When one realizes that the noise standard deviation of a sample at $t=t_n$ is proportional to $\sqrt{(1/N_{scan}(t_n))}$, then the CR lower bounds pertaining to this situation can be obtained by multiplying the original ones by $\sqrt{4}$. This applies to all numbers in Table 1.
2) Truncation of the measurement at N'=N/4, while keeping N_{scan} of each sample at its original value. The attendant CR lower bounds are given in rows 4,5,6 of Table 1, for N'=512/4=128, and in rows 10,11,12 for N''=128/4=32.

TABLE 1. Cramér-Rao lower bounds on the standard deviations of the amplitudes, damping factors, and frequencies of three exponentially damped sinusoids, numbered 1,2,3. The amplitudes are 1.0, 0.1, 0.2, the damping factors -0.015, -0.025, -0.015, the frequencies -0.150, 0.200, 0.210; the standard deviation of both the real part and the imaginary part of each sample is 0.01, except in rows 16,17,18.

Sampling Data points	Sampling Time span	Averaging per Sample	Duration of Measurem.	Amplitude	Damping	Frequency	No.
512	511			0.00242	0.00005	0.00001	1
uniform		uniform	T	0.00450	0.00139	0.00014	2
				0.00363	0.00032	0.00003	3
128	127			0.00259	0.00006	0.00001	1
uniform		uniform	T/4	0.00538	0.00163	0.00015	2
				0.00461	0.00046	0.00004	3
128	511			0.00267	0.00007	0.00001	1
exponential		uniform	T/4	0.00656	0.00197	0.00020	2
				0.00572	0.00060	0.00005	3
32	31			0.00386	0.00024	0.00002	1
uniform		uniform	T/16	0.15942	0.01614	0.00817	2
				0.15796	0.01217	0.00305	3
32	128			0.00320	0.00017	0.00002	1
exponential		uniform	T/16	0.01179	0.00399	0.00054	2
				0.01083	0.00128	0.00013	3
32	128			0.00375	0.00015	0.00002	1
exponential		exponential	T/16	0.00986	0.00355	0.00042	2
				0.00868	0.00096	0.00009	3

3) Following Barna *et al.* [1], the number of samples was reduced by a factor of four, and the remaining samples were *redistributed exponentially* on the time axis, in such a manner that $t'_{N/4\ -1}-t'_0=t_{N-1}-t_0$. N_{scan} was not changed, as in 2). The attendant CR lower bounds are given in rows 7,8,9 of Table 1 for

N'=128 and $t'_{127}-t'_0=t_{511}-t_0$, and in rows 13,14,15 for N''=32 and $t''_{31}-t''_0=t'_{127}-t'_0$.
4) The number of samples is reduced by a factor of four, while the instants of time of the samples and the number of scans per sample were *both* distributed exponentially. Thus, the sample times were distributed as in the case of rows 13,14,15, while N_{scan} of sample n' was multiplied by $a\times\exp(0.015\times t'_{n'})$, in which *a* is a normalizing constant that serves to keep the *total* number of scans (i.e. for all $t'_{n'}$ used) unchanged. The attendant CR lower bounds are given in rows 16,17,18 of Table 1.

4. DISCUSSION AND CONCLUSION

Clearly, the smallest CR lower bounds are attained when all 512 data points are available.

If the duration of the measurement is to be cut by a factor of four, then, in the chosen case, it seems best to simply truncate the measurement at n=127, rather than sampling exponentially. Presumably, this is because the signal has already decayed substantially at n=127, so that the SNR of the samples beyond this number is rather low.

When the duration is to be cut by another factor of four, the situation is reversed. Simple truncation at one quarter of the time yields very large CR lower bounds, at least for sinusoids 2 and 3. A dramatic improvement is found for the latter sinusoids, when exponential sampling, devised by Barna *et al.*, is applied. Apparently, it is possible to improve on this further by combining exponential sampling and exponential averaging. Note that the latter result does not apply to the amplitude of sinusoid 1.

In conclusion, the results in Table 1 indicate that the CR lower bounds can provide a criterion for choosing between various alternative sampling strategies devised for resolution enhancement. This finding should enable one to save time and manpower. To arrive at this result it was assumed that quantification of the model parameters (in the time domain) is the ultimate aim of the experiment. In this scheme, MEM will often be indispensable to identify spectral features and to provide starting values of the model parameters.

ACKNOWLEDGEMENT

This work has been supported by the FOM and STW Research Foundations.

REFERENCES

[1] J.C.J. Barna, E.D. Laue, M.R. Mayger, J. Skilling, and S.J.P. Worrall, J. Magn. Reson., **73** (1987) 69-77.
[2] J.C.J. Barna, S.M. Tan, and E.D. Laue, J. Magn. Reson., **78** (1988) 327-332.
[3] M.B. Priestley, 'Spectral Analysis and Time Series, Vols. 1 and 2, Academic Press, London (1981)
[4] A. van den Bos, in 'Handbook of Measurement Science', P.H. Sydenham, Ed., Vol 1, Wiley, London (1982).
[5] J.P. Norton, 'An Introduction to Identification', Academic Press, London (1986).

THE INVERSE PROBLEM FOR NUCLEAR MAGNETIC RESONANCE

G.J. Daniell AND P.J. Hore
Department of Physics Physical Chemistry Laboratory
University of Southampton Oxford University
SOUTHAMPTON SO9 5NH OXFORD OX1 3QZ

ABSTRACT. We discuss the physical interpretation of nuclear magnetic resonance spectra and the relation between the spectra and the data obtained in pulse experiments. A consideration of a classical theory of NMR suggests how the maximum entropy method can be used to obtain spectra containing lines with arbitrary phases. The quantum modifications are given together with some illustrative results.

1. The Inverse Problem

In this paper we address some fundamental questions about data processing in NMR. The experimental situation is that we have a sample in a magnetic field along the z axis, we apply a radio frequency pulse, which sets the nuclei precessing and we measure the time evolution of the magnetic moment.

The first question is "What are we trying to measure?". If we question a random NMR spectroscopist the most probable answer is "The Fourier transform of the free induction decay", but this is unacceptable as a definition of the spectrum. What we want must not be defined by an operation on the data, since these are usually incomplete and contain noise.

In order to answer our question let us calculate the Fourier transform of the data in a perfect experiment. Consider the simplest ideal experiment with the sample in equilibrium, a very short 90° pulse applied, and a perfectly phased spectrometer.

Let $n(\omega)\, d\omega$ be the number of nuclei with chemical shift corresponding to a precession frequency in the range $\omega \to \omega + d\omega$. The pulse rotates the nuclei to produce a magnetic moment which immediately after the pulse is in the x direction, say.

Let $M_x(\omega)\, d\omega$ be the contribution to the initial moment from nuclei that are going to precess with frequency ω. In this simple experiment

$$M_x(\omega) = K\, n(\omega)$$

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 297–302.

with K independent of ω, real and positive.

The signal at time t is then

$$d(t) = \int_{-\infty}^{\infty} M_x(\omega)\, e^{i\omega t}\, d\omega, \qquad t > 0. \tag{1}$$

There has been some confusion over what happens for $t < 0$. From the point of view of inverse theory the answer is that <u>there are no data for $t < 0$</u>. The function d(t) is not zero; it is not defined.

The spectrometer takes the Fourier transform of d(t). Because we only have d(t) for $t > 0$ we have to take the one sided transform

$$\text{'spectrum'} = \int_{0}^{\infty} d(t)\, e^{-i\omega t}\, dt.$$

If we substitute for d(t) from (1) we see that the spectrum is $M_x(\omega)$ convolved with the Fourier transform of a step function, that is

$$\text{'spectrum'} = M_x(\omega) * (\delta(\omega) - i/\omega).$$

Our experiment does not measure $M_x(\omega)$ or $n(\omega)$ but a convolution.

Normally when we measure a convolution we attempt a deconvolution and we might be tempted to reach for the maximum entropy package. However, in the case we have described $M_x(\omega) = K\, n(\omega)$, and $M_x(\omega)$ is real. This is extra information about the experiment and we can use it to construct a trivial deconvolution algorithm, that is: take the real part of the spectrum.

In more complicated experiments some nuclei may, at the end of the pulse sequence, be pointing in the y direction, so we also have $M_y(\omega)$ and in this case

$$\text{'spectrum'} = (M_x(\omega) + i\, M_y(\omega)) * (\delta(\omega) - i/\omega).$$

In these experiments we lose the formula $M_x(\omega) = K\, n(\omega)$ and moreover M_x and M_y may be negative. Now our simple deconvolution rule fails.

We may have additional information that a particular line should have a phase of 0° or 90° and appear in either M_x or M_y. Then we can deconvolve by picking individual lines from the real or imaginary part of the spectrum, but if lines overlap this is impossible. We must then accept that our experiment measures

$$(M_x(\omega) + i\, M_y(\omega)) * (\delta(\omega) - i/\omega)$$

and there is no unique way of getting M_x and M_y from our data.

We can now define what we are trying to measure. In simple experiments it is $M_x(\omega)$ or $n(\omega)$ and $M_x(\omega) > 0$. In other experiments although we would like $M_x(\omega) + iM_y(\omega)$ we have to settle for $(M_x(\omega) + iM_y(\omega)) * (\delta(\omega) - i/\omega)$.

2. Maximum Entropy

We have not mentioned maximum entropy and the analysis in section 1 applies to any data processing method. In practice we are faced with incomplete knowledge of the function d(t), the data may be missing for large and small t and there is noise.

We need a probability distribution to define entropy. A normalised version of $n(\omega)$ is a probability distribution and in the simple experiments where $M_x(\omega) = K\, n(\omega)$ with $K > 0$ a normalised $M_x(\omega)$ is also a probability distribution. We can therefore define

$$S = - \sum_{\omega} M_x(\omega) \log M_x(\omega)$$

and maximise S subject to the Fourier transform of $M_x(\omega)$ agreeing with the data.

Several people have done this and the results are as expected. This approach can be extended to use positivity to phase the spectrometer. An extensive review of previous work is given by Stevenson (1988).

We are concerned in this paper with the more complicated experiments where K is not a real and positive constant. For some experiments we know $K(\omega)$ in principle; we know the pulse sequence we have used; that is additional knowledge and using it we could still define the entropy on $n(\omega)$. Again this is not the problem we are concerned with. We address the problem where we choose not to know $K(\omega)$ and we want to get the spectrum from the free induction decay alone.

In order to have a probability distribution on which to define entropy we will describe a classical theory of NMR - magnets attached to gyroscopes. Let $n(\omega,\theta,\phi)\, d\omega\, d\theta\, d\phi$ be the number of nuclei, precessing with frequency between ω and $\omega + d\omega$, at $t = 0$ (after the pulse sequence) oriented between θ and $\theta + d\theta$ and ϕ and $\phi + d\phi$, where θ and ϕ are spherical polar coordinates.

We can define an entropy on this distribution:

$$S = - \sum_{\omega\theta\phi} n(\omega,\theta,\phi) \log \frac{n(\omega,\theta,\phi)}{\sin\theta} \,. \tag{2}$$

The $\sin\theta$ factor is needed to make the unconstrained maximum of S isotropic. The magnetic moment is

$$M_x(\omega) + iM_y(\omega) = \sum_{\theta,\phi} n(\omega,\theta,\phi) \sin\theta\, e^{i\phi} \tag{3}$$

and $d(t)$ is the Fourier transform of $M_x(\omega) + i\, M_y(\omega)$.

This is a standard linear maximum entropy problem for the distribution $n(\omega,\theta,\phi)$ and could be solved in the standard way. The only difficulty is that it is in three dimensions with say 1000 points in ω and at least 10 in θ and ϕ giving 10^5 variables. Fortunately there is a much more efficient way of solving the problem.

The data depend only on $M_x + iM_y$ so we can maximise S partially, over θ and ϕ, fixing $M_x + iM_y$ for each value of ω. The result is

$$n(\omega,\theta,\phi) \propto \sin\theta\, e^{(\alpha(\omega)\cos\phi + \beta(\omega)\sin\phi)\cos\theta}$$

where α and β are Lagrange multipliers.

We have determined the θ,ϕ variation of n analytically by maximum entropy, only the ω variation needs to be determined numerically.

If we substitute this form of n into the equations (2) and (3) for S and $M_x + iM_y$ we get equations of the form

$$S = f(\alpha^2 + \beta^2)$$

and

$$(M_x^2 + M_y^2)^{\frac{1}{2}} = g(\alpha^2 + \beta^2)$$

where the functions f and g are rather complicated. We can eliminate $\alpha^2 + \beta^2$ and write

$$S = -\, h(\, (M_x^2 + M_y^2)^{\frac{1}{2}})$$

and

$$d(t) = \sum_{\omega} [M_x(\omega) + iM_y(\omega)]\, e^{i\omega t} .$$

This is now a non standard maximum entropy problem because of the function h; but h is a convex function of both M_x and M_y and so the standard algorithms can be modified.

For the classical theory we have described

$$h(x) = x\, \mathrm{arci}_1(x) - i_0(\mathrm{arci}_1(x))$$

where i_0 and i_1 are the modified spherical Bessel functions and we have used the convenient notation of $\mathrm{arci}_1(x)$ for the inverse function of $i_1(x)$.

For the spin-½ quantum theory the corresponding result is

$$h(x) = x \sinh^{-1}(x) - \sqrt{1 + x^2} .$$

3. Discussion

We need to note two related features of any variational method. Because the data are used only through χ^2 we can add to the spectrum any function whose Fourier transform vanishes for $t > 0$. Maximum entropy will do this if the entropy is thereby raised and the Fourier transform of a maximum entropy spectrum will not be zero for $t < 0$. Another way of stating the same thing is to say that the maximum entropy algorithm attempts to do the deconvolution by $(\delta(\omega) - i/\omega)$.

Bearing these points in mind we can ask what should we display as a maximum entropy spectrum. Of the possible candidates $n(\omega,\theta,\phi)$ is difficult to visualise since it is a function of three variables and $n(\omega)$ omits useful phase information in the experiments we are discussing. $M_x(\omega) + iM_y(\omega)$ is probably the 'best' answer but the shapes of the lines in the spectrum are distorted in comparison with the lines in the Fourier transform spectrum. This is because the corresponding time domain function does not vanish for $t < 0$. An alternative is to accept that, as explained in section 1, we really want $(M_x(\omega) + iM_y(\omega))*(\delta(\omega) - i/\omega))$ and to calculate this quantity by performing the convolution after computing the maximum entropy estimate of $M_x(\omega) + iM_y(\omega)$. This loses a little information but the spectrum now looks familiar and the interpretation is easy.

The figure shows that the method we have described produces a spectrum similar to the Fourier transform but with the expected advantages of the maximum entropy method. The left hand half of each trace is the real part and the right hand the imaginary part. The molecule is the trisaccharide N-acetyl glucosamine-galactose-glucose in a 90° acquire experiment and the large peak in the middle of the spectrum is due to water. The upper trace was computed by Fourier transformation after zero filling in the usual way. For the lower trace $M_x(\omega) + iM_y(\omega)$ was computed by the maximum entropy algorithm we describe and this was then transformed to the time domain, the part with $t < 0$ replaced by zero and the frequency domain spectrum recomputed. The features to note are the lower noise and reduced truncation artifacts.

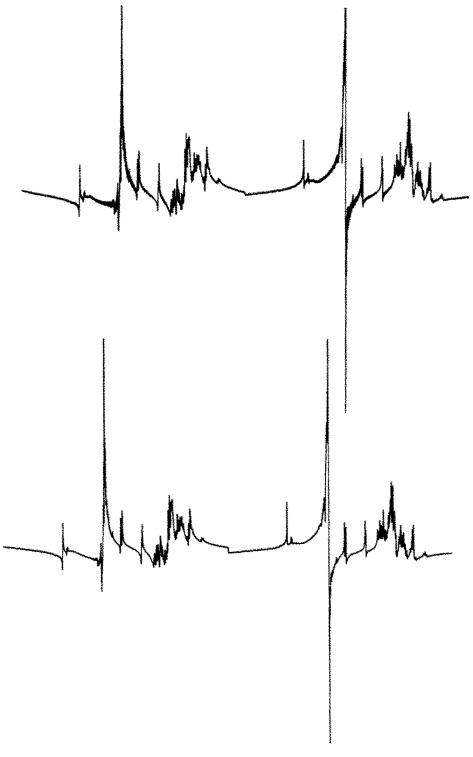
4. Summary

We are concerned in this paper only with the more complicated type of experiments where $M_x(\omega)$ is not essentially the same as $n(\omega)$ and we are assuming we do not build into the analysis the knowledge of the pulse sequence.

By using classical theory there is no doubt thatthe maximum entropy method is correctly applied to determine a probability distribution.

The method produces a complex spectrum like the Fourier transform but we cannot use the method to adjust the phase automatically since automatic methods have to rely on positivity of the spectrum. In the case where all the lines in the spectrum have the same phase then this

additional information means that we should apply the entropy to $n(\omega)$, not to $n(\omega,\theta,\phi)$. There are however many types of NMR experiment (for example relaxation time measurements and coherence transfer experiments) for which our approach should be a new useful method.



Proton NMR spectra of N-acetyl glucosamine-galactose-glucose.
Top: Fourier transform spectrum. Bottom: Maximum Entropy spectrum.

Reference

Stevenson D.S. (1988), 'Linear Prediction and Maximum Entropy Methods in NMR Spectroscopy', Progress in Nuclear Magnetic Resonance Spectroscopy **20** 515.

MAXIMUM ENTROPY CALCULATIONS ON A DISCRETE PROBABILITY SPACE: PREDICTIONS CONFIRMED

PAUL F. FOUGERE
Fern Consultants
461 Old Billerica Road
Bedford, MA 01730
U.S.A.

ABSTRACT. Measurements on the shape of Wolf's dice have recently become available. These measurements compare favorably with Maximal Entropy (ME) predictions made previously. These measurements also yield new constraints which are then used to modify the previously written ME program. The new shape constraints are very nearly as effective as the previously used oblateness constraints, derived from the observed frequencies.

1. Introduction

In the paper "Maximum Entropy Calculations on a Discrete Probability Space" (Fougere,1988) hereinafter called DPS, I analyzed, using the maximum entropy (ME) prescription of Jaynes (1957,1963,1968,1978,1979, 1982), a unique set of experiments performed by Wolf(see Czuber,1908) about 100 years ago.These experiments involved throwing a pair of ordinary playing dice,one white (the" Weisser Würfel"),and one red (the "Roter Würfel") a total of 20,000 times and recording the total number of times that each of the 36 possible combinations (w1,r1),(w2,r1)..... (w6,r6) appeared. Since no correlations were either expected or observed between the white and red results, the red and white marginals were calculated and are repeated for convenience in Table I.

Table I. Marginals and relative frequencies for Wolf's dice.

	White Die		Red Die	
i	Marginal	Frequency	Marginal	Frequency
1	3246	.16230	3407	.17035
2	3449	.17245	3631	.18155
3	2897	.14485	3176	.15880
4	2841	.14205	2916	.14580
5	3635	.18175	3448	.17240
6	3932	.19660	3422	.17110
SUM	20,000	1.00000	20,000	1.00000

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 303–308.

Inspection of these red and white marginals quickly revealed that neither die was "fair". The expected marginals for a fair die are all equal to 20000/6=3333+. Table I shows, for example that w4 occurs 2841 times, 492 less than expected and w6 occurs 3932 times, 599 times more than expected. Furthermore, these deviations from fairness are far greater than expected for normal sampling deviations when casting a fair die. We expect deviations from fairness to be of order square root of 20000/6 ~ 58. Thus the observed deviations are around ten times the expected deviations.

The most likely physical imperfections to be expected in the construction of a real die were then discussed in detail. For convenience we repeat from DPS that the three constraints for the white die were :

1. Spot excavations. Since the higher numbered spots have more material removed, these faces would be "lighter" and more likely to be "up".

2. Prolateness. The 3-4 axis was longer than the other two, approximately equal axes. Thus spots 3 and 4 would be less likely to occur than the others.

3. Corner chip. If the the 2-3-6 corner were chipped off, faces 2 ,3 and 6 would be more likely than the other three.

For the red die, the first two constraints were exactly the same but the third constraint was a little different. Its effect was to make 1 and 4 less likely, 2 and 3 more likely and 5 and 6 were unaffected.

The following predictions were then made:

" We can now see quite clearly, that the white die must have been prolate with the 3-4 dimension being slightly greater than the 1-6 and 2-5 dimensions! "

" The red die is also prolate in exactly the same way as the red die! "

2. The New Observations

Shortly after the camera-ready copy for DPS was sent to the publisher I received a letter from Prof. Jaynes in which he revealed some measurements which had been made by Wolf himself and which were reported to Prof. Jaynes during a visit he made to Wolf's observatory in Zürich. Each measurement was the average of ten micrometer readings of the length of the three sets of axes: the 1-6, 2-5, and 3-4 axes.

Table II quotes these measurements which are reported here for the first time with the express written permission of Professor Jaynes.

Table II. Measurements by Wolf as reported by Jaynes.(millimeters)

Axis	1-6	2-5	3-4
White Die	16.004	16.129	16.402
Red Die	16.303	16.288	16.621

Each measurement is the average of ten micrometer readings.

The measurements in Table II verify both predictions quite well. For the white die, the 3-4 dimension (16.402) is indeed larger than the other two dimensions whose average is 16.067. Similarly for the red die , the 3-4 dimension (16.621) is larger than the the other two dimensions whose average is 16.297.

Of course the dimensional information was not available when DPS was being written, but now that it is available we should be able to use this information as a constraint ,in place of or in addition to the oblateness constraint used in DPS.

Using the dimensions in Table II we can easily find the area of each pair of opposite faces : 1-6, 2-5, and 3-4. For example, the length of the 1-6 axis multiplied by the length of the 2-5 axis gives the area of faces 3 and 4.

The constraints derived in Table III are small integers approximately proportional to Delta, the difference between the area of a face and the average area of all three face (pairs).

Table III. Face areas of the dice; Delta,the difference between an area and the mean area and CON, the approximate constraint.

	White Die			Red Die		
Face	Area	Delta	CON	Area	Delta	CON
1-6	26.455	0.282	7	27.072	0.164	3
2-5	26.250	0.077	2	27.097	0.189	4
3-4	25.813	-0.360	-9	26.554	-0.354	-7
Mean	26.173			26.908		

The computer program discussed at length in DPS was accordingly modified to add a fourth constraint from table III. As before, the constraints can be turned on and off quite simply.

Results for the white die appear in Table IV. The column labelled "significant?" indicates with the symbol">" that the particular set of constraints acting was not sufficient to explain the observed die

frequencies. All sets after the first four are in this category. The first two rows, with all four constraints on , and with the first three on respectively, show very tiny values of chi-squared , 0.11 and 0.39 respectively. With either of these two constraint sets there are definitely no important constraints omitted. The next row shows chi-squared of 8.94 on two degrees of freedom(df). This is significant at the 98% level. That is the constraints numbered 1,2 and 4 are not sufficient to explain the observed frequencies. Similarly the fourth row shows a chi-squared of 9.37 on 3 df, significant at the 97.5% level. The first two constraints by themselves do not quite fully account for the observations.

Table IV. Maxent results on white die. Variation of chi squared (CS) as the constraints are turned on and off. Listed in order of increasing chi squared. 0=constraint off,1=on. ">" under significant indicates very highly significant, at higher than 99.5% level. "df" is number of degrees of freedom.

Serial	I1	I2	I3	I4	CS	df	SIG?
1	1	1	1	1	0.11	1	
2	1	1	1	0	0.39	2	
3	1	1	0	1	8.94	2	~98%
4	1	1	0	0	9.37	3	~98%
5	1	0	1	1	12.88	2	>
6	1	0	0	1	20.86	3	>
7	0	1	1	1	55.58	2	>
8	0	1	1	0	56.28	3	>
9	0	0	1	1	66.91	3	>
10	0	1	0	1	70.99	3	>
11	0	1	0	0	72.01	4	>
12	0	0	0	1	81.12	4	>
13	1	0	1	0	89.77	3	>
14	1	0	0	0	199.42	4	>
15	0	0	1	0	253.85	4	>
16	0	0	0	0	270.90	5	>

From Table IV we can easily read off that the most important constraint was number 2 : oblateness, but close behind was number 4 : the observed shape of the die. Similarly, the most important pair of constraints was numbers 1 and 2, with 1 and 4 close behind. The constraint labelled number 3, the tiny 2-3-6 corner chip was the least important single constraint, but after no 1 and 2 had been taken into account, number 3 was necessary to match the observed frequencies.

Table V lists similar results for the red die. From Table V we can read off that for the red die, the most important constraint was number 4, the observed shape of the die, but close behind it was number 2, the oblateness. The most important pair was 2 and 3, with 2 and 4 close

indeed. In fact either pair (2,3) or pair (3,4) would be completely adequate to explain the observed frequencies.

Table V. Same as Table IV ,but for red die.

Serial	I1	I2	I3	I4	CS	df	SIG?
1	1	1	1	1	0.07	1	
2	1	1	1	0	0.08	2	
3	1	0	1	1	0.61	2	
4	0	1	1	1	2.39	2	
5	0	1	1	0	2.40	2	
6	0	0	1	1	2.91	3	
7	1	1	0	1	13.62	2	>
8	1	0	0	1	15.52	3	>
9	0	1	0	1	15.86	3	>
10	0	0	0	1	17.81	4	>
11	1	1	0	0	18.13	3	>
12	0	1	0	0	20.44	4	>
13	1	0	1	0	74.86	3	>
14	0	0	1	0	77.16	4	>
15	1	0	0	0	91.90	4	>
16	0	0	0	0	94.19	5	>

3. Summary and Conclusions

A. The major predictions made before the shape information became available were nicely borne out by the measurements on the actual dice. Both dice were very close to oblate in the manner indicated in the predictions: the 3-4 dimension was greater than 1-6 and 2-5 dimensions, and this was true for both dice.

B. The smaller constraints : number 3 for both dice, are required to adequately explain the observed frequencies. It is to be hoped that the actual dice used by Wolf over 100 years ago, and believed to be still in existence, can be made available to the scientific community. In that event all of the predictions of Maximum Entropy could be compared with reality.

References

Czuber, R., 'Wahrscheinlichkeitsrechnung',1908.

Fougere,P.F. 'Maximum Entropy Calculations on a Discrete Probability Space',in Maximum Entropy and Bayesian Methods in Science and Engineering,(vol.1) G.J. Erickson and C.R.Smith, editors,Kluwer publishing Co.,1988.

Jaynes,E.T., 'Information Theory and Statistical Mechanics,Part I' ,Phys.Rev.,106,620;'Part II';ibid,108,171, 1957.

Jaynes,E.T., 'Brandeis Lectures' in E.T. Jaynes Papers on Probability, Statistics and Statistical Physics, R.D.Rosenkrantz, Editor, D. Reidel Publishing Co., Boston Mass., 1963.

Jaynes,E.T., 'Prior Probablities',IEEE Trans. Syst. Sci.Cybern.,SSc4, 227,1968.

Jaynes, E.T. 'Where Do We Stand on Maximum Entropy?' in the Maximum Entropy Formalism, R.D. Levine and M. Tribus, Editors, MIT Press, Cambridge,Mass. 1978.

Jaynes, E.T. 'Concentration of Distributions at Entropy Maxima,in E.T. Jaynes' Papers on Probability, Statistics and Statisistical Physics, R.D. Rosenkrantz, Editor,D. Reidel Publishing Co.,Boston Mass.,1979.

Jaynes, E.T. 'On the Rationale of Maximum Entropy Methods',Proc. IEEE 70,939,1982.

APPLICATION OF CLASSICAL, BAYESIAN AND MAXIMUM ENTROPY SPECTRUM ANALYSIS TO NONSTATIONARY TIME SERIES DATA

JUANA SANCHEZ
Washington University
Economics Department
St. Louis, MO.63130

ABSTRACT. This paper contains some preliminary results from an analysis of the sensitivity of Classical, Bayesian and Maximum Entropy Spectrum analysis to detrending procedures. The findings suggest that their performance in discovering periodicities in nonstationary series is affected by the assumption made about the trend. A combination of the three methods when trends are not simple functions of time is desirable and necessary to obtain precise information about the periodic behavior of the data analyzed.

1. Introduction

The most commonly used methods of time series analysis are based on the assumption of stationarity. Many economic time series are nonstationary. Thus in order to apply those methods the nature of the nonstationarities must be discerned and the data transformed accordingly. Recent work in the analysis of economic time series has focused on the effect that wrong assumptions about trends have on the conclusions extracted from data treated with stationarity-dependent methods. Stock & Watson (1988), Serietis (1988) and Nelson & Kang (1981) among others have concluded that the effect can be devastating. Others (e.g. Cochrane 1987) show that for some purposes making the wrong assumption might be even desirable. We analyze in this paper which of these views applies to two methods of analysis of time series in the frequency domain that have not been used extensively with economic time series but could be good alternatives to the well known Classical Spectrum Analysis (CSA) method. These are the Autoregressive or Maximum Entropy Spectrum Analysis (MESA) and the Bayesian Frequency Parameter Estimation (BFPE) methods. The data we use, as far as the results presented here are concerned, are computer generated nonstationary time series. The reason for doing a preliminary analysis with artificial data is very simple: if a method performs

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 309–319.

poorly with data about which we know everything we can use that as a warning against applying that method to real data. If the method does well in some cases and not in other then at least we know what to expect when we apply it.

We have also analyzed the CSA method just for the sake of comparing it with the other two. Our results with it confirm those of Nelson and Kang (1981).

The remainder of the paper is organized as follows: Section 2 briefly reviews the MESA, BFPE and CSA methods and their performance with stationary data. Section 3 presents our preliminary results concerning the sensitivity of the methods to detrending. Conclusions and suggestions for further research are in section 4.

I thank Professor Edward T. Jaynes and Larry Bretthorst for their helpful suggestions and comments.

2. The Methods

We present in this section a brief review of the CSA, MESA and BFPE methods. For extensive detail on them see Jenkins & Watts (1968), Marple (1987) and Bretthorst (1987).

MESA and BFPE are model based, i.e. to apply them a model for the data is initially assumed. This is the only aspect they have in common. On other respects they estimate two very different things.

MESA gives its best estimate of the *spectrum* of the data. The analytic expression for this spectrum is

$$\hat{P}(f) = \frac{T\hat{\rho}_w}{1 + \sum_{n=1}^{p} a_n exp(-i2\pi f n T)} \tag{1}$$

where $\hat{P}(f)$ is the estimated spectral density for each frequency 'f', $\hat{\rho}_w$ is the estimated driving noise variance, and $\hat{a}_n$ are the parameters of the Autoregressive model assumed to represent the data and estimated by means of Burg's algorithm (Burg 1975). To apply it the time series must be stationary and the lag length of the model must be assumed or selected according to some statistical criteria. In any case, what we estimate is the spectrum of both the signal and the noise present in the data.

The BFPE method calculates its best estimate of the *parameters* present in the signal —such as frequencies, decay rates, chirp rates and trends — and also the accuracy of that estimate. It also estimates the noise level of the data. The analytic expression of what it calculates is found by applying the following relation:

$$P(\hat{\theta} \mid DI) = \frac{P(\hat{\theta} \mid I)P(D \mid \hat{\theta}I)}{P(D \mid I)} \tag{2}$$

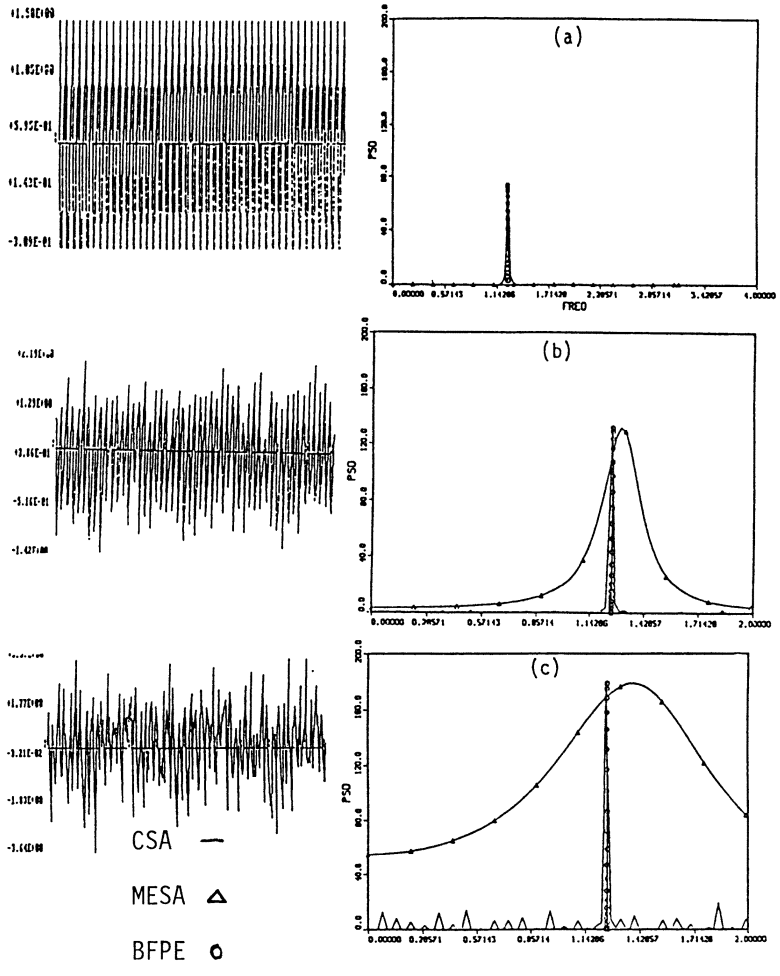


Figure I. MESA, BFPE and CSA of a Stationary Signal with Frequency w = 1.25664 and decreasing signal to noise ratios

where $P(\hat{\theta} \mid DI)$ is the joint posterior distribution of the parameters θ present in a hypothesized model and whose more accurate analytic expression will depend on this model. If some of these θ are nuisance parameters and we are only interested in estimating the set of frequency parameters —f— contained in θ we can integrate out those nuisance parameters and obtain

$$P(\hat{f} \mid \alpha DI) = \int P(\hat{\theta} \mid DI) d\alpha \tag{3}$$

that is, the marginal posterior distribution for the frequencies given the rest of the

parameters —which we call α, the data and the prior information about θ.

The distinctive characteristic of this method with respect to MESA is that it allows us to incorporate any prior information we might have about the data into the estimation process. It does not require stationarity and consequently allows us to consider a wider scope of models to represent the data. The model selection is done according to bayesian probability theory.

CSA estimates the *spectrum* of the data by Fourier transforming it as follows:

$$\hat{P}(f) = \frac{1}{N} \mid \sum_{n=1}^{N} d_n exp(i2\pi f n) \mid^2 \tag{4}$$

where d_n are the data values for $n=1,2,.....,N$

To apply a Classical Fourier Transform to the data the latter must be stationary. No model for the data must be assumed. The method has been extensively compared with MESA (see Childers 1980). As in MESA we estimate the spectrum of both signal and noise together.

2.1 APPLICATION TO STATIONARY DATA

Although stationary data are not the main concern of this paper, we have nevertheless to face it when we have to transform nonstationary series to apply MESA and CSA. It is then convenient to point out an important result that has been documented elsewhere concerning MESA (see Childers 1978; Marple 1987) and that we illustrate with the example in Figure I. Figure I shows the estimates obtained applying the three methods to data generated by the following model:

$$y(t) = cos(1.25664(t-1) + GWN \tag{5}$$

where GWN stands for Gaussian White Noise. When the noise level is nonexistent —plot (a)— the three methods give us the right frequency. As the level of noise increases, —plots (b) to (c)— the MESA spectrum starts to lose resolution while the other two estimates do not.

The bad resolution of MESA as the noise increases is explained by the lag-length assumed. While a lag order of two is good enough to give a high resolution to the estimated espectrum when no noise is present —the peak occurs at w=1.26 in Figure I (a)— that order is no longer appropriate when the noise increases (Figures I (b) and (c)). In Figure I (c), for example, the peak of the MESA spectrum occurs at w=1.398 for lag-length two. We found that it recovers its resolution with a lag-length sixteen.

MESA then is very sensitive to the level of noise in the data and to the lag-length assumed for the autoregressive model representing the process. The method does not provide a model selection criterion and has to rely on ad-hoc criteria to

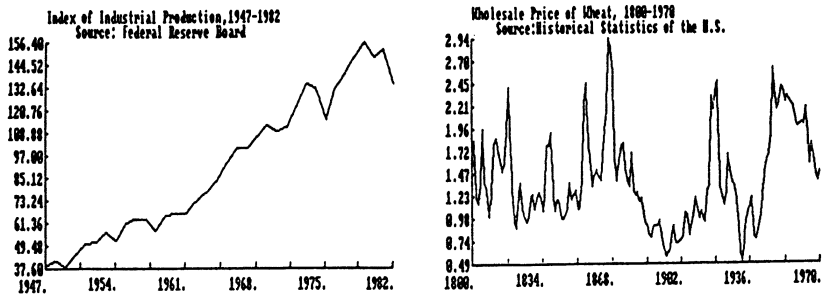


Figure II. Examples of Economic Time Series

decide the lag-length. This must be kept in mind when we add in the next section the problems derived from the presence of trends.

On the other hand the BFPE method provides a model selection procedure based on the same criterion as that used for the estimation of the frequencies, i.e. Bayes theorem. In the example illustrated in Figure I and other that we analyzed this criterion chose the correct model in the majority of cases. When the noise dominated the signal almost completely the posterior probability density function contained spurious peaks that distorted the results.

3. Trends and Cycles

Most economic time series are nonstationary in the sense that they tend to depart from any given value as time goes on. Figure II shows two examples of such series.

Failure to account for nonstationarities has far-reaching consequences in the results obtained from applying to these series methods that assume stationarity. Thus, in Figure III we show what will happen if we apply MESA and Classical Spectrum Analysis to a nonstationary series generated by the following model:

$$y(t) = 1 + 0.3t + cos(0.6283(t-1)) + WN \tag{6}$$

that is, a model with a very simple linear trend and a periodic component of frequency $w = 0.6283$ plus White Gaussian Noise of zero mean and standard deviation equal to 0.9. As it is expected, the results are misleading both for the Classical

(Figure III (b)) and the MESA (Figure III (c)) estimates. The two spectra concentrate all the power at a low frequency. This is a consequence of applying these two methods —which require stationarity— to nonstationary data.

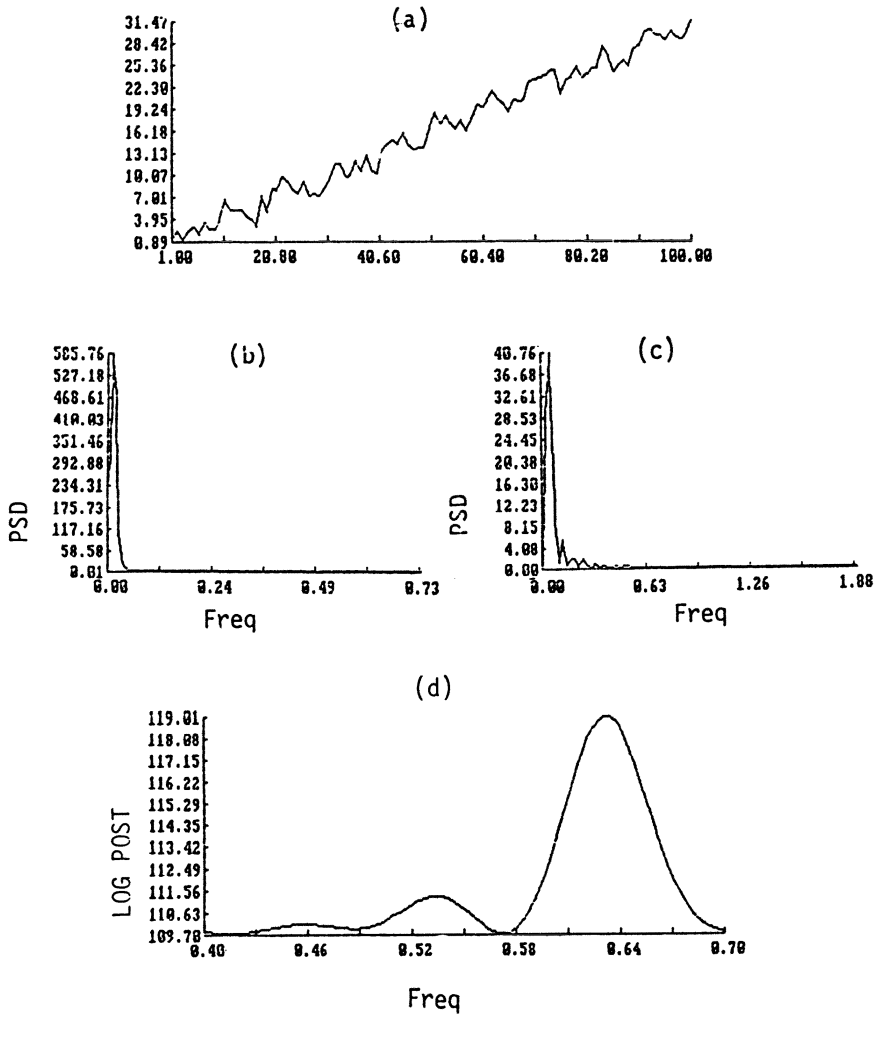


Figure III. MESA, Classical and Bayesian spectra of trend-stationary data

Contrast this result with that obtained from applying the BFPE method to the same data (Figure III (d)). The posterior probability for the frequencies was calculated following the same steps described in Bretthorst (1987) to approach this problem. That is, after seeing the data plotted we considered that an appropriate

model function would be a single frequency plus a trend. Probability theory then prescribed that the removal of a first degree polynomial trend was enough and estimated the right frequency with almost complete accuracy. Removal of higher order trends would not change that result.

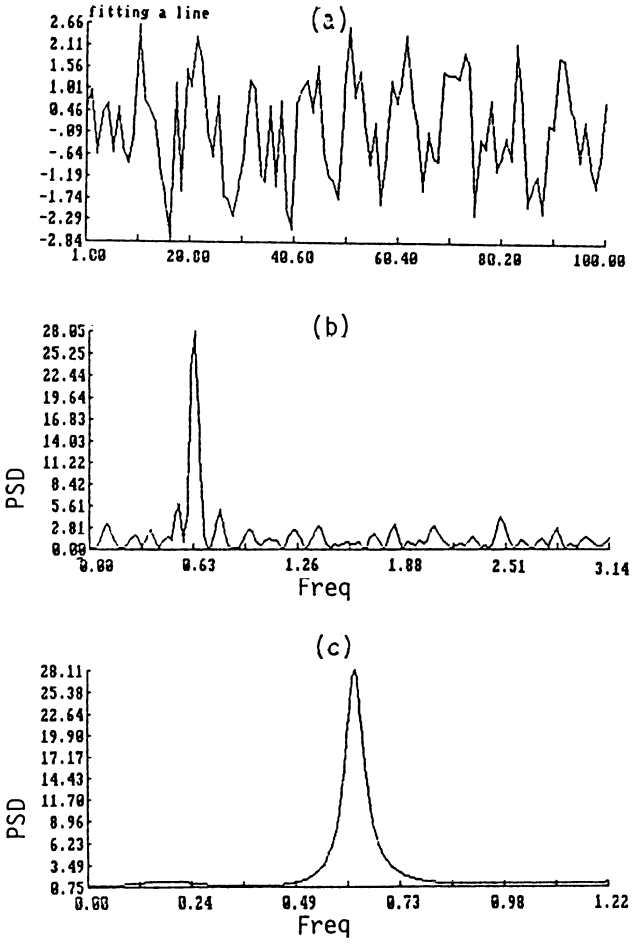


Figure IV. Making the right assumption about the trend of trend-stationary series

3.1 DETRENDING TREND-STATIONARY DATA

Faced with the misleading results obtained from MESA and CSA estimates when applying them directly to nonstationary data, the next step would be to make that data stationary and reapply the methods again. To do that we need

to make some assumption about the type of trend. Economists have been focusing recently on two assumptions: (a) the trend is a function of time around which short-run fluctuations occur —trend-stationary series—; (b) the trend is the accumulation of differences —difference-stationary series—. It is even considered, based on recently developed tests, that most macroeconomic time series are difference-stationary (Stock and Watson 1988). And that making the wrong assumption has far-reaching consequences in applied work (Serietis 1988). We analyze in this section the effect on the frequency estimates of making one assumption when the data has been designed under the other.

Consider first the data in Figure III which is trend-stationary according to the model that generated it. A common procedure to detrend this type of data would be to fit a polynomial function of time of appropriate degree to obtain stationary residuals, such as those in Figure IV (a) obtained after fitting a line to the original data. Figure IV (b) shows the MESA spectrum and Figure IV (c) the CSA one. In this case these two methods perform as well as BFPE (Figure IV (d)).

However, if we make the wrong assumption, i.e. if we consider the series difference-stationary and detrend accordingly, we obtain misleading results from MESA and CSA as we illustrate in Figure V.

Figure V (a) shows the residuals obtained after detrending the data in Figure III (a) by first order differencing. Figure V (b) is the estimated Classical Spectrum. Figure V (c) is the MESA estimate. As we can see many spurious effects have been introduced into the estimated spectrum by assuming the wrong type of nonstationarity for the data.

We carried several other experiments with data that contained higher degree polynomial trends and several frequencies and we reached the same conclusion: if the trend is just a function of time around which a periodic signal and noise evolve, differencing that data will have far-reaching consequences in the estimated spectra. In none of the cases we analyzed did we get a clear answer after differencing trend-stationary data. However, applying the BFPE method to the same data we got a complete information on the periodicities present in the series except in those cases where the noise dominated the signal completely.

3.2 DETRENDING DIFFERENCE-STATIONARY DATA

Many analysts of economic time series agree that most macroeconomic time series are difference-stationary. In particular, it has been considered that the trend of some variables is a random walk (Nelson & Plosser 1982; Stock & Watson 1986). It seems then relevant to analyze how our methods perform in this situation. Figure VI (a) shows artificial data with three periods (100, 10 and 4 years) and a random walk trend contaminated with gaussian white noise with a variance equal to 16. In Figure VI (b) the stationary residuals are obtained after detrending by taking the first difference. Figure VI (c) and (d) show the MESA and CSA spectra. MESA, with an assumed lag equal to sixteen gives much more information about

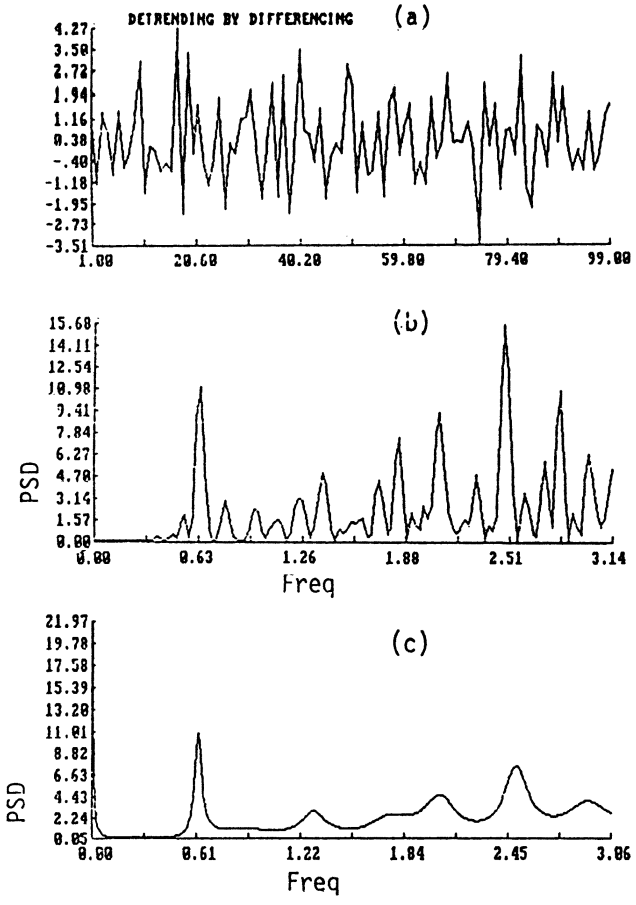


Figure V. Effect of making the wrong assumption about the trend of trend-stationary series on the results obtained from MESA and CSA

the periodicities present in the data than the Classical spectrum. The latter does not pick up the lowest frequency in the the data. When we detrend the data fitting a polynomial, though, the results from the two methods are misleading.

The application of BFPE to this method requires much more work than in the case when the trend was polynomial. The model selection procedure explained in Bretthorst (1987) tells us that besides the three frequencies present in the data the latter has much more structure that needs to be accounted for (Figure VI (e) and (f) respectively). This example suggests that using MESA and BFPE together may be more appropriate when we are dealing with difference- stationary data.

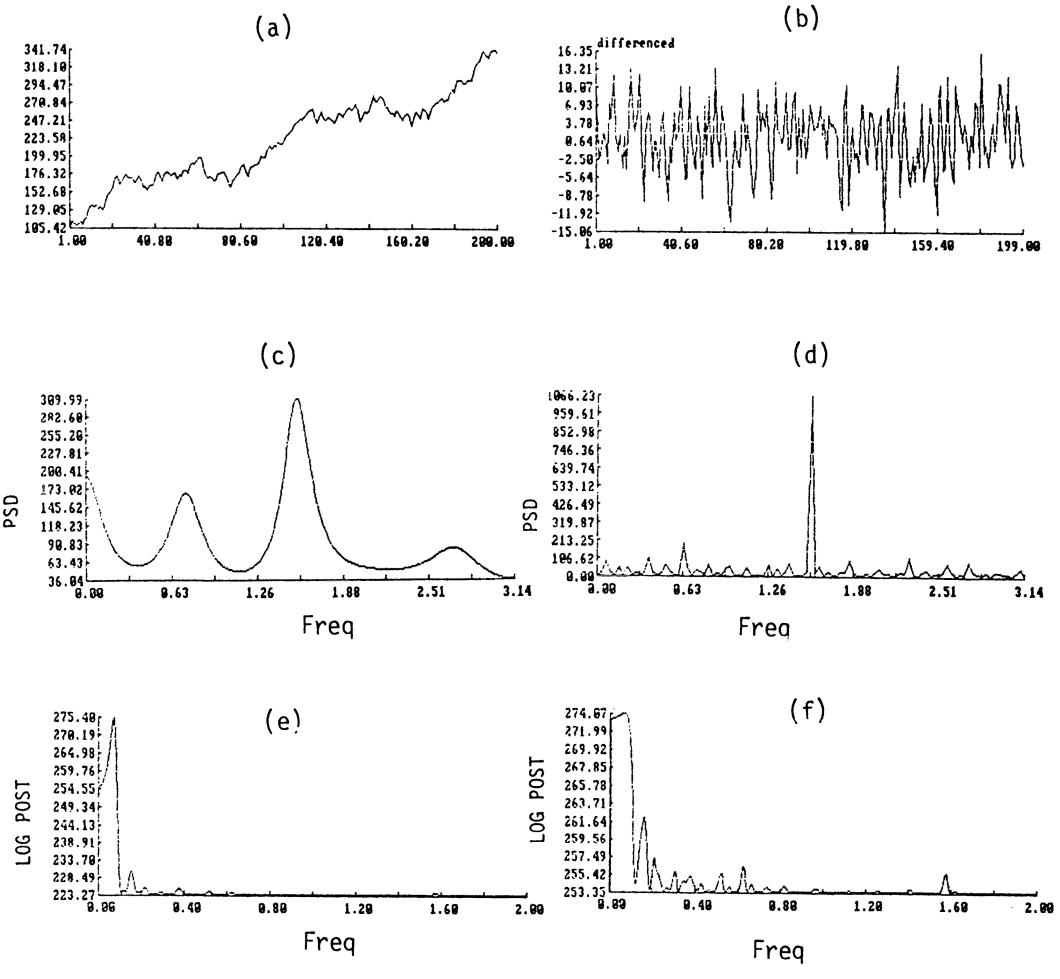


Figure VI. Application to data with a random walk trend plus a periodic function with frequencies 0.06, 0.6 and 1.57

4. Conclusions

We have presented in these pages some preliminary results concerning the sensitivity of the BFPE, MESA and CSA methods to two well known detrending procedures. With the simple examples used we have found that making the wrong assumption about the trend does indeed have undesirable effects on the results obtained from MESA and CSA. This is not the case with BFPE when the trend is a function of time. But when the method is applied to more complicated trends we need some calculations not yet done that would allow the method to incorporate this more complex information into the estimation process. This will be the subject

of further research. In any case using the three methods for the analysis of any time series seems a good advise to follow.

Additional work needs to be done to fully assess the performance of the three methods with other sets of data. The simple results obtained here, though, are an indication of the great care with which they must be applied to economic or any type of nonstationary data.

References

Bretthorst, L., "Bayesian Spectrum Analysis and Parameter Estimation," Ph.D. thesis, University Microfilms Inc., Washington University, St.Louis, MO, Aug. 1987.

Burg, J.P., "Maximum Entropy Spectral Analysis," Ph.D. dissertation, University Microfilms No. 75-25, Stanford University, 1975.

Childers, D.G.,Ed., *Modern Spectrum Analysis.* IEEE Press, 1978.

Cochrane, J.H., "How Big is the Random Walk Component in GNP," *Journal of Political Economy*, forthcoming.

Jenkins, G.M. & Watts, D.G., *Spectral Analysis and Its Applications.* Holden Day, 1968.

Marple, S.L., *Digital Spectral Analysis.* Prentice Hall. Inc., 1987.

Nelson, C.R. & Kang, H., "Spurious Periodicity in Inappropriately Detrended Time Series," *Econometrica*, May 1981, *49*, No.3, 741-751.

Nelson, C.R. & Plosser C.I., "Trends and Random Walks in Macroeconomic Time Series", *Journal of Monetary Economics*, 1982, *10*,139-162.

Serietis, A., "The Empirical Relationship Between Money, Prices and Income Revisited," *Journal of Business & Economic Statistics*, July 1988, *8*, No.3, 351-358.

Stock, J.H. & Watson, M.H., "Variable Trends in Economic Time Series," *The Journal of Economic Perspectives*, Summer 1988, *2*, No.3, 147-174.

Stock, H.H. & Watson, M.H., "Does GNP Have a Unit Root," *Economic Letters*, 1986, *22*, 147-151.

Identifying Discrete Cycles in Economic Data: Maximum Entropy Spectra and the Direct Fitting of Sinusoidal Functions

Claude Hillinger
Monika Sebold

SEMECON - University of Munich
Ludwigstr. 28/Rgb.
8000 München 22

Germany

ABSTRACT. There is a long tradition of observations on business "cycles". Such cycles may be defined as fluctuations about the trend which repeat themselves at roughly constant intervals.

At SEMECON - University of Munich - a methodology was developed by means of which such cycles could be identified in many contemporary time series. The procedure used is the following: First a periodogram of the detrended data is computed in order to identify possibly peaks. Then a function of the form:

$$\sum_t A_t \cos(w_t t + e_t)$$

is fitted to the data by nonlinear least squares, minimizing over the parameter vectors (A_t, w_t, e_t). The entire methodology gave good results, but has the drawback that the discrimination between genuine and spurious peaks of the periodogram is subtle and requires considerable experience.

Using ME-spectra, we have begun to obtain results, which match those obtained by our earlier methodology, but are considerably more elegant and definite than the periodogram. The ME-spectra are therefor a valuable supplement to our methodology. However, as will be shown, there are considerable additional advantages to be derived from the direct fitting of sinusoidal functions.

1. Introduction

Economic fluctuations have traditionally been viewed as having periodic components. This view, which goes back to the work of Juglar in the last quarter of the Nineteenth Century, motivated the term "business *cycle*" to denote such fluctuations. We use *economic fluctuation* as the general term and *economic cycle* to devote periodic components of fluctuations.

Economists devoted considerable efforts to the study of economic cycles until about 1960. While progress was slow and somewhat unsystematic, it was, nevertheless, cummulatively impressive both at the descriptive and explanatory levels. It was discovered that economic cycles are concentrated in the investment components of total output and that specific forms of investment exhibit cycles with characteristic durations. Specifically, there is a 3-4 year

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 321–329.

cycle in inventory investment and a 6–10 year cycle in equipment investment. These cycles are caused by *inertia*, meaning all of the technical and economic factors which prevent the instantaneous adjustment of the capital stock to its optimal level. These developments are closely related to the study of economic growth in which investment is also the driving force. We refer to the entire tradition, encompassing descriptive and explanatory elements as the *Economic Theory of Cycles and Growth (ETCG)*. The view that economic fluctuations are cycles will be called the *Genuine Cycle View (GCV)*. The *GCV* is a part of the *ETCG*.

For a variety of reasons the *ETCG* and the *GCV* were abandoned by almost all economists during the 1960's. The tradition of analyzing discrete cycles in economic data was kept alive during the past quarter century mainly by one of us (Hillinger). Beginning with simple measures of duration between turning points, the methodology soon concentrated on the direct fitting of sinusoidal functions to data by means of nonlinear least squares. This method proved effective for analyzing discrete cycles in the very short economic time series (20–30 years) which are typically available between major wars or other breaks in continuity.[1]

2. Methods used

Among statistical techniques of time series analysis, the *Component Model* is closely associated with the ETCG and the *GCV*. In the case of economic time series the component model assumes that these are additively composed of a smooth trend, economic cycles, as well as seasonal and random components.

2.1 TREND REMOVAL

We have experimented extensively with alternative methods of trend removal. The aim was to determine on the basis of both theoretical and practical considerations, those methods which allow the reliable and robust computation of plausible and replicable spectra for the deviations from the trend.

The simplest method which meets these criteria and has been extensively used is the least squares fitting of polynomial trends. We usually select the trend order k for which the adjusted R^2 reaches it first maximum, given k lower or equal 3. Otherwise the order is 3. This criterium is supplemented by a visual examination of a plot of the trend against the data.

2.2 THE DIRECT FITTING OF SINUSOIDAL FUNCTIONS WITH EXPLORATORY DATA ANALYSIS BASED ON THE PERIODOGRAM

The deviations from trend (or mean) are analyzed to determine if they contain discrete cyclical components. The most intuitive way to proceed is to fit a sum of m cosine functions to the detrended data, x_t, t=1,n, using nonlinear least squares. The sum of squares of the residuals is minimized simultanously with respect to the amplitudes, frequencies and phases:

---

1 The evolution of the *ETCG* and the *GCV* as well as previous work on economic cycles at SEMECON – University of Munich – are discussed in Hillinger (1982, 1986, 1987).

$$\min \sum_{t=1}^{n} \left(x_t - \sum_{i=1}^{m} a_i \cos(w_i t + e_i) \right)^2 \Rightarrow \hat{a}_i, \hat{w}_i, \hat{e}_i, i = 1, m$$

We employ the normalized periodogram to select the order m of the model. The highest peaks of the periodogram generally correspond to the cycles which can be directly fitted. The problem is to distinguish between real and spurious peaks, such as side lobes and some of other phenomena. This could not be dealt successfully by smoothing the periodogram.

Our experience with many empirical analyses and simulation experiments yield the following rules of thumb for the interpretation of the periodogram:

i) A peak below 10% in the neihbourhood of a big peak is a spurious peak.

ii) Flat peaks in the long period range are remainders of the trend.

iii) If two peaks are nearly of the same height, the shorter cycle is more significant.

Diagnostic checks of the residuals test the model selection. Cyclical behaviour of the residuals and their autocorrelations indicate a higher order model. The double standard deviation of a white noise autocorrelation function is another useful instrument for the visual examination of the residuals. A quantitative test against white noise is given by the the cumulated periodogram calculated at the orthogonal frequencies (Durbin, 1969). If the maximal distance between the cumulated periodogram of the residuals and the theoretical one for the white noise exceeds a critical value, determined by a procedure analogous to the Kolmogoroff-Smirnov-test, the hypothesis of white noise residuals is rejected. For our purposes this test is advisable, because in the case of rejection the location of the maximal distance gives an idea of an alternative model for the direct fit.

In conclusion it may be said, that this methodology for identifying economic cycles is complex and requires considerable experience of the user.

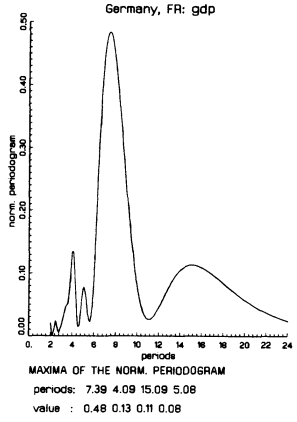
2.3 THE ME-SPECTRUM AND THE DIRECT FITTING OF SINUSOIDAL FUNCTIONS

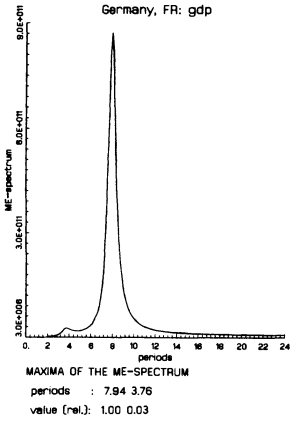
The ME-spectrum is given by the formula (Burg, 1967)

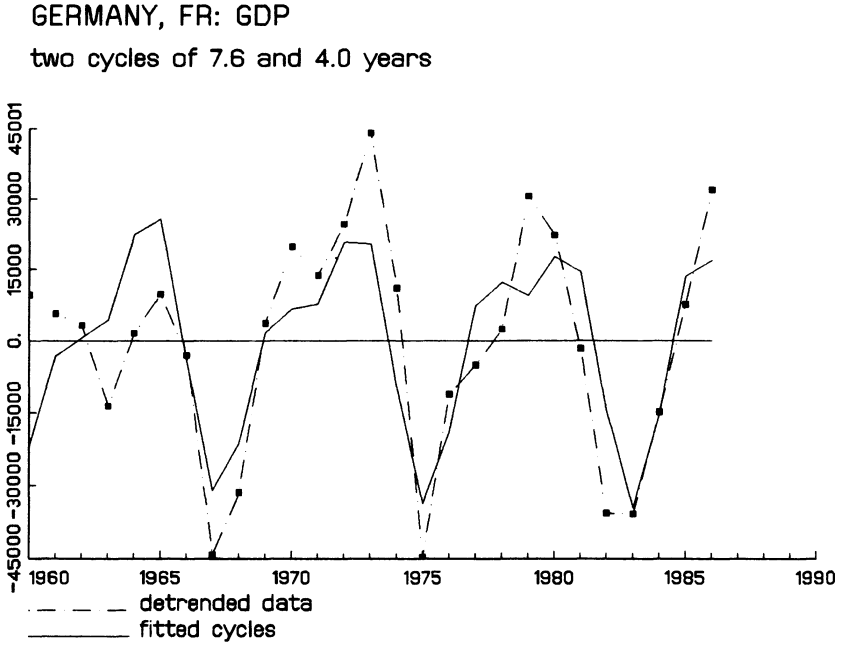
$$S_{ME}(w) = \frac{1}{2\pi} \frac{a_0}{\left| 1 + \sum_{k=1}^{m} a_k e^{-iwk} \right|^2}$$

where the a_i are the coefficients of an AR-process assumed to have generated the data. The coefficients of this process must first be estimated. We use unconstrained least squares, minimizing the unweighted sum of the squared forward and backward forcasting errors of the chosen AR-model (cf.Geyer, 1986, 104-106).

The main problem is to chose the number m of coefficients, which determine the order of the underlying AR-process. For this purpose we use the partial autocorrelation function and the CAT-criterion of Parzen (cf. Priestley, M.B., 1981, 370-380, 600-601). The results of both procedures differ only slightly.


Germany, FR: gdp
norm. periodogram
periods
MAXIMA OF THE NORM. PERIODOGRAM
periods: 7.39 4.09 15.09 5.08
value : 0.48 0.13 0.11 0.08


Germany, FR: gdp
ME-spectrum
periods
MAXIMA OF THE ME-SPECTRUM
periods : 7.94 3.76
value (rel.): 1.00 0.03


GERMANY, FR: GDP
two cycles of 7.6 and 4.0 years
45001
30000
15000
0.
-15000
-30000
-45000
1960
1965
1970
1975
1980
1985
1990
detrended data
fitted cycles

The ME-spectrum substantially eliminates the problem of spurious peaks, such as side lobes, which cause the periodogram to be difficult to work with. Empirically, the peaks produced, match closely our results from the direct fitting of cosine functions.

3. Empirical results

The two methodologies for identifying economic cycles are compared looking at three important economic series, observed annually, for 1960 - 1986, at constant prices: GDP of Germany, inventory investment of the United Kingdom and the equipment investment of the USA (OECD, 1988).

According to economic theory, the investment series are both causally involved in the generation of cycles and also exhibit them more prominently than other ecomomic time series. More specifically, inventory investment exhibits the short (3 - 5 year) cycles and fixed investment the longer (7-10 year) cycles. The GDP contains both of these series and in addition public and private consumption expenditures as well as the foreign sector. GDP is the key variable for evaluating economic performance and enters economic and political decisions. So these three series are analyzed. We chose different countries demonstrate the general relevance of economic cycles.

For all series, the methodology based on the periodogram and the direct fitting of sinusoidal functions is compared with the ME-spectrum. The residuals of the direct fit are tested with the cumulated periodogram at a significance level of 5%.

3.1 GDP, GERMANY

Figure 1 shows the normalized periodogram. In the legend the periods of the main peaks and the corresponding R^2 are given. A prominent cycle at a period of 7.4 years explains nearly 50% of the variation of the detrended data (trend order 3). Lower peaks are located at 4.1, 15.1 and 5.1 years.

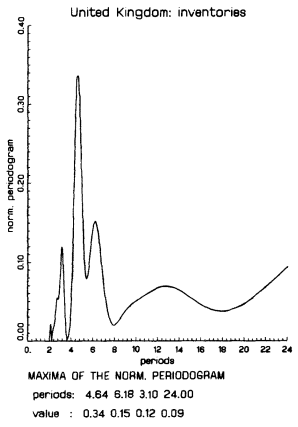
Following the rule of thumb, the long 15.1-year cycle can be interpreted as a remainder of the trend. The 5.1-year cycle looks like a spurious peak caused by the side lobes of the big peak at 7.4 years and the peak at 4.1 years.

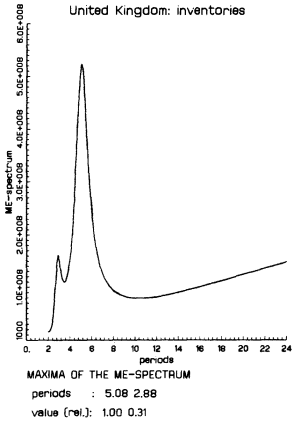
A sum of two cycles is fitted directly. Starting values near 4.1 and 7.4 for the minimizing routine produce a 4.0- and a 7.6-year cycle (Figure 2). The amplitude of the longer cycle is almost double that of the shorter one, as one expects on the basis of the periodogram.

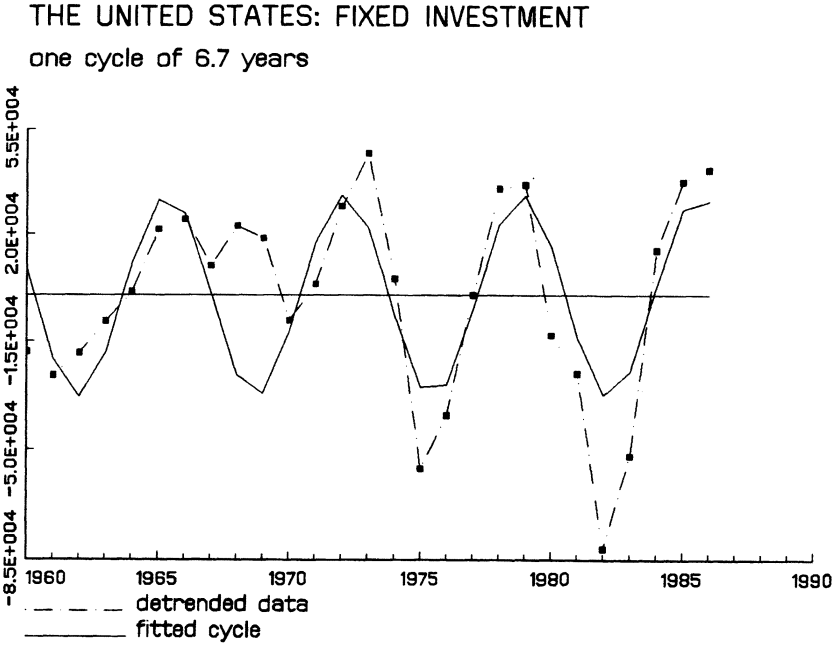
The validity of the regression is confirmed by a high R^2 (0.62) and the significance of the amplitudes indicated by the standard deviations. The test statistic of the cumulated periodogram of 0.413 is lower than the critical value 0.446 and accepts that there is no structure left in the residuals. Also, the residuals and their autocorrelation function lie between the two standard deviation bounds.

The ME-spectrum is shown in Figure 3. For the peaks the periods and height relative to the biggest maximum are written down. A cycle in the range of 8.0 years is suggested. A very low peak at 3.8, which only reaches 3% of the height of the larger peak, indicates a second short cycle.

All tests confirm that only two cycles of about 4 and 7.6 years are relevant for the German data.


United Kingdom: inventories
norm. periodogram
periods
MAXIMA OF THE NORM. PERIODOGRAM
periods: 4.64 6.18 3.10 24.00
value : 0.34 0.15 0.12 0.09


United Kingdom: inventories
ME-spectrum
periods
MAXIMA OF THE ME-SPECTRUM
periods : 5.08 2.88
value (rel.): 1.00 0.31


THE UNITED STATES: FIXED INVESTMENT
one cycle of 6.7 years
-8.5E+004
-5.0E+004
-1.5E+004
2.0E+004
5.5E+004
1960
1965
1970
1975
1980
1985
1990
detrended data
fitted cycle

3.2 INVENTORIES, UNITED KINGDOM

The periodogram of the deviations from mean (Figure 4) indicates three cycles at 4.6, 6.2 and 3.1 years. Two flat peaks in the range of the longer periods could again be interpreted as remainders of the trend. The fact that a 6 year cycle is rarely seen in inventory data suggests some skepticism regarding this peak.

Consequently, three different model are fitted and compared, to examine the relevance of the two lower peaks:

i) two cycles with period 3.2 and 4.8
ii) two cycles with period 4.8 and 6.2
iii) three cycles with period 3.2, 4.8 and 6.2

The first regression passes all tests. The fitted model (Figure 5) matches the major troughs and peaks quite well. The residuals and their autocorrelations indicate no further structur. The value 0.21 for the statistic of the test for white noise lies clearly beyond the critical value 0.38.

Although the R^2 is rising from 45% to 49% for the second fit, the diagnostic checks show that there remains a cyclical pattern in the residuals. A statistic of 0.44 against a critical value of 0.38 rejects this model. The cumulative periodogram indicates that this is due to the excluded 3 year cycle.

The plot of the regression with all three cycles reveals, that a third cycle at 6.2 years does not change the structure of the regression substantially. The additional cycle does not improve the location of the major turning points. For 1969 this regression even places a trough instead of a peak. The first regression is superior to it.

The spurious nature of the 6.2 cycle is confirmed by its absence in the ME-spectrum (Figure 6). Two cycles at 5.1 and 2.9 years are detected. The maximum at the shorter cycle reaches 30% of the longer.

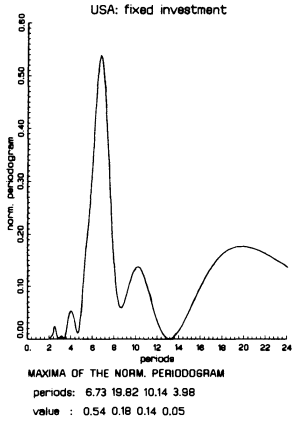
This example confirms the ME-spectrum as an easy to handle instrument for identifying the relevant economic cycles. In the case, just discussed, an inexperienced user of the first methodology could easily be mislead by the weight of the 6.2 - peak in the periodogram and the rise of the R^2. The ME-spectrum identified the correct model.

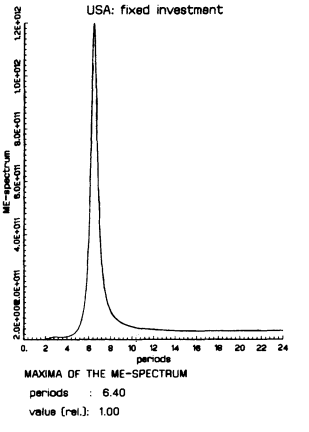
3.3 FIXED INVESTMENT, USA

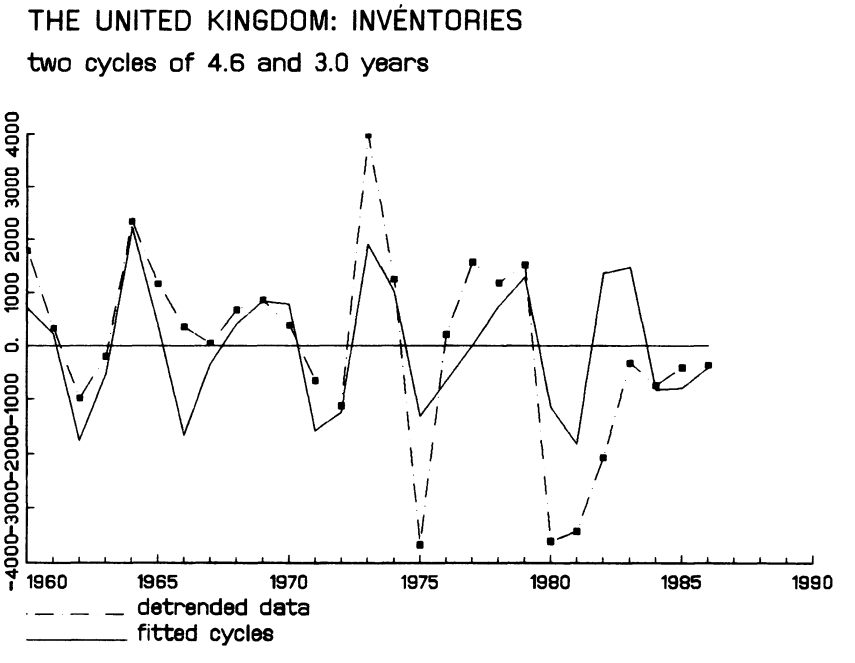
In the periodogram (Figure 7) a peak at 6.7 years overshadows three lower peaks at 19, 10 and 4 years. The flat peak at 19 years again is interpreted as a remainder of the trend. The other two peaks, around the dominant 6.7-maximum, look like side lobes. A failed attempt to fit in addition to the 6.7-cycle a 10.1 cycle confirms this conclusion.

The 6.7 years cycle provides a good fit to deviations from a linear trend (Figure 8). The plot of the detrended data, together with the estimated regression, and the residual check support the hypothesis of no additional cycle. However, the visual examination of the detrended data suggests that the cycle is exploding. In this case a model with varying amplitude would yield better results. A modification of the constant amplitude model in this direction is indicated.

The ME-spectrum (Figure 9) again supports and simplifies the above considerations. It shows a single sharp peak at the period 6.4. With this as a starting value for the direct fit the same results as above are obtained.


USA: fixed investment
norm. periodogram
periods
MAXIMA OF THE NORM. PERIODOGRAM
periods: 6.73 19.82 10.14 3.98
value : 0.54 0.18 0.14 0.05


USA: fixed investment
ME-spectrum
periods
MAXIMA OF THE ME-SPECTRUM
periods : 6.40
value (rel.): 1.00


THE UNITED KINGDOM: INVENTORIES
two cycles of 4.6 and 3.0 years
-4000 -3000 -2000 -1000 0. 1000 2000 3000 4000
1960 1965 1970 1975 1980 1985 1990
detrended data
fitted cycles

4. Conclusion

The positive results obtained by means of the ME-spectra have motivated us to incorporate them into our methodology for identifying economic cycles. This methodology now consists of four steps:

i) Fit a deterministic trend. Visually examine plot of data against trend and of residuals from trend.

ii) Plot the periodogram to obtain an alternative view of the data without imposing any assumptions.

iii) Select the order of an AR-process, which could generate the data, and derive the corresponding ME-spectrum.

iv) On the basis of the cycles selected in the previous steps, fit a sum of sinusoidal functions directly by nonlinear least squares. Use the residuals for diagnostic check.

The advantages of the final step are

a) Estimates of all the parameters, including the phases, as well as of the standard errors, are obtained.

b) The residuals obtained can be used for diagnostic check.

c) The functions can be extrapolated for forecasting purposes, an aspect on which we are currently working.

d) The methodology is easily extended to damped or exploding cycles, which are often encountered in economic data.

The four step methodology suggested, appears to be ideal for analyzing economic cycles. We can think of no reason why it should not work in other areas as well.

5. References

Bard, Yonathan, 1974, *Nonlinear parameter estimation*, Orlando, Academic Press

Burg, John Parker, 1967, 'Maximum entropy spectral analysis'(reprinted), in: Childers, Donald G., eds., *Modern spectrum analysis*, New York, IEEE press 34 -41

Childers, Donald G., 1978, *Modern spectrum analysis*, New York, IEEE press

Geyer, Alois, 1985, *Maximum entropie spektralanalyse oekonomischer Zeitreihen*, Wien, VWGOE

Hillinger, Claude, 1982, 'Business cycles are alive and well', *Economics Letters*, 9133 - 137

Hillinger, Claude, 1986, 'Inventory cycle and equipment cycle interaction', in: Chikan, Attila, eds., *Inventories in theory and practice*, Amsterdam

Hillinger, Claude, 1987, 'Business cycle stylized facts and explanatory models', *Journal of Economic, Dynamics and Control*, 11257 - 263

OECD, 1988, *National accounts main aggregates* , vol **1**, Paris

Press, William ; Flannery, Brian P. ; Teukolsy, Saul A. ; Vetterling, William T., 1986, *Numerical recipes*, Cambridge, Cambridge University Press

Priestley,; M.B., 1981, *Spectral analysis and time series*, vol **1**, New York, Academic Press

Ulrych, Tad J., 1972, 'Maximum entropy power spectrum of truncated sinusoids', *Journal of Geophysical Research*, **77**, 81396 - 1400

Maximum Entropy Spectral Analysis of Hilbert Transformed Complex Data

Michael R. Sturgill
Senior Engineer, Motorola Government Electronics Group,
8201 E. McDowell Rd., Scottsdale, Arizona, 85252 U.S.A.
Louis E. Roemer
Department of Electrical Engineering,
The University of Akron, Akron, Ohio 44315 U.S.A.

Abstract-This paper evaluates Hilbert transformed complex data using complex maximum entropy spectral estimation to determine if a reduction in pole complexity results. Since the computational load of the maximum entropy method (MEM) scales as a function of predictor length, a shorter length predictor may be used in some cases when presented with complex data. This paper deals with the results obtained from simulated real and complex data and the resultant pole locations of the prediction filter.

1. INTRODUCTION

The frequency resolving performance of maximum entropy spectral analysis for short data lengths is well documented. Its use is becoming an increasingly important spectral analysis procedure as the detection and classification of threats, both sonar and radar, become more difficult. One limitation in its use however, is the computational load it places on an imbedded processor operating in real time. These situations require the algorithm to extract as much information as possible with minimal computation.

The thrust in Hilbert transforming the input data is to see whether two complex poles will give as accurate of a spectral estimate as would four real poles. If this was the case, the calculation of the power spectrum could be carried out with half the number of coefficients. Roots of the power spectral equation would also be easier to find due to the pole reduction.

In section 2, a review of the important features of the Hilbert transform is presented. In section 3, the maximum entropy method in its complex form is discussed. Section 4 presents the results and simulation methodology used in the formulation of the results.

2. HILBERT TRANSFORM

The Hilbert transform is derived using the notion causality [1]. Causality is defined in this context as the Fourier transform of a sequence being zero in the frequency range $-\pi \leq \omega < 0$ which is the bottom half of the unit circle in the z domain. Therefore, given a sequence

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 331–338.

$s(n)$ and its resultant Fourier transform $S(e^{j\omega})$,

$$S(e^{j\omega}) \equiv 0, \qquad -\pi \le \omega < 0 \tag{1}$$

From (1) it is required that the input sequence $s(n)$ be complex and denoted by

$$s(n) = s_r(n) + js_i(n) \tag{2}$$

where $s_r(n)$ and $s_i(n)$ are real sequences.

$S_r(e^{j\omega})$ and $S_i(e^{j\omega})$ denote the Fourier transforms of the real sequences $s_r(n)$ and $s_i(n)$, respectively, and are given by

$$S_r(e^{j\omega}) = \frac{1}{2}[S(e^{j\omega}) + S^*(e^{-j\omega})] \tag{3a}$$

and

$$jS_i(e^{j\omega}) = \frac{1}{2}[S(e^{j\omega}) - S^*(e^{-j\omega})] \tag{3b}$$

By Eq. (1) and assuming no overlap between the zero and non-zero portions of $S(e^{j\omega})$ and $S^*(e^{-j\omega})$, $S(e^{j\omega})$ can be completely recovered from either $S_r(e^{j\omega})$ or $S_i(e^{j\omega})$. $S_i(e^{j\omega})$ can now be written in terms of $S_r(e^{j\omega})$ by

$$S_i(e^{j\omega}) = \begin{cases} -jS_r(e^{j\omega}), & 0 \le \omega < \pi \\ jS_r(e^{j\omega}), & -\pi \le \omega < 0 \end{cases} \tag{4}$$

or

$$S_i(e^{j\omega}) = H(e^{j\omega}) S_r(e^{j\omega}) \tag{5}$$

where

$$H(e^{j\omega}) = \begin{cases} -j, & 0 \le \omega < \pi \\ j, & -\pi \le \omega < 0 \end{cases} \tag{6}$$

Eqs. (5) and (6) show that $s_i(n)$ is found directly from $s_r(n)$ by a discrete system having a frequency response $H(e^{j\omega})$. The response is one of unity gain and constant phase angle of $-\pi/2$, $0 \le \omega < \pi$ and $\pi/2$, $-\pi \le \omega < 0$. This 90 degree phase shifter is known as a Hilbert transform.

The impulse response $h(n)$ of the described Hilbert transformer can be obtained by the inverse Fourier transform of Eq. (6) and is

$$h(n) = \begin{cases} \frac{2}{\pi} \frac{\sin^2(\pi n/2)}{n}, & n \ne 0 \\ 0, & n = 0 \end{cases} \tag{7}$$

Eq. (7) is multiplied by the input data to give the Hilbert transformed version of the data. This is then called 'imaginary' data when referring to the input of the MESA procedure.

3. MAXIMUM ENTROPY METHOD

The maximum entropy method is a power spectral density (PSD) calculation technique originally proposed by Burg [3] to process finite lengths of real valued data. The method has been extended to include the processing of complex valued data as well. This paper includes the complex development as outlined in [4].

The MEM for spectral estimation is based upon the extrapolation of a segment of a known autocorrelation function for lags which are not known. By estimating the unknown lags, the problems associated with truncation of the waveform or assumed periodicity found in other spectral analysis techniques can be avoided. The PSD of the input data is found by evaluating an all pole filter as a function of frequency whose coefficients are found using the MEM. The equation for the PSD is given in Eq. (8)

$$P(f) = \frac{P_{M+1}}{\left| 1 - \sum_{k=1}^{M} a_{M,k} \exp^{(-j2\pi f k \Delta t)} \right|^2} \cong \sum_{i=-M}^{M} \phi_i \exp^{(-j2\pi f i \Delta t)} \tag{8}$$

where P_{M+1} and $a_{M,k}$ are the coefficients to be found. Eq. (8) implies a linear set of relations between the autocorrelations ϕ_i and the coefficients $a_{M,k}$. They satisfy the matrix equation of (9).

$$\begin{bmatrix} \phi_0 & \phi_1 & \phi_2 & \cdots & \phi_M \\ \phi_1 & \phi_0 & \phi_1 & \cdots & \phi_{M-1} \\ \phi_2 & \phi_1 & \phi_0 & \cdots & \phi_{M-2} \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \phi_M & \phi_{M-1} & \phi_{M-2} & \cdots & \phi_0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_n \end{bmatrix} = \begin{bmatrix} P_{M+1} \\ 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \tag{9}$$

The matrix of Eq. (9) is a symmetric Toeplitz matrix where the ϕ_i and a_{Mk} are in general complex. The Toeplitz matrix is of order [MxM] and the maximum entropy method gives an estimate of the autocorrelation value ϕ_{M+1}. An efficient algorithm for the calculation of the unknown coefficients was originated by Burg and is presented here.

Initially, dummy parameters $b_{Mk}, b'_{Mk}; k = 1, 2, \ldots, N-M$ where N is the input data record length, are calculated as follows.

$$b_{Mk} = b_{M-1,k} - a^*_{M-1,M-1} \cdot b'_{M-1,k} \tag{10}$$

$$b'_{Mk} = b'_{M-1,k+1} - a_{M-1,M-1} \cdot b_{M-1,k+1} \tag{11}$$

where $b_{M-1,k}, b'_{M-1,k}$ are previously defined values. Now the a_{Mk} are calculated as follows.

$$a_{MM} = \frac{2 \sum_{k=1}^{N-M} b^*_{Mk} b'_{Mk}}{\sum_{t=1}^{N-M} \left(|b_{Mk}|^2 + |b'_{Mk}|^2 \right)} \tag{12}$$

$$a_{Mk} = a_{M-1,k} - a_{MM} \cdot a^*_{M-1,M-k} \qquad 1 \le k \le M-1 \tag{13}$$

and

$$P_{M+1} = P_M \cdot \left(1 - |a_{MM}|^2\right) \tag{14}$$

The initial values are

$b_{1k} = x_k \qquad k = 1, 2, \ldots, N-1$

$b'_{1k} = x_{k+1} \qquad k = 1, 2, \ldots, N-1$

$a_{M0} = -1$

$a_{Mk} = 0, \qquad k > M$

and

$$P_0 = \sum_{k=1}^{N} |x_k|^2 / N \tag{15}$$

Equations (10) and (11) are not used for $M=1$. This iterative method described was implemented in complex form using the programming language TURBO Pascal. The results presented are an outcome of this simulation.

4. RESULTS

Equation (8) shows the power spectral density relationship of the coefficients a_{Mk}. Eq. (8) is a polynomial in z where $z=e^{j\omega t}$. By expanding the denominator of (8) inside the $|\ |^2$ brackets, the following results.

$$1 - \sum_{k=1}^{M} a_{Mk} z^{-k} = 1 - a_1 z^{-1} - a_2 z^{-2} - \ldots - a_M z^{-M} \tag{16}$$

where M is the length of the prediction error filter and N is the length of the input data record, $M<N$. The roots of Eq. (16) are complex conjugates when the input data is real only. It will be seen that if the input consists of complex data where the imaginary part is obtained from the Hilbert transform of the real data, the resultant PSD found from the MEM will be zero for frequencies $-\pi \le \omega < 0$. This result was derived in section 2.

The roots of Eq. (16) can be found for any order M. In this paper, only $M=1, 2,$ and 4 will be considered for real and complex data. Fig. (1) shows the block diagram of the methodology used in the simulated data results. It shows that either one or two sinusoids were generated with or without Gaussian random noise summed. The signal is then input into the MEM procedure for the case of real data. For the case of complex data, this output is run through the Hilbert transformer and then into the MEM procedure. In either case the MEM produces a power spectral density plot and the coefficients of Eq. (8). These coefficients are then used in order to find the roots of the denominator polynomial in Eq. (16). The plots for the simulated cases can be found in figures 4-6. These figures contain six plots each and each plot is denoted by a different symbol. The annotations on the

plots such as **1Cmplx** and **4Real** mean that a first order estimate of Hilbert transformed complex data and a fourth order estimate using real data was used, respectively. The other annotations follow the same format. The data length in all cases is 64 points and the amplitudes of the sine waves are all unity.

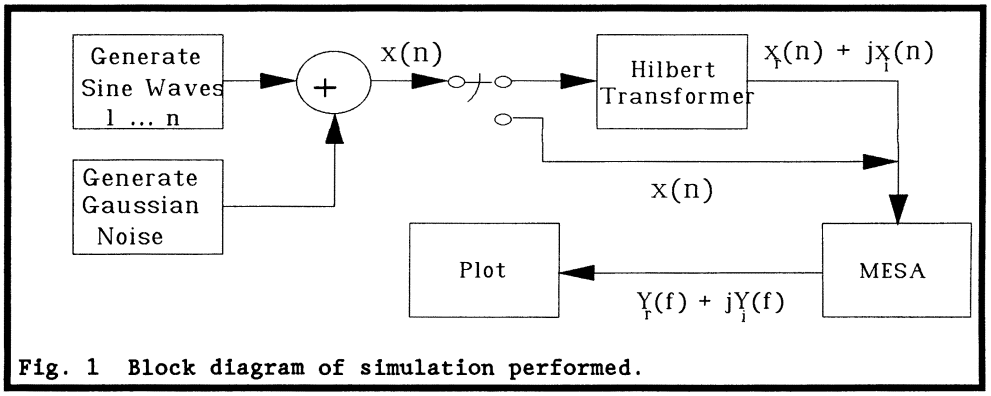


Fig. 1 Block diagram of simulation performed.

Case #1: *M=1* For the case of M=1 and real data, Eq. (16) reduces to

$$1 - a_1 z^{-1} \tag{17}$$

Eq. (17) has one real pole located on the real z axis at a value of $z = a_1$.

The plots obtained for a single order estimation with added Gaussian noise having a standard deviation of 0.5 and 1.0 and a single sinusoidal input is shown in Figures 4 and 5, respectively. It can be seen from these plots that little information about the true spectrum is obtained from a first order estimation using real data. There is no frequency information obtained from a single pole lying on the real z axis.

For the case of M=1 and complex data, Eq. (17) has one complex pole associated with it. The pole location is *0.435042 + j0.763544* and is shown in Fig. 2 on the z axis plot. The resultant complex MEM spectrum is much different from the ones in the real data case. In this case the MEM was able to detect the presence of a tone in the data. Figures 3 and 4 show the resultant spectral characteristics of this complex estimate. These PSD plots also show the nonexistence of any appreciable spectral power in the range $f_s/2 \le f < f_s$ (where f_s is the sample rate of the system) and corresponds to the negative frequencies $-\pi \le \omega < 0$. This result agrees with that which was defined in Eq. (1).

Case #2 *M=2* For the case of M=2 and real data, Eq. (16) reduces to

$$1 - a_1 z^{-1} - a_2 z^{-2} \tag{18}$$

Eq. (18) has two poles located on the complex z plane at complex conjugate values.

Eq. (18) can be factored using the quadratic equation and the resultant pole locations are given by

$$P_1 = \frac{a_1}{2} \pm \sqrt{\frac{a_1^2}{4} + a_2} \tag{19}$$

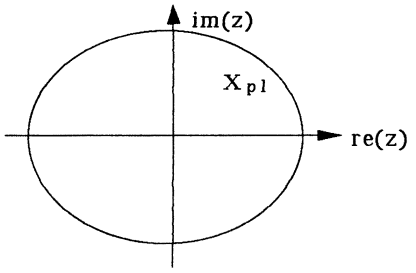


Fig. 2 Single complex pole representation in the z domain.

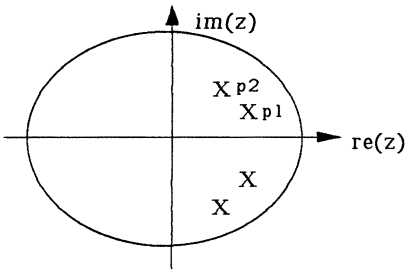


Fig. 3 Fourth order estimate, real data input.

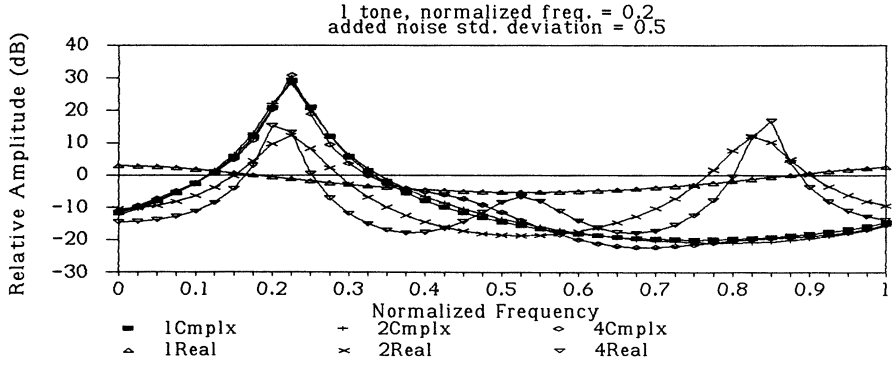


Fig. 4 Real and Complex Data Comparisons.

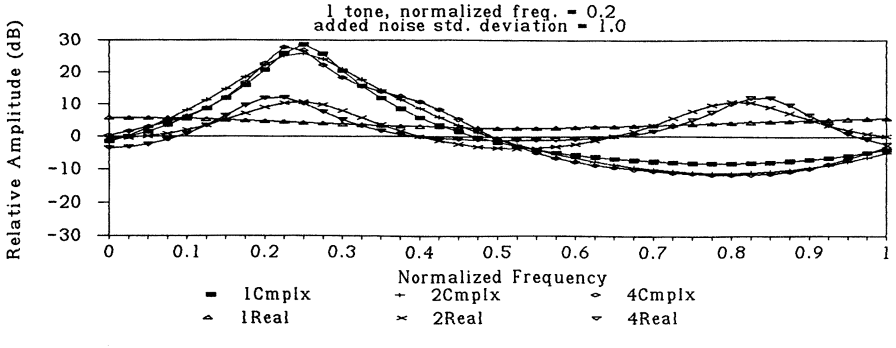


Fig. 5 Real and Complex Data Comparisons.

With the coefficients calculated from the simulated data, the poles are $p_k = 0.5616 \pm j0.8273$ which shows that the roots do indeed occur in complex conjugate pairs. Again, figures 4 and 5 represent the plots for a single sinusoid with additive noise for this case. For the

case *M=2* and complex input data, Eq. (18) has two complex poles associated with it. Unlike the case for real data, these complex poles do not occur in complex conjugate pairs.

Eq. (19) can be used to find the roots of the second order estimate for the coefficients obtained by using the complex form of the MEM. Figures 4 and 5 contain the plots obtained from this estimate.

Case #3 *M=4* For the case of M=4 and real data, Eq. (16) reduces to

$$1 - a_1 z^{-1} - a_2 z^{-2} - a_3 z^{-3} - a_4 z^{-4} \qquad (20)$$

Eq. (20) contains two pairs of complex conjugate poles located in the z plane as shown in Fig. (3).

Factoring Eq. (20) is not as simple as using the quadratic equation, instead *Laguerre's* method for finding roots of polynomials having complex coefficients was employed. This method led to the following sets of roots for the fourth order case : $p_1 = 0.558250 \pm j0.829649$, $p_2 = 0.572610 \pm j0.819750$. These roots correspond to those shown in Fig. 3. The spectral plots can be seen in Figs. 4 and 5 also.

For the case *M=4* and complex input data, Eq. (20) has four complex poles associated with it. Unlike the case for real data, these complex poles do not occur in complex conjugate pairs.

Laguerre's method is again used here to factor the polynomial of Eq. (20). The poles associated with this set of coefficients are : $p_1 = 0.555286 + j0.830766$, $p_2 = 0.135987 + j0.028344$, $p_3 = -0.050553 + j-0.139076$, $p_4 = -0.085556 + j0.109077$. The resultant spectra are shown in Figs. 4 and 5.

Figure 6 shows plots for the cases of 2nd, 4th, and 8th order estimates of two sinusoids summed with Gaussian random noise. The sinusoids are of unity amplitude and the noise has a standard deviation of 0.5. The data length is again 64 points.

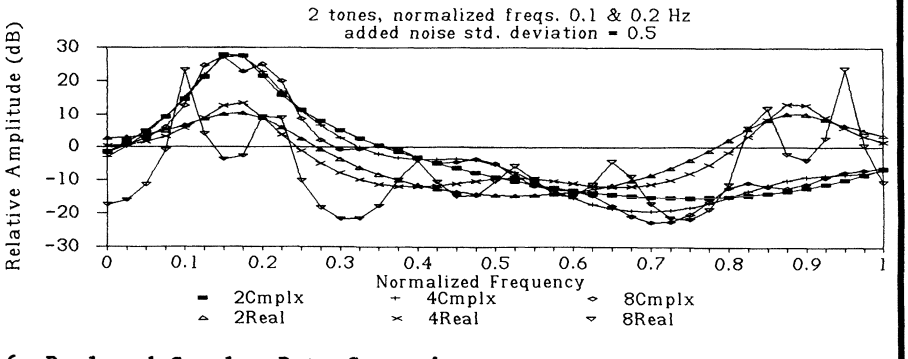


Fig. 6 Real and Complex Data Comparisons.

5. DISCUSSION

This paper presented the derivation of the Hilbert transform for its use in generating complex data from real data. It also discussed the generation of the complex form of the maximum entropy method. The results were mixed as to the effectiveness of the complex form of the

maximum entropy method in its use in frequency discrimination. MEM's effectiveness is dependent upon the data being analyzed [7]. In some cases the complex form of MEM was clearly superior and in others it was clearly inferior. In instances where a single tone buried in noise needs to be found quickly, a first order complex estimate does a very good job of extracting the tone. The first order complex MEM finds the dominant pole associated with the data. This can be extremely beneficial if one is trying to detect the presence of a single short duration tone. By being able to execute this detection with one coefficient, this method can be executed in real time on a number of existing digital signal processors in a few milliseconds. Observing figures 4 and 5 also show that little additional benefit is gained when higher order complex estimates are used. In fact, the plots for the complex cases in figs. 4 and 5 are almost directly on top of one another. This is because the dominant pole is essentially the same in all the cases tested and therefore the remaining poles have little effect on the resultant spectrum. These figures also show that for the purposes of detecting a single tone, the first order complex estimate provides as much detectability as the fourth order real data case.

The results of the complex MEM procedure for greater than one tone does not do any better than the real data case. In most situations in fact, the real case performed better.

As for the pole locations of the estimated filter, the complex form yielded results consistent with those found in Fourier analysis. The complex conjugate nature of the poles disappears when the input data is passed through a Hilbert transformer. The same general results held true for finding roots of the denominator polynomial in Eq. 8. When the fundamental pole location must be found quickly, the first order complex estimate is a good candidate solution.

REFERENCES

[1] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing*. Englewood Cliffs, NJ: Prentice Hall, 1975.
[2] R. Ansari, 'IIR Discrete Time Hilbert Transformers,' *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. **ASSP-35**, No. 8, pp. 1116-1119, August 1987.
[3] J. P. Burg, 'Maximum Entropy Spectrum Analysis,' Ph.D. dissertation, Stanford University, Stanford CA, May 1975.
[4] S. Haykin and S. Kesler, 'The Complex Form of the Maximum Entropy Method for Spectral Estimation,' *Proceedings of the IEEE*, pp. 822-823, May 1976.
[5] W. H. Press, B. P. Flannery, S. A. Tenkolsky, W. T. Vetterling, *Numerical Recipes, The Art of Scientific Computing*. New York, NY: Cambridge Press, pp. 430-435, 1986.
[6] N. Anderson, 'On the calculation of Filter Coefficients for Maximum Entropy Spectral Analysis,' *Geophysics*, vol. **39**, No. **1**, pp. 69-72, February 1974.
[7] E. K. L. Hung, R. W. Herring, 'Simulation Experiments to Compare the Signal Detection Properties of DFT and MEM Spectra,' *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. **ASSP-29**, No. **5**, pp. 1084-1089, October 1981.

MAXIMUM ENTROPY TOMOGRAPHY OF ACCRETION DISCS FROM THEIR EMISSION LINES

T.R. Marsh
Royal Greenwich Observatory
Herstmonceux Castle
Hailsham
East Sussex BN27 1RP
UK

Keith Horne
Space Telescope Science Institute
3700 San Martin Drive
Baltimore
MD 21218
USA

ABSTRACT. The formation of emission lines in close binary stars is shown to be equivalent to the problem of X-ray tomography in medical imaging. We apply the maximum entropy method to find the image of a binary from a series of line profiles taken at different parts of its orbit. The problem is well constrained, however statistical noise does allow the form of prior information (through a default image) to influence some features of the image. We compare the effects of different defaults upon the reconstruction.

1. Introduction

The close binary stars are a rich source of astrophysical phenomena. Interaction between the two stars in close binaries can produce classical nova eruptions, type I supernova eruptions, powerful X-ray sources and the millisecond pulsars. In many systems mass transfer takes place between the two stars, and as the material loses angular momentum accretion discs are formed. Accretion discs play many rôles in astrophysics, and probably occur in the formation of the planets and the power sources of quasars. Close binaries, particularly cataclysmic variable stars, provide the best observational examples of accretion discs.

Emission lines produced by sources stationary in the rotating frame of the binary move sinusoidally in wavelength as the binary rotates. With many sources, the profiles are difficult to interpret. In this paper we show how observations of line profiles can be inverted to measure the pattern of emission as a function of velocity in the binary. We apply this to observations of cataclysmic binaries in which the emission lines come largely from an accretion disc surrounding the white dwarf component of these binaries. The accretion disc in these variables is supplied with gas from a near main-sequence red dwarf.

2. Line profiles from close binaries

The profile of an emission line from a binary can be computed by summing the contribution from every part of the system with the appropriate line width and Doppler shift. The line intensity at each point in the system is the image of the binary.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 339–345.

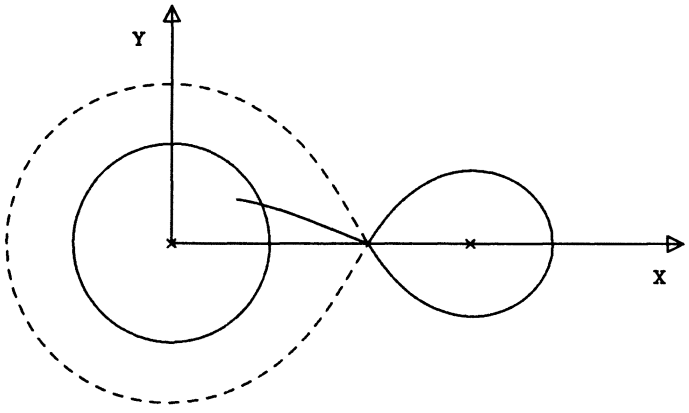


Figure 1: Schematic picture of a close binary with an accretion disc. The circle represents the outer edge of the disc and the curved line represents the gas stream from the red star into the disc.

Computation of a line profile from an image which is a function of position requires a relation between velocity and position. Such a relation is well determined for uniformly rotating systems (Vogt & Penrod 1988), but in general, and in the particular case of accretion discs, this is not the case. It is easier in such cases to consider the image to be a function of velocity in the binary. The velocity vectors of the binary continuously rotate with respect to an inertial frame, and so we define the velocity of a point to be the velocity as measured in a frame at rest with respect to the centre of mass of the binary at orbital phase zero. For a cataclysmic variable phase zero is defined as the phase at which the white dwarf moves perpendicular to our line of sight and is the further of the two stars; the orbital phase increases by one every cycle.

We define the X-axis to point from the white dwarf to the red dwarf and the Y-axis to point in the direction of the red star's orbital motion. These are illustrated schematically in figure 1. With these definitions the radial velocity V_R of a point with velocity (V_X,V_Y) at orbital phase ϕ is given by

$$V_R = \gamma - V_X \cos 2\pi\phi + V_Y \sin 2\pi\phi,$$

where γ is the radial velocity of the centre of mass with respect to the observer. The flux in the line at V_R from the line centre from an image in velocity coordinates is obtained by adding in flux from all the points (V_X,V_Y) satisfying the equation above. These lie on a straight line with gradient determined by the orbital phase ϕ and an offset determined by the value of V_R.

Figure 2 shows the formation of the line profile at two different orbital phases, illustrating some features of the discussion above. Each profile is the projection of the image, represented here by a greyscale image. The image has a smoothly varying background which rises towards the centre until a critical radius below which it falls, and a spot of emission

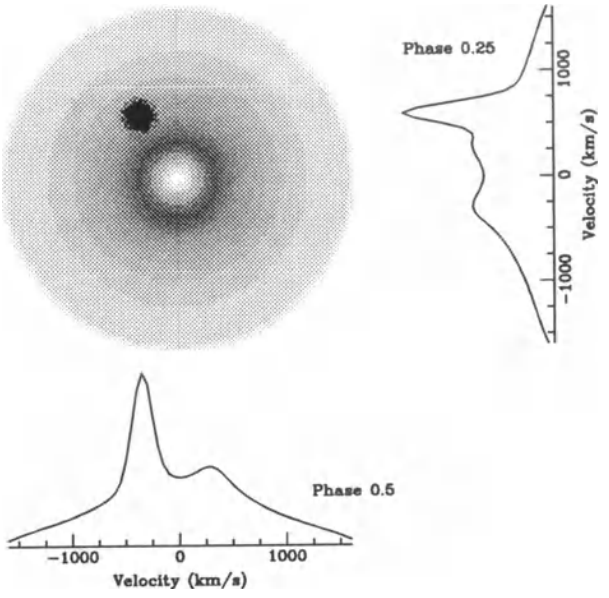


Figure 2: A simulated greyscale image of a binary in velocity coordinates and two line profiles plotted to show their formation as projections of the image.

close to the critical radius. The spot approximates the region where the gas stream hits the accretion disc and the critical radius is equivalent to the outer edge of the disc.

Having shown that the line profiles are projections of the desired image, the problem of finding the image is exactly the same as that faced when trying to reconstruct an image of the human head from a series of X-ray projections. The different projection angles used in an X-ray scanner are equivalent to different orbital phases and the opacity to X-rays translates to the emission line intensity.

3. Maximum Entropy Inversion

The problem of reconstruction of an image from a set of projections can be solved by linear methods (Rowland 1979). However the linear method, while fast, is difficult to modify for effects such as optically thick line emission. Further disadvantages of the linear inversion are the possibility of negative or complex data values and the propagation of statistical noise into the image. Instead we apply the maximum entropy method to carry out the inversion, using the FORTRAN code MEMSYS (Skilling & Bryan 1984).

The image is modelled by a polar grid of pixels. The radius of the image is matched to the highest velocities seen in the data (typically 2000 km s^{-1}), and the pixels are sufficiently small to match the spectral resolution and may typically be 50×50 km s^{-1} square. Predicted

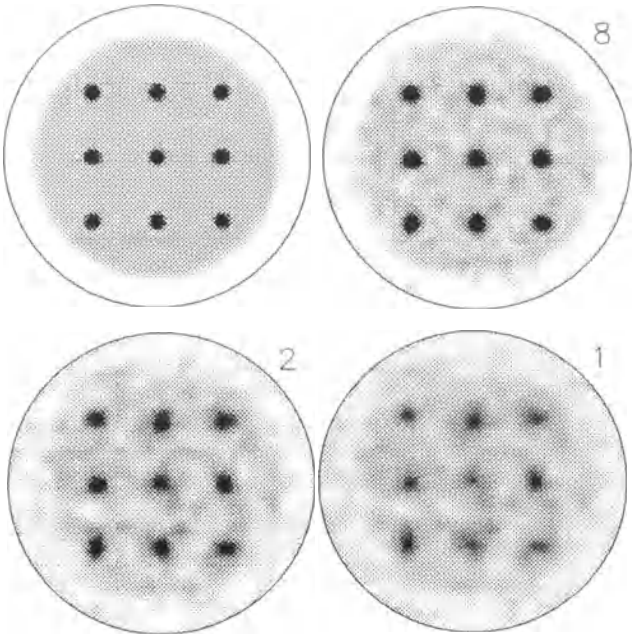


Figure 3: The test image and three reconstructions from noisy data using a uniform default image. The circles mark the edge of the image array.

data are computed from the image by applying the physical model of the line formation described in the previous section. The image is then adjusted to match the predicted with the true data to within a pre-set value of the χ^2 statistic. In general there are many such images, and of these we choose the one of maximum entropy, $S = \sum_i I_i - D_i - I_i \ln(I_i/D_i)$. The I_i are the image levels and D_i are the default image values. In the absence of any data constraints, the maximum of S occurs when the image I is the same as D. Therefore, in this definition of entropy, D is the default which I tends towards as the data become noisier.

To test the method we generated data at 40 orbital phases from an image consisting of nine spots on a uniform background (top left, figure 3). The following three images were then reconstructed from the data after the addition of increasing amounts of gaussian noise. Figure 4 shows the profile at phase zero for the data with and without noise to illustrate the amount added. The noise was generated with a pseudo-random number generator and the same seed integer was used so that the pattern of noise is identical for every reconstruction. The numbers next to the reconstructions are the signal-to-noise ratio in the continuum in each case. Figure 3, reconstructed with a uniform default image, shows that the important features of the image can be recovered, even with large amounts of added noise.

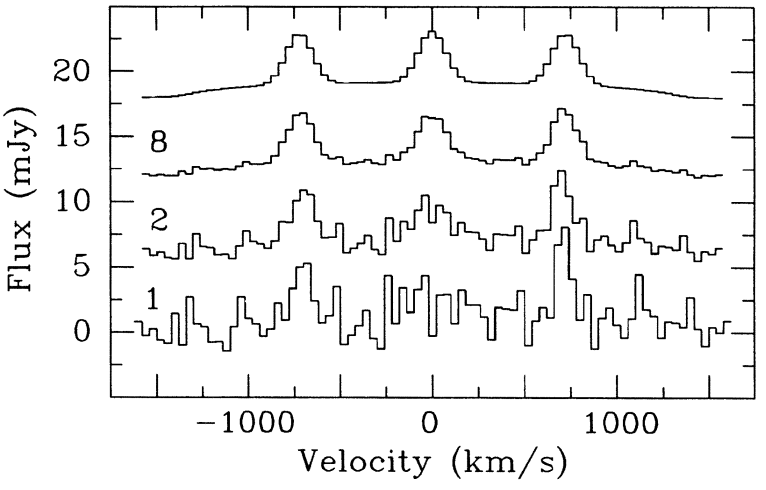


Figure 4: The profile at phase 0 for the model and each reconstruction.

4. The choice of default

The uniform default used for figure 3 pulls the outer regions of the reconstructions higher than in the original image, while pulling the peaks lower. The effect becomes more important as the noise level increases.

The emission line intensity as a function of radius from the centre of the image is a useful constraint on theories of emission line formation. As we have seen above, a uniform default biases the result towards a flat radial distribution. We can do better than this by modifying the default. Following Horne (1985) we consider iteratively defined defaults which are a function of the reconstruction.

Figure 5 shows the reconstructions with three different defaults. The default for image B was computed by taking the average of the reconstructed image at each radius. Such a definition removes the entropy constraint on the radial profile of the image while still acting on the azimuthal variations. The improved fit to the radial profile is seen in the outer part of the image compared to the equivalent lower left image in figure 3, however, this default introduces rings at the same radii as the spots.

If we compute the default from the median at each radius, a great improvement is seen (image C, figure 5). This is based on an initial guess that variations in real systems will be isolated spots and so their effect can be removed with a median. Azimuthal defaults all suffer from the anisotropy of the entropy constraint which constrains the image more in the azimuthal direction than the radial. This can cause artefacts which mimic expected real features in the disc. To avoid this, we consider a final example, with the default derived from convolution of the reconstruction with a gaussian. Such a default removes the entropy constraint from large scales in the image, while retaining it for short scale structure. Image D of figure 5 shows the effect of this. The gaussian we used had a full width half maximum 0.2 times the radius of the image.

The smoothed default captures some of the best features of the uniform and axi-

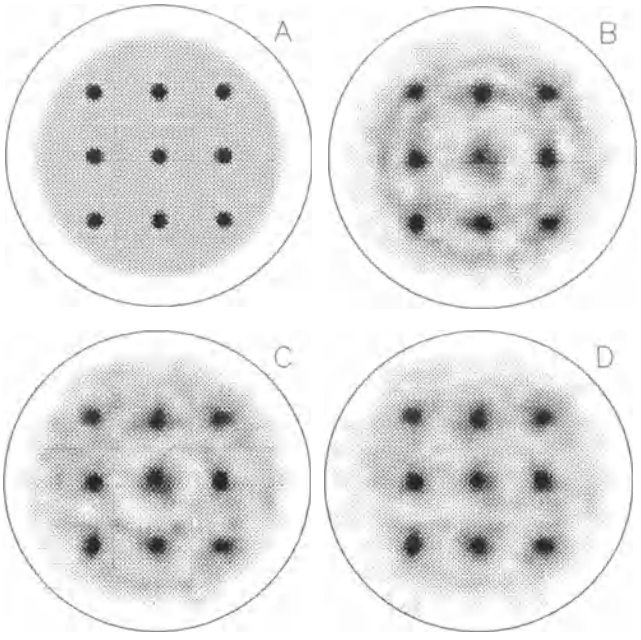


Figure 5: The test image (A), and reconstructions with (B) azimuthal average, (C) median and (D) gaussian smoothed default images for a signal-to-noise ratio = 2.

symmetric defaults. It represents the radial distribution more faithfully than the uniform default and does not produce the asymmetric smearing of the axi-symmetric default.

5. Results on Cataclysmic Variable stars

Images of the cataclysmic variable star, IP Peg are shown in figure 6. Images of the Hβ and Hγ Balmer lines of hydrogen are shown during an outburst on the left and during the normal faint state on the left. We have also drawn the predicted positions of the red star and the gas stream that falls from it.

The outburst images of IP Peg show no features connected with the gas stream, but there is strong emission from the red star caused by irradiation from the centre of the accretion disc. The quiescent images show a bright region near the region of impact of gas stream and disc. Further examination shows that this feature penetrates far into the disc, presumably as the gas stream coasts over the surface of the disc. We have examined the similar system WZ Sge (data kindly provided by Dr. Schlegel) and have found gas stream coasting here as well. Other unexplained features are the subject of current work.

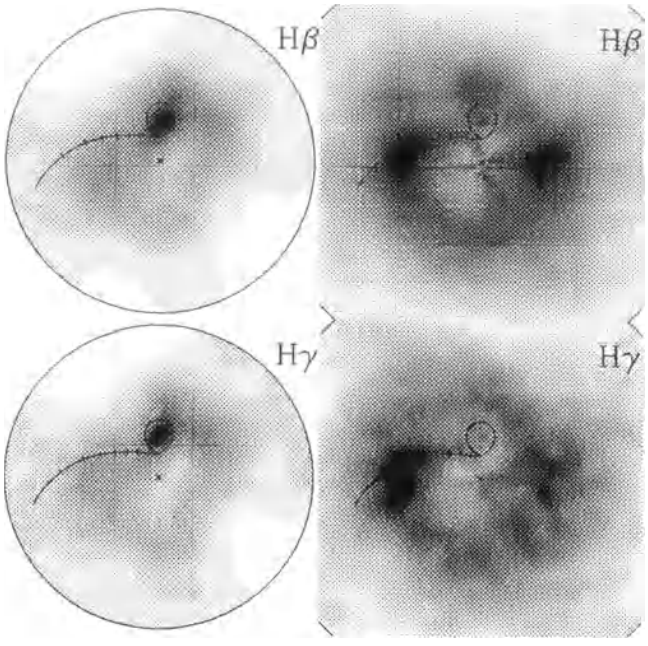


Figure 6: Images of IP Peg in outburst (on the right) and quiescence. The outburst images are $\approx$ 100 times brighter but have been rescaled for display purposes. The size of the quiescent images has been increased to give the same velocity scale.

6. Discussion

The maximum entropy method can be applied successfully to imaging the emission lines in close binary stars. For this to be successful, the Doppler broadening should dominate over any intrinsic line broadening. The short periods of cataclysmic variables ensure this condition, and typical velocities are $\approx 500\,\mathrm{km\,s^{-1}}$ compared to thermal broadening of $\approx 10\,\mathrm{km\,s^{-1}}$.

To avoid the bias caused by a uniform default, we applied the maximum entropy method with an adjustable default. However, one has to be careful to avoid introducing spurious features into the image, which for binary stars, led us to use gaussian convolution to compute the default.

References

Horne, K., 1985. *Mon. Not. R. astr. Soc.*, **213**, 129.

Rowland, S.W., 1979. *Image reconstruction from projections*, p9, ed. Herman, G.T., Springer-Verlag

Skilling, J. & Bryan, R.K., 1984. *Mon. Not. R. astr. Soc.*, **211**, 111.

Vogt, S.S., Penrod, G.D. & Hatzes, A.P., 1987. *Astrophys. J.*, **321**, 496.

DISTANCES TO CLUSTERS OF GALAXIES USING MAXIMUM ENTROPY

Ofer Lahav[1], Stephen F. Gull[2] and Donald Lynden-Bell[1]
1 Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK
2 Mullard Radio Astronomy Observatory, Cavendish Laboratory,
Madingley Road, Cambridge CB3 0HE, UK

ABSTRACT. We present a method of estimating distances to clusters of galaxies from two-dimensional catalogues by using the Maximum Entropy Method.

1. Introduction

A basic problem in astronomy is the deduction of a 3-dimensional distribution from a 2-dimensional distribution projected over the sky. The problem is well illustrated in the study of the large-scale galaxy distribution. Magnitude (or angular diameter) limited catalogues list the angular position of galaxies with a high precision almost over the entire celestial sphere. On the other hand, the information on the third dimension is very limited. The usual way of obtaining a 3-dimensional picture of the local universe is to carry out redshift surveys and to deduce distance from velocity by using Hubble's law. Such redshift surveys, however, cover as yet only a small fraction of the sky. Furthermore, the distribution of galaxies as seen in redshift surveys is distorted by local gravitational fields. Angular diameters (or magnitudes) of galaxies, which are listed in 2-D catalogues in a complete way, can be used as distance indicators as well. Astronomers have used the nth brightest galaxy (e.g. the first ranked or the tenth brightest) in a cluster as a standard candle and have deduced distances by fitting luminosities of cluster galaxies to the entire luminosity function. Here we study further the mapping from 2-D to 3-D by using a diameter function (or a luminosity function), which is deduced from a redshift survey at a section of the sky. We give a new solution to this inversion problem by using the Maximum Entropy (MaxEnt) Method.

2. The Inversion Problem

In a universe in which all galaxies have the same metric diameter D, the distance to each galaxy is simply (neglecting relativistic corrections) $r = D/\theta$, where θ is the apparent angular diameter. However, in our own universe there is a broad 'natural' distribution of galaxy metric diameters.

We define the diameter function $\phi(D)$ in analogy with the luminosity function, such that the number of galaxies per volume element d^3r and with a metric diameter D in

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 347–354.

the interval $(D, D+dD)$ is

$$dN = \frac{n(\mathbf{r})}{\langle n \rangle} d^3r \, \phi(D) dD \,, \tag{2.1}$$

where $\mathbf{n}(\mathbf{r})$ is the 'true' number density of galaxies at position $\mathbf{r}$, and $\langle n \rangle$ is the mean number density of galaxies in the universe. By writing eq. (2.1) in this form we assume that metric diameters are uncorrelated with spatial position and local density. In this work we assume that the diameter function is well-known and we seek a solution for $n(\mathbf{r})$.

We assume now that within a narrow cone of solid angle ω, $n(\mathbf{r}) = n(r)$, and consider $\hat{N}(\geq \theta)$, the expected number of galaxies with angular diameter greater than θ:

$$\hat{N}(\geq \theta) = \omega \int_0^\infty n(r) \, \varphi(r, \theta) \, r^2 \, dr \,, \tag{2.2}$$

where $\varphi(r, \theta)$ is a selection function which expresses the probability of finding a galaxy with an angular diameter $\geq \theta$ at a distance r. For a catalogue with a lower cutoff in angular diameter this function is:

$$\varphi(r, \theta) = \varphi(r\theta) = \frac{1}{\langle n \rangle} \int_{r\theta}^\infty \phi(D) dD \,, \tag{2.3}$$

where we neglect Galactic obscuration.

In a discrete form we write the relation as

$$\hat{N}_k \equiv \hat{N}(\geq \theta_k) = \sum_i n_i P_{ik} \,, \tag{2.4}$$

where n_i is the density at the ith distance bin and P_{ik} is our 'Point-Spread Function' (PSF):

$$P_{ik} = V_i \, \varphi(r_i \theta_k) \,, \tag{2.5}$$

where $V_i = \omega \, [(r_i + \Delta r)^3 - r_i^3] \, /3$ is the volume of the ith distance bin of thickness Δr.

Our task now is to find the density vector n_i given the counts vector $\hat{N}_k$. The deviations of the measurements N_k from the predictions (2.4) can be expressed in terms of the χ^2 statistic:

$$\chi^2(n) = \sum_k (N_k - \hat{N}_k)^2 / \sigma_k^2 \,, \tag{2.6}$$

where σ_k is the standard error on the data. A naive approach might be to set $\hat{N}_k = N_k$ and to invert the set of linear equations (2.4) directly to find n_i. However, an inversion of relations which involve noisy data is unstable and has no unique solution. We suggest instead using the Maximum Entropy Method.

3. Maximum Entropy Solution

Our inversion problem can be viewed as analogous to a problem in image processing. We wish to reconstruct the true radial density profile from a 'blurred' image. The blurring is caused by a large spread in the distribution of metric diameters. The PSF tells us the counts of angular diameters when all cluster galaxies are concentrated at one distance. This close analogy to the reconstruction of true images from distorted pictures suggests the application of a reconstruction technique like the MaxEnt Method to our problem.

To cope with the fact that the quantity n is unnormalized we adopt Skilling's generalisation of Shannon's entropy (Skilling & Gull 1988, in this volume):

$$S = \sum_i [n_i - m_i - n_i \, log(n_i/m_i)] \, , \tag{3.1}$$

where m_i is an initial model for n_i. When the image n_i matches the model perfectly the entropy is zero. For example one can take $m_i = \langle n \rangle$, i.e. that galaxies are distributed uniformly. The procedure now is to define a space of the dimension of the number of image cells and to maximize the entropy (3.1) under the constraint:

$$\chi^2 \leq C_{\rm aim} \, , \tag{3.2}$$

where $C_{\rm aim}$ is a constant defined in advance (see below); any image vector n_i with $\chi^2 > C_{\rm aim}$ is contradicted by the data. Maximizing the entropy alone will give us a uniform distribution (a 'grey map') in the case of an initial uniform model. Therefore, maximizing the entropy under the data constraint will give us the most conservative picture of deviations from uniformity allowed by the data.

As our MaxEnt algorithm we use the algorithm MEMSYS (Skilling & Bryan 1984). This algorithm iterates towards the image that maximizes the entropy under the constraint. A crucial question is how to choose $C_{\rm aim}$. We present the data as cumulative distributions, hence we are not allowed to set $C_{\rm aim}$ to be the number of bins because there are correlations between the bins. Instead we apply the following procedure, in which we run the algorithm twice. The first run ('pass 1') is used in order to find an empirical $C_{\rm aim}$, which is then used for the second run ('pass 2'). 'Pass 2' is the 'proper' MEMSYS run.

In 'pass 1' we set $\sigma_k^2 = N_k$ and choose an arbitrary very low value for $C_{\rm aim}$. We monitor the iterations by the following statistical test which aims to check whether our predicted number counts given in eq. (2.4), $\hat{N}_k$, *after each iteration* are not too different from the measurements N_k. After each iteration we compare the observed and predicted distributions by the Kolmogorov-Smirnov (KS) statistic, which tells us about the shape of the distributions. The smaller the KS probability $p_{\rm KS}$, the larger the difference between the two distributions. We also calculate a Poisson probability, $p_{\rm P}$, for the total number of galaxies in the cone, to verify that the amplitudes of the predicted and measured distributions agree. Since the KS and Poisson probabilities are independent we calculate the joint probability simply by

$$p = p_{\rm KS} \, p_{\rm P} . \tag{3.3}$$

In order to decide when to reject the null hypothesis (that the two distributions are the same) and therefore to continue the iterations we specify a significance level γ. The null hypothesis is rejected if $p \leq \gamma$, while the iterations are stopped if $p > \gamma$. We then record the 'final' χ^2 as our new $C_{\rm aim}$.

We then run the algorithm again ('pass 2') with the new $C_{\rm aim}$ as an input. In this pass we do not monitor the above probabilities but simply iterate many times. We only verify that after many iterations the algorithm has converged, i.e. χ^2 has reached $C_{\rm aim}$ and the MEMSYS parameter TEST (a dimensionless number measuring the angle between the gradients of χ^2 and the entropy S) is nearly zero. The only free parameter in our procedure is the significance level γ (or in fact $C_{\rm aim}$). We would like to emphasise that the final output of this procedure is a probability function for the density field along the line of sight, *not* the positions of individual galaxies.

4. Applications

As an example we have applied the method to galaxies from the diameter limited UGC catalogue (Nilson 1973).

The diameter function is well fitted by the analytic form (Lahav *et al.* 1988, Appendix A):

$$\phi(D)dD = \phi_* \, t^{-\mu} \, [1 + t/\nu]^{-\nu} \, \{\mu/t \, + \, [1 + t/\nu]^{-1}\} \, dt \, , \tag{4.1}$$

where $t \equiv (D/D_*)^2$ and ϕ_*, μ, ν and D_* are free parameters. The estimated parameters are $\mu = 0.16$, $\nu = 3.78$, and $D_* = 60.7 \; h^{-1}$Mpc· arcmin ($= 17.7 \; h^{-1}$ Kpc). Hereafter we specify distances in units of Mpc/h ($H_0 = 100 \; h$ km s^{-1} Mpc^{-1}), or alternatively in velocity units of 100 km s^{-1}. Our PSF is therefore:

$$P_{ik} = t_{\min}^{\mu} \, [1 + t_{\min}/\nu]^{\nu} \, V_i \, t_{ik}^{-\mu} \, [1 + t_{ik}/\nu]^{-\nu} \, , \tag{4.2}$$

where $t_{ik} = (r_i \, \theta_k \, / D_*)^2$ and $t_{\min} = (D_{\min}/D_*)^2$

We bin the data in a cumulative way such that the bin boundaries are defined according to the θ's of the data. This binning extracts maximum information from the data. The distance r is binned in steps of 1 Mpc/h and covers the range 0-200 Mpc/h. We extrapolate the diameter function down to $D_{\min} = 1$ Mpc/h· arcmin. For our initial model we adopt as a fiducial value $m_i = \langle n \rangle = 0.015$ galaxies per (Mpc/h)^3.

We now direct our 'Maximum Entropy telescope' towards the Virgo cluster (l = 284°; b = 74°). We perform number counts for all galaxies within 6° of Virgo's centre which have major diameter $\theta \geq 1$ arcmin. The number counts are shown as dots in Fig. 1a. There are 320 galaxies in 53 cumulative bins. For a uniform distribution the expected slope in such a logarithmic plot is -3. In Fig. 1b we show results for 3 values of the significance level γ (our free parameter). As γ is increased from 0.1 (a conservative value) the density structure becomes more detailed, and by $\gamma = 0.7$ (a very liberal value), 3 peaks are apparent at 9 , 21 and 63 Mpc/h. In all the above cases the parameter TEST is very small (of the order of 10^{-5}), therefore indicating good convergence to the unique MaxEnt solution. In Fig. 1a we show the reconstruction of number counts for each of the 3 density profiles. The case $\gamma = 0.1$ does not fit the data well, whilst the other two cases show good fits. The distance to Virgo is estimated to be about 12 Mpc/h (see e.g. Tammann 1987) . Therefore, the $\gamma = 0.5$ case gives a reasonable answer (13 Mpc/h). The density profile corresponding to $\gamma = 0.7$ is compatible as well, at least qualitatively, with other studies of the Virgo cluster. We interpret the peak at 63 Mpc/h as a background cluster, A1367. While the position of the peaks has a simple interpretation, the meaning of the amplitude and breadth of the density bumps is less trivial. We currently use the method mainly for the identification of peaks.

Another application of our MaxEnt Method is to find rough estimates of distances to clusters in new deep 2-D catalogues. The Cambridge Automatic Plate Measuring (APM) machine has been used by Maddox *et al.* (1988) to produce a deep catalogue of several million galaxies from the UK Schmidt Southern Sky Survey. Our algorithm finds the distance of a cluster in less than few minutes CPU time (on a VAX), so it is an efficient way of getting some knowledge of the distance, We have explored that possibility by applying our procedure to APM clusters with known redshift. Redshifts to 14 APM clusters have been measured by Colless (1987). For the APM galaxies we have used magnitudes instead of diameters as distance indicators and a Schechter luminosity function. While in most statistical applications the significance level should be defined in advance of the experiment, we use it here as a control parameter. We use the 14 clusters as calibrators to tune our procedure in order to give a good fit of predicted to observed number counts, as well as 'correct' distances. We find the value $\gamma = 0.1$ as an

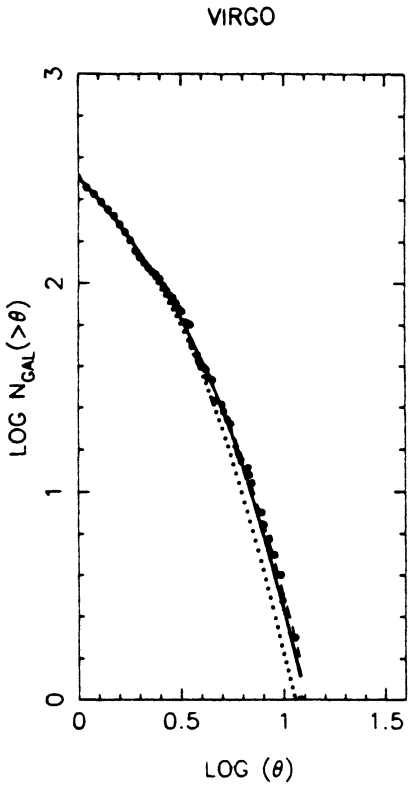


Figure 1a . Number counts as a function of angular diameter for the Virgo cluster (6° in radius). The dots are the measurements and the dotted, solid and dashed lines represent the reconstruction (according to the density profiles in Fig. 1b) corresponding to significance level of 0.1, 0.5 and 0.7, respectively. The slope expected for a uniform distribution is −3.

optimal one. Fig. 2 shows our estimates r_{MEM} versus the redshift distance r_Z for the 14 clusters. Our distance estimates r_{MEM} deviate by $2-40\%$ from the redshift estimates r_Z. The new method slightly improves the distance estimate in comparison with the 10-*th* brightest galaxy estimator. While for some of the clusters we get good number counts fitting as well as a 'correct' distance for $\gamma = 0.1$, we find a trend in other clusters for *lower* prediction of number counts at the bright-end compared with what is observed. If we increase γ we get a better fit but (in some cases) a wrong distance. It is difficult to find a 'universal' behaviour of all clusters. A possible explanation is that some of the bright galaxies do not fit the 'universal' luminosity function, either because there are large cD galaxies or because there are line-of-sight overlaps (galaxy-galaxy or galaxy-star). Clearly, a better classification of the APM bright galaxies is required.

5. An Alternative Approach

As an alternative to the KS-Poisson test for finding $C_{\rm aim}$ we are currently developing a new procedure, the 'classic' MaxEnt (Gull & Skilling 1988, in this volume). The idea

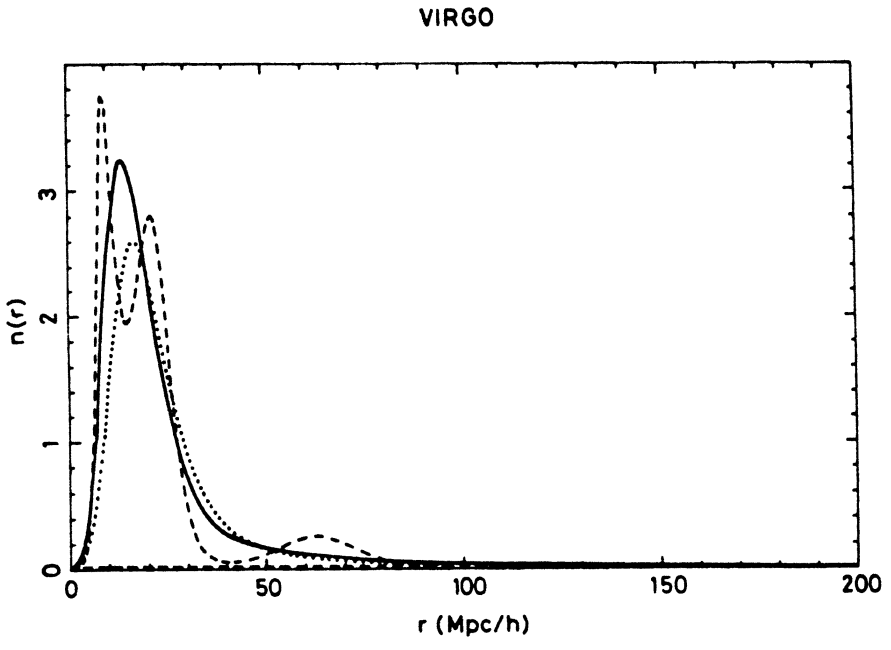


Figure 1b . The radial density profile towards Virgo as deduced by our MaxEnt algorithm. The dotted, solid and dashed lines correspond to significance level of 0.1, 0.5 and 0.7, respectively. The horizontal dashed line at $n = 0.015$ $(\mathrm{Mpc}/h)^{-3}$ is the initial model. Note that there is a great excess of galaxies in this direction.

here is to fix the Lagrange-multiplier α, which measures the weight of the entropy S relative to the log-likelihood function $-L$. It turns out from Bayesian arguments that the value of α can be fixed by the relation

$$-2\,\alpha\,S(\hat{n},m) = \sum_j \frac{\lambda_j}{\lambda_j+\alpha}\,, \tag{5.1}$$

where $\hat{n}$ is the 'best' image and m is the initial model. The λ_j's are the eigenvalues of a matrix which involves $\nabla\nabla L$ and the image n. Since $\lambda_j > \alpha$ represents a 'good' observation, $\sum_j \frac{\lambda_j}{\lambda_j+\alpha} \approx N_{\mathrm{good}}$, the number of 'good' observations. Therefore, the 'best' Lagrange multiplier is $\hat{\alpha} \approx -N_{\mathrm{good}}/(2S)$. In principle the right-hand-side of eq. (5.1) can be calculated explicitly, but as a simple estimate we guess N_{good} . This is possible for our case because a Singular Value Decomposition of the PSF matrix shows that the eigenvalues are uniformly spaced in the logarithm, hence for a wide range of problems we expect $N_{\mathrm{good}} \approx 4$.

The likelihood in this procedure is taken to be

$$\exp(-L) \propto \prod_k \exp(-\hat{N}_k)\,\frac{(\hat{N}_k)^{N_k}}{(N_k)!}, \tag{5.2}$$

where the $\hat{N}_k$ and N_k are the predicted and measured *differential* counts, respectively. The gradients of L are evaluated and incorporated directly into an α-controlled variant of MEMSYS.

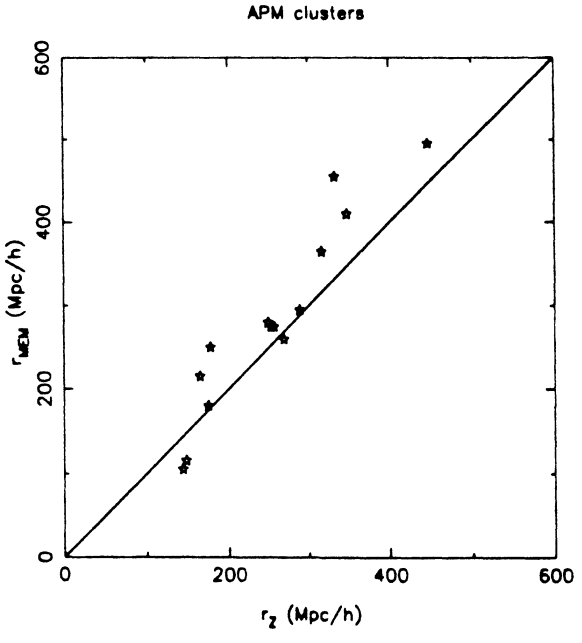


Figure 2 . Distances to 14 APM clusters as estimated by our Maximum Entropy Method versus redshift estimates of Colless (1987).

In Fig. 3 we show Virgo's profile as deduced by this method (using diameters) for $N_{good} \approx 4$. Note the similarity to the results shown in Fig. 1b. If we take N_{good} to be a factor 2 larger or smaller we are getting profiles which are more 'conservative' or more 'liberal' respectively. This method is attractive since it involves no free parameters and does not depend on the binning of the data.

6. Discussion

We have presented a new non-parametric method of estimating radial density profiles from magnitude/diameter limited catalogues. Currently our method is useful for identifying peaks along the line of sight. It is important to know more about the errors in the estimation of the number counts, particularly at the bright-end. Another important issue is the universality of the luminosity function. Even if the data were free of noise, a variation in the luminosity function from cluster to cluster would be expected.

Simple modifications to our procedure are possible, for example, changing the cone's radius. Here the trade-off is to keep the high angular resolution provided by the catalogues with the need to have a large number of galaxies in a cone for the statistics. Another modification is to split the sample according to morphological types. That might give a narrower Point-Spread Function for each morphological type, but would decrease the number of galaxies in each cone.

More fundamental modifications are required on both the statistical and astronomical aspects of the problem. On the statistical side, we intend to use the formalism

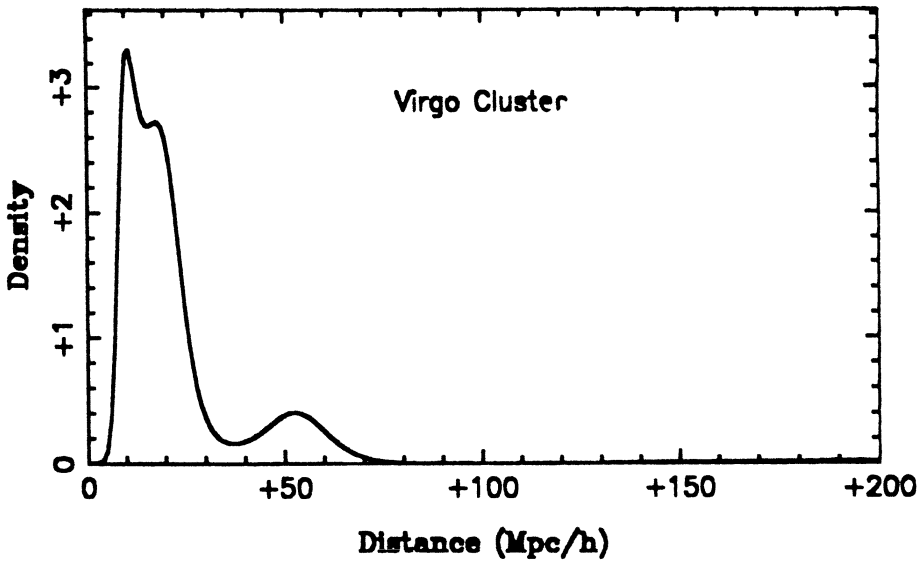


Figure 8 . The density profile towards Virgo, as deduced by the '$2\alpha S$' method for $N_{good} \approx 4$. Compare with Fig. 1b.

which fixes the Lagrange-multiplier $\hat{\alpha}$. On the astronomical side, it is important to find a better distance indicator. Furthermore, even the most accurate distance indicators (e.g. Faber-Jackson and Tully-Fisher relations) are in fact distribution functions. The MaxEnt Method might be useful in converting these narrow distributions functions into an unbiased distance estimators.

Acknowledgements. We thank M. Colless, G. Efstathiou, N. Kaiser and S. Maddox for helpful discussions.

References

Colless, M., 1987. Ph.D. thesis, Cambridge University.

Gull, S.F. and Skilling, J. & , 1988. in *Maximum Entropy and Bayesian Methods, Cambridge 1988*, ed. J. Skilling, Kluwer Academic Publishers, Dordrecht.

Lahav, O., Rowan-Robinson, M. & Lynden-Bell, 1988. *Mon. Not. R. astr. Soc.* , in press.

Maddox, S.J., Efstathiou, G., & Loveday, J. 1988. in *Large Scale Structures of the Universe*, IAU Symp. No. 130, eds. J. Audouze, M.C. Pelletan, & A. Szalay, Reidel, Dordrecht.

Nilson, P., 1973. *Uppsala General Catalogue of Galaxies*, Uppsala astr. Obs. ann., **6**.

Skilling, J. & Gull, S.F., 1988. in *Maximum Entropy and Bayesian Methods, Cambridge 1988*, ed. J. Skilling, Kluwer Academic Publishers, Dordrecht.

Skilling, J., & Bryan, R.K., 1984. *Mon. Not. R. astr. Soc.* , **211**, 111.

Tammann, G.A., 1987. in IAU Symp. No. 124, *Observational Cosmology*, eds. A. Hewitt *et al.*, Reidel, Dordrecht.

THE APPLICATION OF MAXIMUM ENTROPY TECHNIQUES TO CHOPPED ASTRONOMICAL INFRARED DATA[1]

C. Burrows[2], J. Koornneef[2]
Space Telescope Science Institute
3700 San Martin Drive
Baltimore, Maryland 21218, U.S.A.

ABSTRACT. We discuss the application of maximum entropy techniques to chopped astronomical infrared data. The resulting maps are much better than those obtained with the only existing published algorithm. Nevertheless, we feel that the technique requires further improvement before it can yield all the information inherent in such data. In the course of our investigation of the method, an interesting limit of the maximum entropy solution was discovered. This enables a solution of the maximum entropy equations for a large class of problems. The resulting solution illustrates the freezing of degrees of freedom discussed elsewhere in these proceedings. Finally, we discuss an apparently attractive method for assessing the errors on a reconstruction, and why we feel that such a formalism is incomplete.

1. Introduction

If one observes the northern winter night sky in the region where the equatorial plane meets the galactic plane one is looking at the Serpens-Ophuchius giant molecular cloud. In this region, there is evidence for extensive dust and gas, Herbig Haro emission line objects, bipolar molecular outflows, and heavily reddened embedded sources. All of these phenomena are frequently associated with recent or ongoing star formation. One particular region of the cloud that shows evidence for extended optical emission was mapped in the near infrared by Churchwell and Koornneef[1]. They identified a large number of embedded point sources, many of which are not evident on deep CCD frames at visible wavelengths. Greater optical depth through the cloud implies a source is less visible at shorter wavelengths— it is reddened. The least reddened source IRS3 is apparently nearest to the surface of the cloud. It seems

[1] Based on data obtained at the European Southern Observatory.

[2] Affiliated with the Astrophysics Division, Space Science Department of the European Space Agency

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 355–362.

to be associated with the extended optical emission that originally drew attention to the region and appears to be the center of a bipolar molecular outflow. The working hypothesis is therefore that this source represents a star near the surface of the cloud in the process of formation.

We decided to obtain much more detailed infrared maps of IRS 3 in order to understand the near infrared morphology and photometric properties of the source and its environs. We describe here some of the techniques that we used to reduce the maps, and discuss the problems that we encountered.

A second point of this contribution is to describe an interesting limit of the maximum entropy equations in which the solution reduces to a linear problem. This illustrates many of the new properties of the full solution discussed at this workshop, and gives a concrete example of how the tradeoff between prior and posterior knowledge is made. Finally we make some comments concerning the possible estimation of the errors associated with a reconstruction.

2. Summary of Astronomical Data

A fundamental problem in making infrared photometric measurements in astronomy is the presence of a large background, from the sky and telescope. The classical technique for overcoming this is to rapidly move the field of view of the telescope between the object of interest and a nearby reference point on the sky. This can be done for example by wobbling the telescope secondary mirror through a suitable 'chopping distance.' A phase sensitive amplifier then gives the flux difference and effectively removes the background. This works well for point objects but fails for extended sources, because nearby points with no intrinsic flux are then not available.

The solution generally adopted is to map a region in a rectangular grid, and thus obtain a map of differences. A point source would appear first as a positive contribution, when it appears in the positive beam or position of the secondary mirror. Then as the telescope is moved, it will appear in the negative beam. The effect is that when the system is suitably set up and calibrated, the point spread function (PSF) consists of a positive lobe separated by the chopping distance from a symmetrical equal negative lobe. Such a PSF convolved with a uniform background produces zero in the observed data. Hence, the method is insensitive to the background present. This fact is one of the crucial properties of such data, and can be thought of as a generic property of data with a zero volume point spread function.

In Figure 1 is a montage of our near infrared maps of the source, together with part of a red CCD image obtained at the Palomar 200 inch telescope by Bel Campbell. It can be seen that point sources such as number 9 contribute approximately equal positive and negative lobes. The bright, complex and extended structure around source 3 is our main interest.

In Figure 2 is the K band scan line through the center of source 3. Each point has an associated error bar which was obtained experimentally by repeating the measurement several times. The small estimated systematic errors are added to such error bars in quadrature to produce an overall error estimate for each data point. It can be easily seen that the data has large dynamic range, and that the errors are largest when the data or particularly its slope is large.

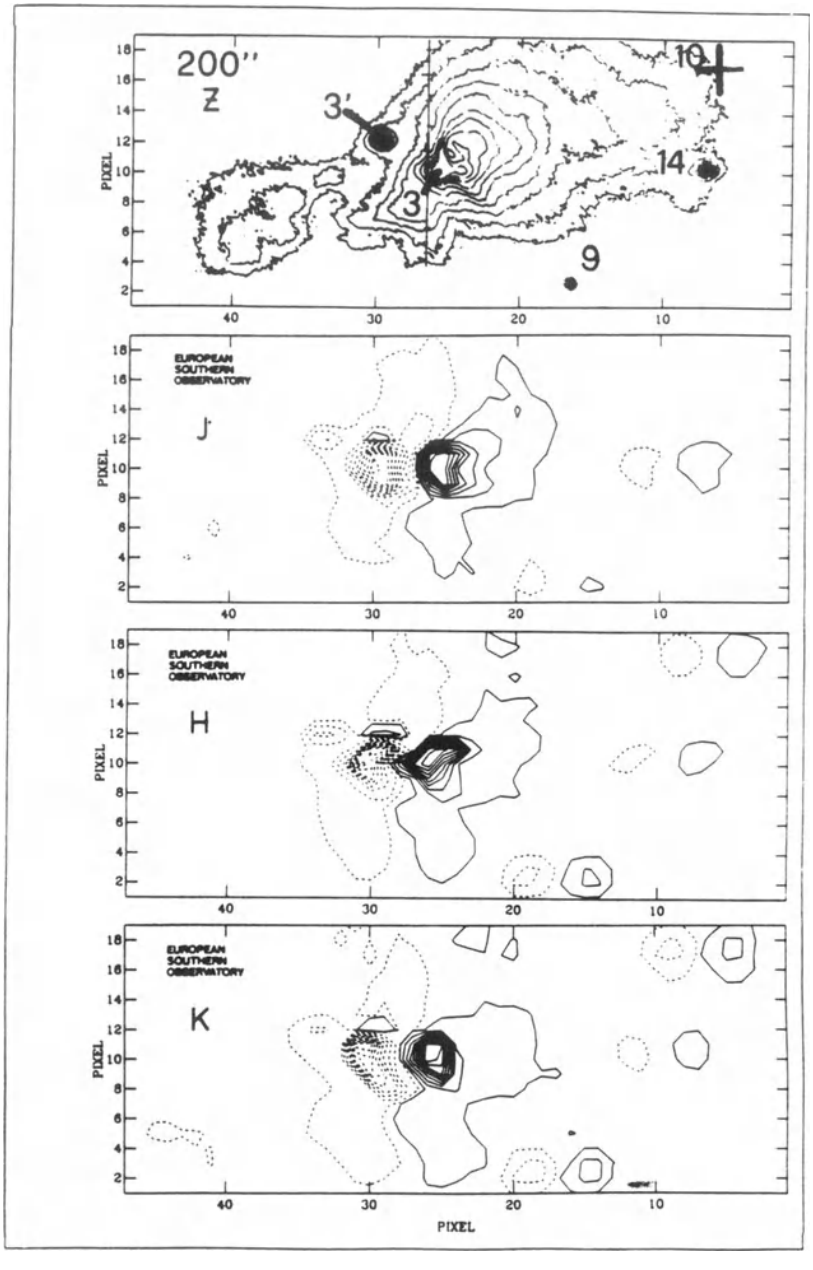


Figure 1. The top panel shows a Gunn z CCD image due to Bel Campbell. The scale is 2 arcseconds per pixel. The three remaining panels show our 3.6m ESO/InSB observations in their original (chopped) form. The lowest contours are at 3, 2 and 0.8% of the peak signal for J, H and K, respectively.

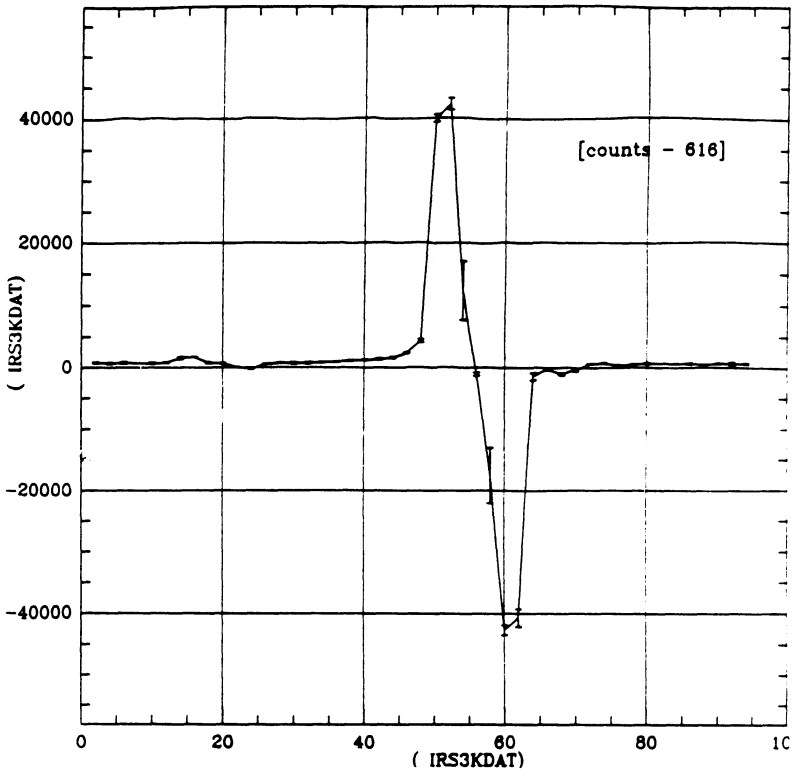


Figure 2. The data for the scan line through the center of the K band map.

The only published method [2] for reducing the data does not work very well because noise tends to propagate across the image. The details of the extended emission from source 3 are lost in noise that propagates from the bright peak. We were therefore attracted to try the maximum entropy approach because it should tend to damp such oscillations, and will automatically ensure that the derived fluxes are positive.

3. Limit of Maximum Entropy Solution for Undetermined Background

We write the log likelihood as proportional to

$$W = -\sum_j \frac{O_j}{s} \log \frac{O_j}{s} + \lambda \chi^2 \quad . \tag{1}$$

Where, O_j are the unknown fluxes at pixel j, and j labels the two dimensional coordinate suitably. S is the sum of the O_j and χ^2 is the sum of the normalised residuals squared. Maximizing this function with respect to the O_j gives the formal solution

$$O_j = S \; exp\left\{ \sum_k \frac{O_k}{S} \log \frac{O_k}{S} - \lambda S \frac{\partial \chi^2}{\partial O_j} \right\} \quad . \tag{2}$$

We attempted to solve the above equation iteratively, by starting with a flat image, inserting in the right hand side to get a new estimate and iterating. The variable λ was gradually increased at the same time to implement the constraint that χ^2 tend to its expectation value, N the number of data points. This development follows that of Gull and Daniell [3]. Unfortunately, in this case, the iteration does not converge. The iterations exhibit oscillatory divergences, and the total flux becomes large. The oscillations can be largely controlled by averaging successive iterates before repeating the iteration. The variable S, the total flux is not constrained by the data, and tends to larger and larger values in order to make the relative fluctuations about a flat image as small as possible. Given a set of O_j, that imply a certain χ^2, the image $O_j + C$ for any positive constant C has higher entropy, and because the PSF has zero volume, the same χ^2.

Faced with this situation, in which a solution to the maximum entropy problem does not formally exist, we can take two possible approaches. The first is to constrain the total flux to some particular fixed value, perhaps based on some prior information. The second is to investigate the limit of the solution as S becomes large, and ask what the fluctuations about the mean value tend to. The second approach is equivalent to the first if in that case, the constraint imposes a large total flux relative to the size of the fluctuations in the solution. This might well be the case for thermal infrared maps for example, where the sky background can swamp the source fluctuations.

In the reconstruction of radio maps by maximum entropy methods, the background is also undetermined, and frequently it is made as small as can be reasonably imposed. This arbitrary choice leads to good reconstructions, but seems to have little theoretical basis. Similarly, we have found that arbitrarily imposing a small background leads to better results, but we have no justification for this, and indeed our prior information will frequently be that the background is large.

Therefore, we initially investigated the solutions that are obtained if the background is large, or the total flux is unconstrained but only the fluctuations are solved for. In this case, we can put $O_i = S/M + \delta_i$, and expand the above expression for $\delta_i << S/M$, where M is the number of points in the image to be determined. The lowest order terms cancel, and the result is

$$\delta_j = -\lambda \sum_k P_{jk} \left(\frac{\sum_l P_{lk}\delta_l - I_k}{T_k} \right)^2 \quad , \qquad 3$$

where the arbitrary constant λ has been replaced by $M\lambda/s^2$ and P_{jk} is the contribution of image point O_k to data point I_j which has variance σ_j^2. This equation is linear unlike the maximum entropy expression from which it is derived. It does give the solution essentially exactly of a well posed maximum entropy problem, and as such provides an interesting and solvable model for the general case. The equation illustrates the bias inherent in maximum entropy solutions. In this case, the solution is proportional to the normalised residuals convolved with the PSF.

As an aside, Equation 3 gives a Weiner filter solution in the particular case of a position independent point spread function, and data with additive white noise. In that case, the convolutions turn into products in Fourier space, and the solution is

$$\tilde{\delta}(\omega) = \frac{\tilde{I}(\omega)}{\tilde{P}(\omega)}\left[1 + \frac{\sigma^2}{\lambda \mid \tilde{P}(\omega) \mid^2}\right]^{-1} \quad . \tag{4}$$

In the above it is clear that when $\tilde{P}(\omega) >> \sigma^2/\lambda$ the Fourier inverse is obtained, but noise amplification is avoided for small $\tilde{P}$ by the modification to the Fourier inverse given in the square brackets.

Returning to the general case, we can write Equation 3 in the form

$$\underline{\delta} = -\lambda(\underline{\underline{A}}\underline{\delta} - \underline{B}) \quad , \tag{5}$$

where

$$\underline{\underline{A}}_{ij} = \sum_k \frac{P_{ik}P_{jk}}{\sigma_k^2} \quad \underline{B}_j = \sum_k \frac{I_k}{T_k^2}P_{jk} \quad . \tag{6}$$

A is a positive definite real symmetric matrix, with at least one zero eigenvalue. Expanding everything in terms of the complete orthonormal basis set x^α, we get $\delta_i = \sum_\alpha \delta_\alpha x_i^\alpha$ and so on, and the formal solution is given by

$$\delta^\alpha = \frac{\lambda}{1 + \lambda\lambda^\alpha}B^\alpha \quad . \tag{7}$$

By substituting this solution in the expression for χ^2, one can show that

$$\chi^2 = \sum_\alpha \frac{B_\alpha^2}{\lambda^\alpha}\left[\frac{1}{(1 + \lambda\lambda^\alpha)^2}\right] \quad . \tag{8}$$

The term in the square brackets is zero if $\lambda << 1/\lambda^\alpha$ and is one if the converse holds. Thus when λ is chosen, χ^2 receives contributions from important degrees of freedom. Gull and Siblisi in these proceedings have shown separately that a posterior canonical choice for λ exists, and that it measures the effective number of degrees of freedom in the data. This example illustrates the close relationship between the choice of lambda and the effective number of degrees of freedom, but leaves the choice of that number open.

This formal solution also illustrates why convergence difficulties can be encountered with the iterative solution. When $\lambda\lambda^\alpha$ is less than unity, the iteration converges. When λ is increased in order to decrease χ^2, the coefficients of the eigenvectors that violate this condition oscillate divergently about the correct solution. Averaging the successive iterates, and starting with a good initial solution reduces but does not entirely solve the problem.

4. Results and discussion

The convergence difficulties mentioned above are of a numerical nature, and do not reflect on the existence or uniqueness of the solution once found. Indeed, for values

of $\chi^2 = 3N$, we were able to iteratively solve the equations by a combination of the methods described above.

Unfortunately, the resulting maps are not good representations of the expected reality, although many correct features are present. The essential problem is that because the reconstructed image is superimposed on a very high background, the positivity constraint is not affecting the results. The solutions exhibit oscillations that propagate across the image with only gradually decreasing amplitude away from the bright sources.

We have made considerable progress in solving this problem by arbitrarily constraining the total flux to a small value to force positivity, replacing the point spread function with a narrower pair of positive and negative lobes to reduce the ringing caused by superresolution, by using Skilling's MEM software to improve convergence and by changing the prior or default image to reflect our preference for isolated sources on a more uniform background. For the last modification, the method adopted was to use a Clean algorithm followed by Gaussian blurring to produce a default map. Unfortunately, this of course means that many arbitrary steps and choices are made on the way to constructing the default map. The stopping point of the Clean algorithm, blur diameter, background, and PSF width are all free parameters. Further, the method chosen to construct the default is itself arbitrary. For example the results of a MEM reconstruction could be used instead of Clean, or the blurring applied could be position or flux dependent.

The result is that very reasonable maps have been constructed. We have been able to unambiguously determine that IRS3 is extended in the near infrared, and to identify some faint point sources in the data that Churchwell and Koornneef did not find. However, it is clear that a tradeoff is possible for example between assigned flux in a central point like source for IRS3, and the flux from its environs, for example by choosing different default images. We can proceed no further until a quantitative measure of the inherent errors is available.

In maximum likelihood estimation, one can estimate the covariance of the estimated parameters from the inverse of the matrix of second derivatives of the likelihood function. Roughly speaking, the width of the likelihood function defines the possible range for the parameters. Elsewhere in these proceedings, Skilling proposes to use this approach to estimate the errors in an image reconstruction. We are concerned that this approach will tend to underestimate the errors, because it does not take into account the uncertainty in the prior information used. This information has been variously described as the measure on the parameter space, and as the default image. Different choices of priors can lead to markedly different results, even if the uncertainty in the reconstruction for a given prior is small. To take an extreme example, if the prior happens to provide an acceptable fit to the data, the resulting reconstruction will be equal to the prior. The uncertainty estimate would be essentially zero, and yet any prior consistent with the data could have been chosen. When we get into the business of improving the map by changing the prior, we are getting dangerously close to this limit.

No multicolour calibrated near infrared maps of regions such as IRS 3 existed at the time of our measurements. Since then the use of staring infrared arrays has

become more widespread, and such maps have been produced. However, the staring arrays presently do not operate effectively in the thermal infrared, particularly in the atmospheric windows at 10 and 20 microns. Chopping and synchronised readout will be an essential technique to overcome this limitation. Further, the techniques described here will be useful in order to produce photometrically accurate maps with arrays even in the near infrared, in cases where the source contrast against the background is low.

References

1. Churchwell, E., and Koornneef, J. 1986 *Astrophys. J.,* **300**, 729.
2. Simon T. 1976 *Astron. J.,* **81**, 136.
3. Gull, S. F., and Daniell, G. J. 1978 *Nature*, **272**, 686.

THE USE OF BAYESIAN AND ENTROPIC METHODS IN NEURAL NETWORK THEORY

S. P. Luttrell
Royal Signals and Radar Establishment
St. Andrews Road
Malvern, WORCS.
WR14 3PS
U.K.

ABSTRACT. There has been much interest recently in the use of neural networks to solve complicated information processing problems such as those which arise in signal and image processing. In this paper we review Markov random field (MRF) neural network techniques for representing joint probability density functions (PDF). The "Boltzmann machine" serves as the paradigm, and we present a generalised version of its learning algorithm. We also present a technique for designing MRF potentials with low information redundancy for modelling image texture. To improve further the computational efficiency of such neural networks we introduce a novel method of cluster decomposing a PDF by using topographic mappings. The outcome of this programme is a means of designing sampling functions for extracting information from datasets (typically images).

1. Introduction

The image processing community has shown much interest in the use of Markov random field models to describe probability density functions for use in Bayesian image reconstruction schemes [Geman and Geman 1984, Geman and Graffigne 1987]. If we denote the field state as $\mathbf{x}$ and the PDF over states as $P(\mathbf{x})$ then an MRF is defined by a consistent set of conditional PDFs (called characteristics) amongst the components of $\mathbf{x}$. It follows from the Hammersley–Clifford theorem that corresponding to each consistently defined MRF there is an equivalent Gibbs distribution [Besag 1974, Kindermann and Snell 1980, Preston 1974], so $P(\mathbf{x})$ may be written as

$$P(\mathbf{x}) = \frac{1}{Z}\exp\left[-\mathbf{k}.\mathbf{s}(\mathbf{x})\right] \tag{1}$$

where $\mathbf{s}(\mathbf{x})$ is a vector potential, $\mathbf{k}$ is a vector of coefficients, and Z is a partition function. We use an unconventional symbol $\mathbf{s}$ to denote the potential because it is in fact a set of sampling functions of $\mathbf{x}$. Equation (1) defines a $P(\mathbf{x})$ from which samples $\mathbf{x}$ may be drawn by using a Monte Carlo scheme such as the Metropolis algorithm [Metropolis et al, 1953] or some variant thereof.

$P(\mathbf{x})$ is, of course, the maximum entropy PDF (with a uniform prior) which is consistent with the set of constraints $<\mathbf{s}(\mathbf{x})>=\mathbf{s}_0$, where $<..>$ denotes an average

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 363–370.

over $\mathbf{x}$ [Jaynes 1957, 1968, 1982]. Note that $P(\mathbf{x})$ has the form given in equation (1) if, and only if, the functional derivative $\delta H/\delta P(\mathbf{x})$ lies in the function subspace spanned by the vector of functional derivatives $\delta\langle \mathbf{s}(\mathbf{x})\rangle/\delta P(\mathbf{x})$, where H denotes the entropy of $P(\mathbf{x})$.

The purpose of this paper is to extend the above MRF scheme by introducing a greater degree of adaptability into the model. Thus in section 2.1 we shall explain how the Boltzmann machine neural network (and its generalisations) can be used to learn MRF models adaptively, and in section 2.2 we shall explain how economical MRF models of image texture can be constructed. In section 3.1 we shall introduce a novel form of multilayer neural network which allows maximum entropy reconstructions of the input PDF to be constructed with minimal computational effort, and in section 3.2 we shall explain how topographic mappings can be used to implement the layer to layer transformations in such a network.

2. G–maximisation models

Equation (1) is inflexible because $\mathbf{s}(\mathbf{x})$ must be selected by hand: there is no means of deriving $\mathbf{s}(\mathbf{x})$ adaptively from a training set of samples $\mathbf{x}$ following some observed PDF $P_o(\mathbf{x})$ ($\neq P(\mathbf{x})$ in general). In order to acquire $\mathbf{s}(\mathbf{x})$ adaptively we need a measure of the similarity of the (true) observed PDF $P_o(\mathbf{x})$ and the (maximum entropy) hypothesised PDF $P(\mathbf{x})$ defined in equation (1). Define the relative entropy G

$$G \equiv -\int d\mathbf{x}\, P_o(\mathbf{x}) \log\left[\frac{P_o(\mathbf{x})}{P(\mathbf{x})}\right] \tag{2}$$

Assuming base 2 logarithms, 2^{nG} is the probability that the hypothesised $P(\mathbf{x})$ will generate high probability n–sample sequences of states $\mathbf{x}$ which belong to the high probability set generated by the true $P_o(\mathbf{x})$, where n is asympotically large. Note that $G \leq 0$, with $G=0$ iff $P(\mathbf{x})=P_o(\mathbf{x})$. We shall deal with two types of adaptation in sections 2.1 and 2.2.

2.1. THE BOLTZMANN MACHINE

For a fixed set of potentials $\mathbf{s}(\mathbf{x})$ we may optimise $\mathbf{k}$ by G–maximisation using

$$\frac{\partial G}{\partial k_i} = \int d\mathbf{x}\, \frac{P_o(\mathbf{x})}{P(\mathbf{x})}\, \frac{\partial P(\mathbf{x})}{\partial k_i} = \left\langle s_i(\mathbf{x})\right\rangle_{P(\mathbf{x})} - \left\langle s_i(\mathbf{x})\right\rangle_{P_o(\mathbf{x})} \tag{3}$$

$\partial G/\partial k_i = 0$ when the constraints $\langle \mathbf{s}(\mathbf{x})\rangle = \mathbf{s}_o$ are satisfied, so hill–climbing G in $\mathbf{k}$–space yields the required maximum entropy PDF estimate. The first term in equation (3) is estimated from Monte Carlo samples of $P(\mathbf{x})$ defined in equation (1), whereas the second term is estimated from the training set which implicitly defines $P_o(\mathbf{x})$.

We may construct a version of equation (3) in which $\mathbf{s}(\mathbf{x})$ itself is effectively

learnt at the same time as the coefficient vector **k**. Thus we introduce hidden variables **h** by augmenting **s**(**x**) to become **s**(**x**,**h**). Denoting the associated Gibbs distribution as P(**x**,**h**) leads to

$$P(\mathbf{x}) = \frac{1}{Z}\int d\mathbf{h}\, \exp\left[\mathbf{k}.\mathbf{s}(\mathbf{x},\mathbf{h})\right] \tag{4}$$

where Z is now the partition function over all states (**x**,**h**). Defining G as in equation (1) then leads to

$$\frac{\partial G}{\partial k_i} = \left\langle s_i(\mathbf{x},\mathbf{h})\right\rangle_{P(\mathbf{x},\mathbf{h})} - \left\langle s_i(\mathbf{x},\mathbf{h})\right\rangle_{P(\mathbf{h}|\mathbf{x})P_o(\mathbf{x})} \tag{5}$$

The first term in equation (5) is estimated by using a Monte Carlo procedure, whereas the second term is a hybrid which uses the training set to provide samples from $P_o(\mathbf{x})$ and a Monte Carlo procedure (with **x** held constant) to provide samples from P(**h**|**x**). The advantage of introducing the hidden variables **h** is that a complicated P(**x**) can be generated by using simple **s**(**x**,**h**) because the effect of **h**–**h** and **h**–**x** interactions "dresses" the bare **x**–**x** interactions. This amounts to learning **s**(**x**) adaptively by adjusting the strengths of the interactions with and amongst the hidden variables.

The so–called Boltzmann machine [Ackley et al, 1985] is a simple form of hidden variable model which uses binary variables **x** and **h** and quadratic interactions **s**(**x**,**h**) together with G–maximisation. More general hidden variable models have been discussed elsewhere [Luttrell, 1985; Sejnowski, 1986]. Whilst the Boltzmann machine is very flexible in its ability to adapt to the statistical properties of $P_o(\mathbf{x})$, it is computationally very inefficient due to the extensive Monte Carlo simulations which are required.

2.2. DESIGNING POTENTIALS

There is another G–maximisation approach to learning **s**(**x**) for which the constraints are not on $\langle \mathbf{s}(\mathbf{x})\rangle$ but on the whole PDF $p_o(\mathbf{s})$ of **s**(**x**). In general $p_o(\mathbf{s})$ is given by

$$p_o(\mathbf{s}) \equiv \int d\mathbf{x}\, P_o(\mathbf{x})\, \delta(\mathbf{s}-\mathbf{s}(\mathbf{x})) \tag{6}$$

where the Dirac delta function constrains the **x** integral as required. If the **x** dependence of the observed $P_o(\mathbf{x})$ can be expressed entirely in terms of **s**(**x**), then **s**(**x**) is a sufficient set of statistics [DeGroot, 1970; Luttrell, 1987a]. The maximum entropy reconstruction (with a uniform prior) will then be $P_o(\mathbf{x})$ if the entire PDF $p_o(\mathbf{s})$ is used as a constraint. On the other hand, when **s**(**x**) is not a sufficient set of statistics, the maximum entropy method will, as usual, give the least committal reconstruction P(**x**) of $P_o(\mathbf{x})$ which is consistent with $p_o(\mathbf{s})$ [Luttrell, 1988a]

$$P(\mathbf{x}) = \frac{p_o(\mathbf{s}(\mathbf{x}))}{Z(\mathbf{s}(\mathbf{x}))} \quad ; \quad Z(\mathbf{s}) \equiv \int d\mathbf{x}\ \delta(\mathbf{s}-\mathbf{s}(\mathbf{x})) \tag{7}$$

$Z(\mathbf{s})$ is proportional to the number of states $\mathbf{x}$ which map to $\mathbf{s}$: it plays the role of a normalisation factor (it is not a partition function). With this interpretation the expression for $P(\mathbf{x})$ in equation (7) is intuitively obvious.

Using the definition of G in equation (2), substituting in $P(\mathbf{x})$ from equation (7), and using the definition of $p_o(\mathbf{s})$ in equation (6) yields

$$G = G_o + \int d\mathbf{s}\ p_o(\mathbf{s}) \log\left[\frac{p_o(\mathbf{s})}{Z(\mathbf{s})}\right] \tag{8}$$

where G_o is a constant. In order to optimise G we envisage two distinct types of change to $\mathbf{s}(\mathbf{x})$: a dimensionality preserving perturbation $\mathbf{s}(\mathbf{x}) \rightarrow \mathbf{s}(\mathbf{x}) + \epsilon \mathbf{t}(\mathbf{x})$, and a dimensionality increasing change $\mathbf{s}(\mathbf{x}) \rightarrow (\mathbf{s}(\mathbf{x}), \mathbf{t}(\mathbf{x}))$. We shall now present the results for these two cases.

For perturbations of the form $\mathbf{s}(\mathbf{x}) \rightarrow \mathbf{s}(\mathbf{x}) + \epsilon \mathbf{t}(\mathbf{x})$ we require the functional derivative $\delta G / \delta \mathbf{s}(\mathbf{x})$ which, in turn, requires the results

$$\frac{\delta p_o(\mathbf{s})}{\delta \mathbf{s}(\mathbf{x})} = -P_o(\mathbf{x})\, \nabla_{\mathbf{s}} \delta(\mathbf{s}-\mathbf{s}(\mathbf{x})) \quad ; \quad \frac{\delta Z(\mathbf{s})}{\delta \mathbf{s}(\mathbf{x})} = -\nabla_{\mathbf{s}} \delta(\mathbf{s}-\mathbf{s}(\mathbf{x})) \tag{9}$$

where $\nabla_{\mathbf{s}}$ is the derivative operator wrt $\mathbf{s}$: these results permit functional differentiation to be replaced by ordinary differentiation. After some manipulation we then obtain the functional derivative in the form [Luttrell, 1988a]

$$\frac{\delta G}{\delta \mathbf{s}(\mathbf{x})} = \left[P_o(\mathbf{x}) - P(\mathbf{x})\right] \nabla_{\mathbf{s}} \log\left[\frac{p_o(\mathbf{s})}{Z(\mathbf{s})}\right]_{\mathbf{s}=\mathbf{s}(\mathbf{x})} \tag{10}$$

which yields a change ΔG_1 in G given by

$$\Delta G_1 = \epsilon \int d\mathbf{s}\, d\mathbf{t} \left[p_o(\mathbf{s},\mathbf{t}) - p(\mathbf{s},\mathbf{t})\right] \mathbf{t}.\nabla_{\mathbf{s}} \log\left[\frac{p_o(\mathbf{s})}{Z(\mathbf{s})}\right] \tag{11}$$

where

$$p_o(\mathbf{s},\mathbf{t}) \equiv \int d\mathbf{x}\, P_o(\mathbf{x})\, \delta(\mathbf{s}-\mathbf{s}(\mathbf{x}))\, \delta(\mathbf{t}-\mathbf{t}(\mathbf{x}))$$

$$p(s,t) \equiv \int dx\, P(x)\, \delta(s-s(x))\, \delta(t-t(x)) \tag{12}$$

Alternatively, for changes of the form $s(x) \to (s(x),t(x))$ we obtain a change ΔG_2 in G given by [Luttrell, 1988a]

$$\Delta G_2 = \int ds\, dt\, p_o(s,t) \log\left[\frac{p_o(s,t)}{p(s,t)} \right] \tag{13}$$

Both ΔG_1 and ΔG_2 depend on a comparison of $p_o(s,t)$ and $p(s,t)$. Any differences between $p_o(s,t)$ and $p(s,t)$ are caused by the presence of structure in $P_o(x)$ which is not measured by $s(x)$ alone.

We have used the techniques outlined in this subsection to design MRF coherent image texture models [Luttrell, 1987b,c,d], where we assumed that $P_o(x)$ describes spatially stationary statistics (ie $P_o(Lx)=P_o(x)$ where L is any image translation operator). It then suffices to consider only $s(x)$ for which $s(Lx)=s(x)$, which severely restricts the set of feasible $s(x)$. The approach which we have presented in this subsection does not involve any hidden variables, so it has difficulty dealing with subtle properties of $P_o(x)$ which are better described by introducing "spectator variables". However it does successfully model short range textural properties.

3. Cluster decomposition model

We now propose a novel scheme for representing PDFs which completely eliminates the need for Monte Carlo simulations, whilst retaining the flexibility of the adaptive approach. This improvement is obtained at the cost of imposing an artificial hierarchical structure on the PDF reconstruction.

3.1. MULTILAYER NEURAL NETWORK

For simplicity consider the following situation

$$x \equiv (x_1,x_2) \quad ; \quad s(x) \equiv (s_1(x_1),s_2(x_2)) \tag{14}$$

Now suppose that the estimated values of $P_o(x_1)$, $P_o(x_2)$ and $p_o(s_1,s_2)$ are used as constraints on a maximum entropy reconstruction $P(x)$ of $P_o(x)$ (with a uniform prior). After some algebra which is similar to that which led to equation (7) we obtain

$$P(x) = P_o(x_1)\, P_o(x_2) \left[\frac{p_o(s_1(x_1),s_2(x_2))}{p_o(s_1(x_1))\, p_o(s_2(x_2))} \right] \tag{15}$$

This expression has a natural interpretation. If x_1 and x_2 are independent random

variables then so also are s_1 and s_2, yielding $p_o(s_1,s_2)=p_o(s_1)p_o(s_2)$, hence $P(x)=P_o(x_1)P_o(x_2)$, as expected. On the other hand, if x_1 and x_2 are mutually dependent then there is an additional correction term.

This approach to estimating $P_o(x)$ is usually simpler than specifying $P_o(x)$ directly when $\dim(s_1)<\dim(x_1)$ and $\dim(s_2)<\dim(x_2)$. This is because the cost of exhaustively specifying a PDF increases exponentially with the dimensionality of its underlying space, so specifying three low dimensional PDFs $P_o(x_1)$, $P_o(x_2)$ and $p_o(s_1,s_2)$ is usually cheaper than specifying one high dimensional PDF $P_o(x)$.

The above decomposition of $P_o(x)$ immediately generalises to

$$P(x)=\left[\prod_{i=1}^{n}P_o(x_i)\right]\left[\frac{p_o(s(x))}{\prod_{i=1}^{n}p_o(s_i(x_i))}\right] \tag{16}$$

where $x\equiv(x_1,x_2,\ldots,x_n)$, $s\equiv(s_1,s_2,\ldots,s_n)$, $s_i\equiv s_i(x_i)$ and $\dim(s_i)<\dim(x_i)$ for $i=1,2,\ldots,n$. Now suppose that $\dim(x_i)$ and $\dim(s_i)$ are small enough that $P_o(x_i)$ and $p_o(s_i)$ are easy to estimate ($i=1,2,\ldots,n$), then it remains to estimate $p_o(s_1,s_2,\ldots,s_n)$. The original problem of estimating $P_o(x)$ has been replaced by an analogous (but simpler) problem of estimating $p_o(s)$ where $\dim(s)<\dim(x)$. The maximum entropy procedure may be iterated to yield an estimate of $p_o(s)$ itself, and so on until the dimensionalities encountered are low enough for a direct estimation of the remaining PDFs to be made. This produces a hierarchical cluster decomposition of the original x because the layers of sampling functions form a tree–structure. This is a type of multilayer neural network.

3.2. TOPOGRAPHIC SAMPLING FUNCTIONS

The main problem with this type of cluster decomposition is the selection of sampling functions. The $s_i(x_i)$ must not only be good sampling functions insofar as the statistical properties of x are concerned, but also the $s_i(x_i)$ must stand as reduced dimension representations of the x_i themselves so that sampling process can be iterated.

A novel means of deriving a reduced dimension representation $s(x)$ of an input x is to use topographic sampling functions [Kohonen, 1984]. Define a vector quantisation $s(x)$ of x thus

$$s_o \text{ minimises } \left|x-x(s)\right|^2 \text{ wrt } s \quad ; \quad s_o\equiv s(x) \tag{17}$$

where $x(s)$ is a code book of quantisation vectors parameterised by s. An update scheme which improves $x(s)$ in response to samples x drawn from $P_o(x)$ is

$$x(s)\longrightarrow x(s)+\epsilon\left(\left|s-s(x)\right|\right)\left[x-x(s)\right] \tag{18}$$

where $\epsilon(r)$ is a non−negative monotonically decreasing function of r. The update function $\epsilon(r)$ must have a finite width (in r) to ensure that $\mathbf{x}(\mathbf{s})$ is a continuous function of $\mathbf{s}$. The converse ($\mathbf{s}(\mathbf{x})$ a continuous function of $\mathbf{x}$) is usually not possible when $\dim(\mathbf{s})<\dim(\mathbf{x})$. Physically $\mathbf{x}(\mathbf{s})$ can be thought of as a $\dim(\mathbf{s})$ dimensional manifold embedded in $\dim(\mathbf{x})$ dimensions. Equation (18) describes the dynamical behaviour of this manifold in response to being "pulled" by $\mathbf{x}$: the response of the manifold is rather like that of a stiff sheet. A particularly desirable property of this learning algorithm is that $p_o(\mathbf{s})$ tends to become constant, thus maximising the output entropy. Furthermore, the algorithm can be shown to minimise $<\log(V(\mathbf{x}))>$ where $V(\mathbf{x})$ is the error volume associated with the reconstruction of $\mathbf{x}$ from $\mathbf{s}$ using $\mathbf{x}(\mathbf{s})$ [Luttrell, 1988b]. It is these properties of the learning algorithm which lead to $\mathbf{s}(\mathbf{x})$ being called a topographic sampling function.

A hierarchy of topographic sampling functions can be derived by extending this optimisation scheme. This produces a cluster decomposed representation $P(\mathbf{x})$ of $P_o(\mathbf{x})$ as explained after equation (16). Furthermore, various improvements can also be introduced which enormously speed up the convergence of the update scheme as originally proposed [Luttrell, 1988c].

4. Conclusions

There is a pressing need for representations of PDFs in situations where direct physical insight fails to provide a complete model. We have discussed two alternative techniques: MRFs and cluster decomposition. The MRF technique potentially can represent a PDF very accurately by using a sufficiently complicated set of potentials and/or hidden variables, but the computational cost of Monte Carlo simulation of MRFs can be unacceptable. The cluster decomposition technique imposes an artificial hierarchical structure on the PDF which can lead to inaccuracies in representation, but it involves no Monte Carlo simulations. For real time applications we recommend the use of cluster decomposition.

5. References

Ackley D H, Hinton G E and Sejnowski T J, 1985, Cogn. Sci., **9**, 147−169, 'A learning algorithm for Boltzmann machines'.

Besag J, 1974, *J. R. Statist. Soc, Ser. B*, **36**, 192−236, 'Spatial interaction and the statistical analysis of lattice systems'.

DeGroot, 1970, *Optimal statistical decisions*, New York, McGraw−Hill.

Geman S and Geman D, 1984, *IEEE PAMI*, **6(6)**, 721−741, 'Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images'.

Geman S and Graffigne C, *Proc. Int. Cong. Math. 1986*, Ed. Gleason A M, Am. Math. Soc., Providence, 1987, 'Markov random field image models and their applications to computer vision'.

Jaynes E T, 1957, *Phys. Rev.*, **106(4)**, 620−630, 'Information theory and statistical mechanics'.

−, 1957, *Phys. Rev.*, **108(2)**, 171−190, ' Information theory and statistical mechanics. II'.

−, 1968, *IEEE Trans. SSC*, **4(3)**, 227−241, 'Prior probabilities'.

−, 1982, *Proc. IEEE*, **70(9)**, 939−952, 'On the rationale of maximum−entropy methods'.

Kindermann R and Snell J L, 1980, *Markov Random Fields and their Applications, Contemporary Mathematics*, Vol. 1, Am. Math. Soc., Providence, Rhode Island.

Kohonen T, 1984, *Self organisation and associative memory*, Springer–Verlag.

Luttrell S P, 1985, *RSRE Memo.*, **3815**, 'The implications of Boltzmann–type machines for SAR data processing'.

–, 1987a, *Inv. Prob.*, **3**, 289–300, 'The use of Markov random field models to derive sampling schemes for inverse texture problems'.

–, 1987b, *Proc. AGARD Conf. on Scattering and Propagation in Random Media*, 'Markov random fields: a strategy for clutter modelling'.

–, 1987c, *Proc. SPIE Int. Symp. on Inverse Problems in Optics*, **808**, 182–188, Ed. Pike E R, 'The use of Markov Random field models in sampling scheme design'.

–, 1987d, *Proc. IEE RADAR–87 Conf.*, 222–226, 'Designing Markov random field structures for clutter modelling'.

–, 1988a, to be published in *Inv. Prob.*, 'A maximum entropy approach to sampling function design'.

–, 1988b, *Proc. IGARSS'88 Conf. on Remote Sending Moving Towards the 21st Century*, 'Image compression using a neural network'.

–, 1988c, submitted to *Patt. Recog. Letts.*, 'Image compression using a multilayer neural network'.

Metropolis N, Rosenbluth A W, Rosenbluth M N and Teller A H, 1953, *J. Chem. Phys.*, **21**, 1087–1092, 'Equation of state calculations by fast computing machines'.

Preston C J, 1974, *Gibbs States on Countable Sets*, Cambridge University Press.

Sejnowski T J, 1986, *Proc. Conf. on Neural networks for Computing*, Vol. 151, Am. Inst. Phys., Snowbird, Utah, 'Higher order Boltzmann machines'.

ELECTRONIC 'NEURAL' NETS FOR SOLVING ILL-POSED PROBLEMS WITH AN ENTROPY REGULARISER

C.R.K. MARRIAN, M.C. PECKERAR and I.A. MACK
Electronics Science and Technology Division,
Naval Research Laboratory,
Washington DC, USA.

and Y.C. PATI
University of Maryland,
College Park MD, USA.

ABSTRACT. Multiply connected analogue electronic circuits ('neural' nets) are characterised by having a large number of simple processing nodes such as threshold or summing devices which are connected to many other nodes through weighted interconnects. Under certain conditions the transient and equilibrium behaviour of these nets can be described in terms of a stability or Lyapunov function which is minimised at the equilibrium conditions of the net. This paper describes a circuit which is capable of solving ill-posed problems through use of an informational entropy regulariser which is incorporated into the stability function of the net. A circuit has been constructed which provides maximum entropy solutions to the loaded dice problem. The performance in terms of accuracy and speed of such circuits is discussed.

1. Introduction

1.1 'NEURAL' NETS.

The concept of an electronic 'neural' net is based on some greatly simplified ideas of the processing architecture of neurobiological systems. Essentially a large number of simple processing nodes are connected to a large number of other nodes. The processing capability of such nets is determined by such factors as the specific architecture, node characteristics, interconnect values etc.. Impressive demonstrations of the power of these nets in areas such as associative memories, artificial vision systems, pattern classifiers, combinatorial optimisation etc. abound in the literature [1].

The work of J.J. Hopfield and D.W. Tank [2-4] first demonstrated that one could build actual electronic circuits of this type which exhibited substantial computational power. These circuits are characterised by one layer of nodes which have interconnects from outputs to inputs. This is in contrast to the multi-layered or perceptron like architectures where the outputs of one layer are connected to the inputs of the next layer, an example of which is given in reference [5].

Under certain conditions, the behaviour of a single layer, fully interconnected net can be described in terms of a stability or Lyapunov function [4,6]. This allows a net to

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 371–376.

be designed to minimise a specific cost function which then determines the equilibrium state of the net.

1.2 MAXIMUM ENTROPY ALGORITHM

In terms of a deconvolution problem (assumed to be ill-posed), the algorithm [7] determines a solution by minimising a cost function E of the form:

$$E = |\mathbf{O}*\mathbf{T} - \mathbf{I}|^2 - \beta S.$$

T is the *known* convolution function, **I** the observed (noisy) convolved image of an unknown object **O** assumed to be describable by a positive additive distribution (PAD), β is a constant and S is the informational entropy of **O**. The first term in the expression for E can be considered as a measure of the amount which the solution **O** violates the constraints defined by **I**. In cases where **I** is contaminated by noise which can be considered gaussian and stationary, this term is proportional to the log of the likelihood of **O**. For S, we have used a form of cross entropy due to J. Skilling [8] which in discrete notation is given by:

$$S = \sum_i O_i - \sum_i M_i - \sum_i O_i \log(O_i/M_i)$$

where the PAD, **M**, can be considered as a prior estimate of **O**. Note as ΣM_i is constant it can be ignored in minimising E.

The algorithm is based on a gradient search so it is desired that:

$$\frac{dO_i}{dt} \propto -\frac{\partial E}{\partial O_i} = -2\sum_j T_{ij}\left(\sum_k T_{kj} O_k - I_j\right) - \beta \log(O_i/M_i) \quad (1)$$

In section 2, the implementation in a multiply connected circuit of this algorithm is described. In section 3, a circuit constructed to provide the maximum entropy solution to the 'loaded dice' problem [9] is presented and its accuracy and transient behaviour described. Section 5 contains a discussion of the net's properties and a summary.

3. Net Formalism

The first bracketed term on the right hand side of equation (1) can be achieved by summing currents at the inputs of N_c (number of pixels in the image data **I**) virtual earth amplifiers. For example, node j (=1 to N_c) will have the voltages **O** connected to its input through conductances defined by the matrix **T** and an external current input equal to $-I_j$. The output voltage f_j of this node is a voltage proportional to the net current at its input. These are referred to as the constraint nodes as each I_j can be considered a constraint which the outputs of the signal nodes must meet.

The outputs of the constraint nodes are inverted and fed into the inputs of N_s signal nodes through conductances again defined by the matrix **T** to provide a net current input corresponding to the first term on the right hand side of equation (1). The factor of 2 can be considered as part of the constant β. The time differential is achieved by shunting the input with a capacitor C. The second term on the right of (1) is applied as an external current input of $-\beta\log(1/M_i)$ and by a resistor to make the total resistance

shunting the input of each signal node $1/\beta$. If the voltage at the input to signal node i (both i and k are taken to be indexed from 1 to N_S) is represented by u_i, the current through the capacitor C can be written as:

$$C \frac{du_i}{dt} = - \sum_j T_{ij} f_j - \beta \log(1/M_i) - \beta u_i$$

Now if

$$u_i = g^{-1}(O_i) = \log(O_i),$$

i.e. the input-output characteristic, g, of the signal nodes is made exponential, the current through the capacitor will be of the form $\partial E/\partial O_i$. However du_i/dt can be written:

$$\frac{du_i}{dt} = \frac{1}{O_i} \frac{dO_i}{dt} \qquad (2)$$

As **O** is a PAD which is, of course, ensured by the exponential signal nodes, the circuit will make a gradient search for **O**. The circuit then has the form shown schematically in figure 1, where two signal and two constraint nodes are shown.

E can be shown to be the Lyapunov function of the circuit by summing the integral of the current with respect to the voltage over each capacitor in the circuit with the interconnects and node characteristics described above [6]. By way of confirmation, it can be demonstrated that E will decrease monotonically with time from equation (2), i.e.

$$\frac{dE}{dt} = \frac{\partial E}{\partial \mathbf{O}} \frac{d\mathbf{O}}{dt} = - \sum_i \frac{1}{O_i} \frac{dO_i}{dt} \frac{dO_i}{dt}$$

which is clearly ≤ 0 for all t as O_i is >0 for all i.

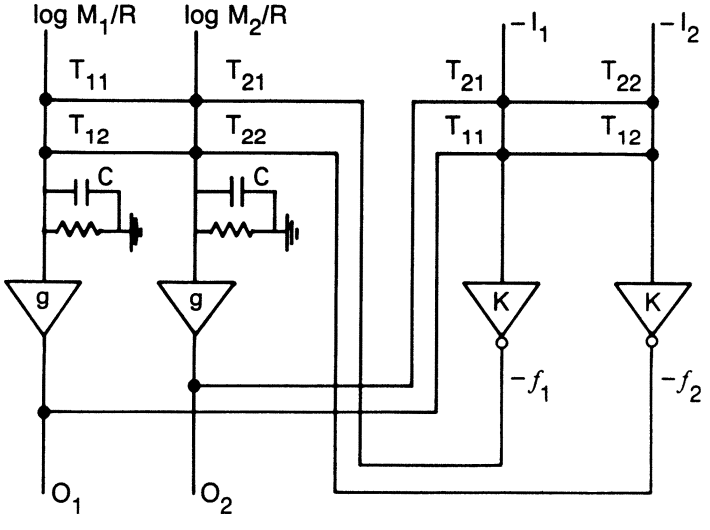


Figure 1. Schematic of the MaxEnt 'Neural' net

4. The Loaded Dice Problem.

4.1 CIRCUIT DESCRIPTION

It should be emphasised that this problem was chosen for the purpose of a demonstration as it has a well accepted maximum entropy solution. Thus the performance of the net can be compared to a theoretical solution as opposed to simply the results of a computer simulation. The application of the algorithm and actual nets to real problems such as spectral deconvolution [7] and tactile sensors [6] is described elsewhere.

The loaded dice problem [9] requires an estimate of the biases on a six sided dice given only information on the average throw. The net will therefore have six signal nodes (i.e. N_s=6) with outputs corresponding to the biases of each face and two constraint nodes (N_c=2) corresponding to normalisation and the observed average τ. Thus the outputs from the two constraint nodes can be written:

$$f_1 = K_1 (\Sigma_i O_i - 1) \quad \text{and} \quad f_2 = K_2 (\Sigma_i O_i i - \tau) \qquad (3)$$

i.e. I_1=1 and I_2=τ. K_1 and K_2 are the gains of the two constraint nodes. Thus the interconnects **T** are given by:

$$T_{i1} = 1 \quad \text{and} \quad T_{i2} = i$$

The steady state output of the net is given by:

$$O_i = M_i \, e^{-f_1 - f_2 i}$$

where M_i is the prior which was set to 1/6. Note that this is similar in form to the maximum entropy solution which can be written as:

$$O_i = 1/Z \; e^{-\lambda i}$$

where Z is the partition function and λ the Lagrange multiplier. Thus the Lagrange multiplier will be simply the voltage at the output of the second constraint node !

The exact maximum entropy solution will, however, only be attained in the limit of infinite gains on the constraint nodes. For finite gains the accuracy can be written from equations (3) in terms of the closeness to normalisation and the observed average throw. For the case where τ=4.5, the requirement of 0.1% accuracy implies that $K_1 \approx 1300$ and $K_2 \approx 370$. In practice these are close to the maximum gains realisable with standard operational amplifiers.

A circuit was constructed with the parameters defined above. The signal nodes used two op-amps and a diode to provide the exponential characteristic. Each constraint node consisted of three op-amps to provide high gain. The interconnects were 1% resistors selected to be within 0.1% of their nominal values. The size of the capacitors was chosen to prevent oscillations in the circuit caused by phase shifts introduced by the constraint nodes. As a result of the high gains, a rather high value for C (3 μF) was required.

4.2 STEADY STATE AND TRANSIENT MEASUREMENTS.

The output voltages measured for various values of the input τ to the average throw constraint node are compared to the calculated maximum entropy solution in Table I. The

case for τ=2.5 shows the greatest discrepancies with the exact solution. This is due to the op-amps not having an input offset voltage nulling facility. The offsets at the two constraint nodes can be compensated for by adjusting their external inputs corresponding to normalisation (i.e. 1) and the average throw (τ). This results in the reduced errors shown for τ=4.5 and 3.5, where close to the 0.1% accuracy is shown.

TABLE I Steady State Circuit Outputs in Volts

Average	τ=4.5		τ=3.5		τ=2.5	
	Calc.	Meas.	Calc.	Meas.	Calc.	Meas.
Face					no offset	null
1	0.054	0.053	0.167	0.168	0.348	0.368
2	0.079	0.078	0.167	0.167	0.240	0.243
3	0.114	0.114	0.167	0.168	0.166	0.161
4	0.166	0.167	0.167	0.166	0.114	0.098
5	0.240	0.242	0.167	0.166	0.079	0.068
6	0.348	0.347	0.167	0.166	0.054	0.043
f_1	1.491	1.41	0.000	0.00	-1.106	-1.34
$\lambda=f_2$	-0.371	-0.35	0.000	0.00	0.371	0.42

The transient response was monitored by applying a square wave to the average constraint node. The transient was limited by the time required to charge the capacitors which is determined by their size and the maximum voltage swing of the constraint nodes. Settling times of the order of 15 ms were observed. Full details will be published separately [10].

5. Discussion.

5.1 TRANSIENT TIME

If it were possible to build perfect op-amps (i.e. with no phase shift) the capacitors C could be made arbitrarily small and the transient response of the circuit would be correspondingly arbitrarily fast. The loaded dice problem has two constraints which are 'hard', i.e. they must be satisfied exactly. As a result, high gains in the constraint nodes are required and large capacitors are necessary for stability. In the case of the deconvolution of noisy data, for example, the constraints (the data in the observed image) are 'soft', i.e. it is not required for them all to be satisfied exactly. This is reflected in lower constraint node gains which reduce the phase shift problem. Consequently, smaller capacitors can be used and transient times are decreased. A circuit built to deconvolve the stress-strain kernel of a tactile sensor [6] has shown transient times of 20 μs.

5.2 UNIQUENESS OF SOLUTION

One can show that there exists a one parameter family of solutions for the equilibrium equations of the net. The solutions are parameterized by the constant

determining the weight given to the entropy term in the cost function. Furthermore, the cost function is strictly convex in terms of the output from the net **O**. Thus it is clear that the minimum in the cost function will be the global minimum rather than a spurious local minimum. So for a given **I**, **T** and β it is guaranteed that the net will settle into the same solution irrespective of the initial condition of the net. The problem of undesired localized minima is a significant problem with many 'neural' net architectures. These nets reach a point of equilibrium but there is often no way of determining whether it is an undesired spurious minimum. As a result one cannot be certain the solution given by the net is that for which one is searching. In addition, there are no restrictions on the form of the **T** matrix as in reference [2], for example. Thus **T** may be asymmetric, non-square and have non-zero diagonals and reflect a measured convolution function from, for example, an energy dispersive spectrometer or an imaging system.

5.3 SUMMARY

We have described an algorithm based on an ideal multiply connected analogue circuit which solves ill-posed problems, such as the deconvolution of noisy data, through the use of an informational entropy regulariser. The behaviour of the circuit is characterised by a Lyapunov function which allows a specific cost function to be minimised. The cost function is a weighted difference between the amount the constraints on the solution are broken and the informational entropy of the solution.

Actual circuits have been built to implement the algorithm using standard electronic components. Overall accuracies of 0.1% and transient times of 20 μs have been demonstrated. We are presently working towards the construction of an integrated circuit which will implement this algorithm in a compact form.

6. References.

[1] For example, 'Proceedings of the First International Conference on Neural Nets' ed. M. Caudhill & C. Butler, IEEE, 1987.
[2] J.J. Hopfield Proc. Nat. Acad. Sci. USA, **79**, p. 2554, 1982.
[3] J.J. Hopfield & D.W. Tank, Biological Cybern., **52**, p. 141, 1985.
[4] D.W. Tank & J.J. Hopfield, IEEE Trans. on Circ. and Sys., **CAS-33**, #5, p. 533, 1986.
[5] T.J. Sejnowski & C.R. Rosenberg in 'Neurocomputing' ed. J.A. Anderson & E. Rosenfeld, MIT Press, Cambridge MA, 1988.
[6] Y.C. Pati, D. Friedman, P.S. Krishnaprasad, C.T. Yao, M.C. Peckerar, R. Yang and C.R.K. Marrian, 'Neural Networks for Tactile Perception', Proc. IEEE Automation and Robotics Conference, Philadelphia PA, 1988.
[7] C.R.K. Marrian and M.C. Peckerar, 'Electronic "Neural" Net Algorithm for Maximum Entropy Solutions of Ill-posed Problems', to appear in IEEE Trans. on Circuits and Systems.
[8] J. Skilling, 'The axiom of Maximum Entropy' presented at 1986 Maximum Entropy Workshop, Seattle WA, 1986.
[9] 'E.T. Jaynes: Papers on Probability Statistics and Statistical Physics', p. 39 & 211, ed. R.D. Rosenkrantz, D. Reidel, Dordrecht, Holland, 1983.
[10] C.R.K. Marrian, I. Mack, C. Banks & M.C. Peckerar, to be published.

BAYESIAN MODEL SELECTION: EXAMPLES RELEVANT TO NMR

G. LARRY BRETTHORST
Department of Chemistry
Campus Box 1134
Washington University
1 Brookings Drive
St. Louis, MO 63130

Abstract. The model selection problem is one of the most basic problems in data analysis. Given a data set one can always expand the model almost indefinitely. How does one pick a model which explains the data, but does not contain spurious features relating to the noise? Here we present the results of a Bayesian model selection calculation started in [1] and then extended in [2], and show that the Bayesian answer to this question is essentially a quantitative statement of Occams razor: When two models fit the evidence in the data equally well, choose the simpler model.

Introduction

When analyzing the results of an experiment it is not always known which model function applies. We need a way to choose between several possible models. This is easily done using Bayes' theorem. The first step in answering this question is to enumerate the possible models. Suppose we have a set S of s possible models with model functions $\{f_1, \cdots, f_s\}$. We are hardly ever sure that the "true" model is actually contained in this set. Indeed, the "set of all possible models" is not only infinite, but it is also quite undefined. It is not even clear what one could mean by a "true" model. Both questions may take us into an area more like theology than science.

The only questions we seek to examine are the ones that are answerable because they are mathematically well-posed. Such questions are of the form: "Given a specified set S_s of possible models $\{f_1, \cdots, f_s\}$ and looking only within that set, which model is most probable in view of all the data and prior information, and how strongly is it supported relative to the alternatives in that set?" Bayesian analysis can give a definite answer to such a question – see [2], [3], [4]. Here we give the results of the calculation done in [2] and present two numerical examples of its use.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 377–388.

The Relative Probability of Model f_j

Given a set S_s of models $\{f_1, \cdots, f_s\}$ and looking only within that set, which model best accounts for the data? We will take

$$f_j(t) = \sum_{k=1}^{m} A_k H_k(t, \{\omega\}) \tag{1}$$

as our model, where H_k are the orthonormal model functions defined in [1]. The subscript "j" refers to the jth member of the set S_s of models $\{f_1, \cdots, f_s\}$ with the understanding that the amplitudes $\{A\}$, the nonlinear $\{\omega\}$ parameters, the total number of model functions m, the total number of nonlinear parameters r, and the model functions $H_k(t, \{\omega\})$ are different for every f_j.

The use of the orthogonal models does not change the generality of the calculation because any arbitrary model may be transformed into an orthogonal model. If we have a nonorthogonal model

$$f_j(t) = \sum_{k=1}^{m} B_k G_k(t, \{\omega\}), \tag{2}$$

where G_k is the model function (for example a sine or Bessel function) and B_k is its amplitude, then we may transform this model into an orthogonal model, Eq. (1), as follows: compute the interaction matrix

$$g_{kl} = \sum_{i=1}^{N} G_k(t_i) G_l(t_i) \tag{3}$$

and from the kth eigenvalue λ_k of the interaction matrix, Eq. (3), and the lth component of the kth eigenvector e_{kl} compute the orthogonal model functions H_k given by

$$H_k(t) = \frac{1}{\sqrt{\lambda_k}} \sum_{l=1}^{m} e_{kl} G_l(t).$$

The orthogonal amplitudes A_k may be computed from a linear combination of the B_l:

$$A_k = \sqrt{\lambda_k} \sum_{l=1}^{m} B_l e_{kl}.$$

From Bayes' theorem we may compute the posterior probability of model f_j:

$$P(f_j|D, I) = \frac{P(f_j|I) P(D|f_j, I)}{P(D|I)} \quad \text{and} \quad P(D|I) = \sum_{j=1}^{s} P(f_j|I) P(D|f_j, I). \tag{4}$$

We will assume, for now, that the variance of the noise σ^2 is known and derive $P(f_j|\sigma, D, I)$; then at the end of the calculation, if σ is not known we will remove it. Thus symbolically the heart of the problem is to compute

$$P(D|\sigma, f_j, I) = \int d\{A\} d\{\omega\} P(\{A\}, \{\omega\}|I) P(D|\{A\}, \{\omega\}, \sigma, f_j, I). \tag{5}$$

When we solved this problem, there were several places where prior information had to be incorporated: first, when we assigned a noise prior; second, when we removed the amplitudes; third, when we removed the nonlinear $\{\omega\}$ parameters; and fourth, when we removed the variance of the noise. When we assigned a noise prior we assumed the second moment of the noise was given and using maximum entropy arrived at a Gaussian prior as the least informative prior probability for the noise for a given second moment. The amplitudes are location parameters, and when we removed the amplitudes, we used a Gaussian, centered at zero, whose variance δ^2 expressed how strongly we believed the amplitudes to be near zero. From the form of the model, we do not know if the nonlinear $\{\omega\}$ parameters were location parameters or scale parameters. However, when we did this calculation we made a local Gaussian approximation to the posterior probability (i.e. we assumed the data determine the parameters well and we assumed the data determine the parameters much more precisely than the prior information). In this approximation the $\{\omega\}$ parameters are location parameters; thus we used a Gaussian centered at zero with variance γ^2 to represent the prior information about the nonlinear parameters. Last, if the three variances σ^2, δ^2, and γ^2 were not known, we removed them using a normalized Jeffreys prior. The normalization constant for the Jeffreys prior expresses the prior information in the form of a permissible range of values for the variances. This range of values appears in the problem as a natural logarithm of the upper limit divided by the lower limit. We designated this ratio as R_σ for the variance of the noise and similarly for γ and δ.

If the three variances are actually known, then the direct probability of the data is approximately given by

$$\begin{aligned} P(D|\gamma,\delta,\sigma,f_j,I) &\approx (2\pi\delta^2)^{-\frac{m}{2}} \exp\left\{-\frac{m\overline{h^2}(\{\hat{\omega}\})}{2\delta^2}\right\} \\ &\times (2\pi\gamma^2)^{-\frac{r}{2}} \exp\left\{-\frac{r\overline{\omega^2}}{2\gamma^2}\right\} v_1^{-\frac{1}{2}} \cdots v_r^{-\frac{1}{2}} \\ &\times (2\pi\sigma^2)^{-\frac{N-m-r}{2}} \exp\left\{-\frac{N\overline{d^2} - m\overline{h^2}(\{\hat{\omega}\})}{2\sigma^2}\right\} \end{aligned} \tag{6}$$

where $\overline{\omega^2}$ is the mean-square estimated $\{\omega\}$ parameter

$$\overline{\omega^2} = (1/r)\sum_{k=1}^{r} \hat{\omega}_k^2,$$

$\overline{h^2}(\{\hat{\omega}\})$ is the mean-square value of the h_k functions

$$\overline{h^2}(\{\hat{\omega}\}) \equiv \frac{1}{m}\sum_{k=1}^{m} h_k^2\Big|_{\{\hat{\omega}\}},$$

h_k is the projection of the data onto the orthonormal model functions H_k

$$h_k \equiv \sum_{i=1}^{N} d(t_i) H_k(t_i, \{\hat{\omega}\}),$$

$\{\hat{\omega}\}$ is the location of the maximum posterior probability digitized as $\{\hat{\omega}_1, \cdots, \hat{\omega}_r\}$, and v_k is one of the eigenvalues of the matrix

$$b_{jk} \equiv -\frac{m}{2} \frac{\partial^2 \overline{h^2}}{\partial\omega_j \partial\omega_k}\bigg|_{\{\hat{\omega}\}}.$$

If the three variances σ^2, δ^2, and γ^2 are not known, then they may be removed using a normalized Jeffreys prior. The global likelihood of the data is then approximately

$$\begin{aligned} P(D|f_j, I) \approx & \frac{\Gamma(m/2)}{2\log(R_\delta)} \left[\frac{m\overline{h^2}(\{\hat{\omega}\})}{2}\right]^{-\frac{m}{2}} \frac{\Gamma(r/2)}{2\log(R_\gamma)} \left[\frac{r\overline{\omega^2}}{2}\right]^{-\frac{r}{2}} v_1^{-\frac{1}{2}} \cdots v_r^{-\frac{1}{2}} \\ & \times \frac{\Gamma([N-m-r]/2)}{2\log(R_\sigma)} \left[\frac{N\overline{d^2} - m\overline{h^2}(\{\omega\})}{2}\right]^{\frac{m+r-N}{2}} \end{aligned} \tag{7}$$

where $\Gamma(x)$ is a gamma function of argument x. The three factors involved in normalizing the Jeffreys priors (R_σ, R_γ, R_δ) appear in every model; they always cancel as long as we are dealing with models having all three types of parameters. However, as soon as we try to compare a model involving two types of parameters to a model involving three types of parameters (e.g. a regression model to a nonlinear model) they no longer cancel: the prior ranges become important. One must think carefully about just what prior information one actually has about σ, γ, and δ and use that information to set the prior ranges.

Example – Multiple Exponential Decays

Two major problems in NMR are: determining the characteristic decay time of a signal (the so-called T_2 time), and determining how many resonances (frequencies) are in a free induction decay. We will give two examples of the use of Eq. (4), one on simulated multiple decaying exponential data, and one on simulated multiple nonstationary frequencies.

The data in T_2 experiments are a time series (typically nonuniformly sampled) which decays away exponentially. The problem is to determine how many exponentials are in the data, and to estimate the decay rates and their amplitudes. This is frequently done in one of two ways: least squares or curve stripping. In least squares, the experimenter will fit a model having one, two, three, etc. decaying exponentials to his data and then stop the process when (1) the model looks like the data, (2) the parameters are not physically meaningful, or (3) the parameters are not statistically

significant. In curve stripping, the data are plotted on a semi-logarithmic plot. On this plot a single exponential will appear as a straight line. The experimenter looks for how many straight lines are necessary to represent the data. The stopping criterion is typically set by the human eye. Neither least squares nor curve stripping has any theoretically justifiable stopping criterion; rather, they are intuitive procedures that give reasonable results.

Let us apply the procedures we have developed and see what probability theory can do on this problem. We take as our data, Fig. 1 (A), derived from

$$d(t_i) = 100\mathrm{e}^{-0.05t_i} + 50\mathrm{e}^{-0.02t_i} + n(0,1),$$

where $n(0,1)$ is a Gaussian random number with zero mean and standard deviation one. In this example we take $t_i = \{0, 0.5, 1, \cdots 100\}$, $N = 201$, with signal-to-noise ratio of approximately 60. We have displayed the data in Fig. 1 (A) and a semi-logarithmic plot of the data in Fig. 1 (B). We see little evidence of two decaying exponentials in these data.

To apply the procedures given here, we specify the set of models to be examined. We take

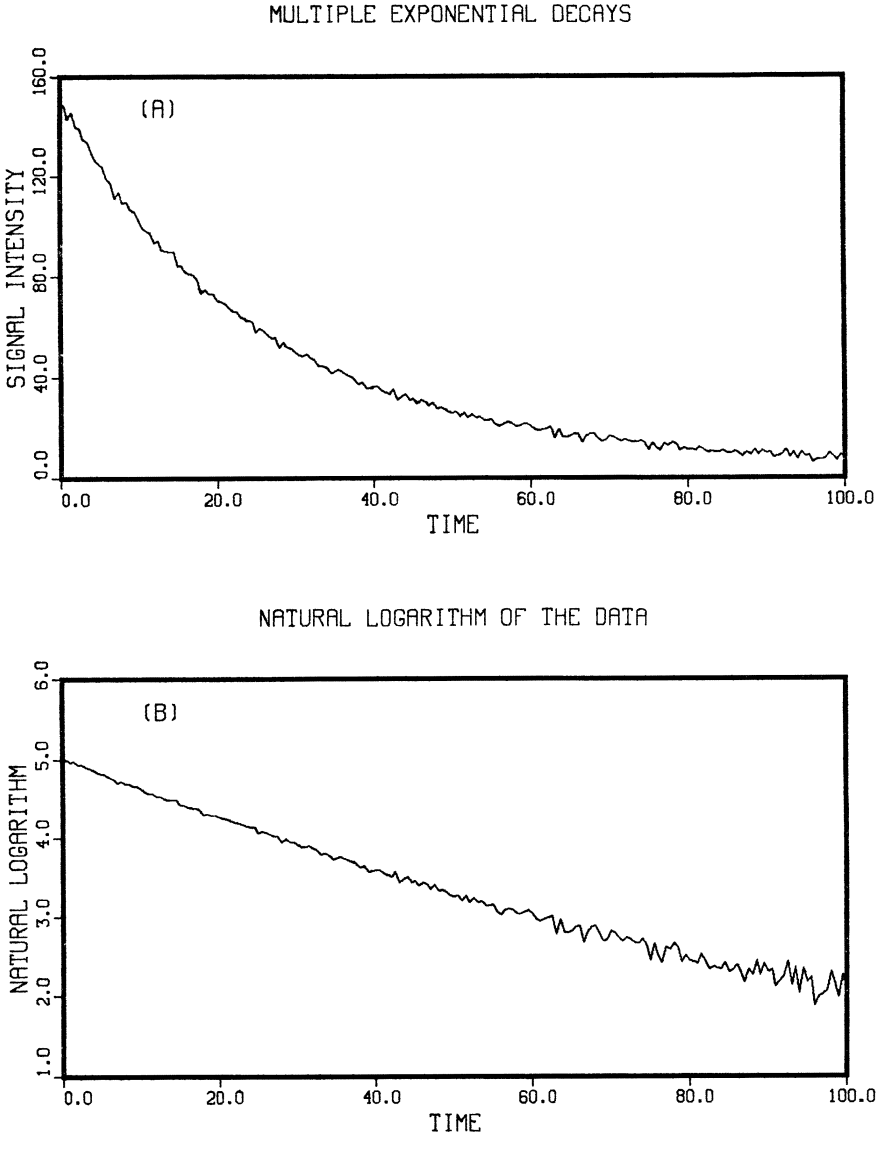
$$f_j(t) = \sum_{k=1}^{j} B_k \mathrm{e}^{-\alpha_k t} \qquad (j = 1, 2, \cdots)$$

as the nonorthogonal models, Eq. (2). The question we would like to answer is: "What value of j is necessary to account for all systematic effects in the data?" To answer this question we compute the orthogonal model, Eq. (1), and from the orthogonal model we compute the global likelihood of the data, Eq. (7). From the global likelihood we compute the posterior probability, Eq. (4). To compute the posterior probability we must assign a prior probability, $P(j|D, I)$, for the number of exponential components in the data. Having little prior information about this, we use a uniform prior probability.

The results of this calculation are displayed in Fig. 2. There are six plots in this figure occurring in pairs. We have plotted the data (dotted line) in panels (A), (C), and (E). Included on these panels as the solid line is a one exponential model (A), a two exponential model (C), and a three exponential model (E). Panels (B), (D), and (F) contain a plot of the residuals (i.e. the difference between the data and the model) for the different models. The residuals for the one exponential model, Fig. 2 (B), show a systematic effect, and the posterior probability of this model is zero to eight decimal places. The residuals for the two exponential model, Fig. 2 (D), look like white noise, and the posterior probability of this model is 0.9984. The residuals for the three exponential model, Fig. 2 (F), show no noticeable improvement compared to the two exponential model. The posterior probability of this model is 0.0016. Given the three choices, Bayesian probability theory has correctly determined the number of exponentials in the data.

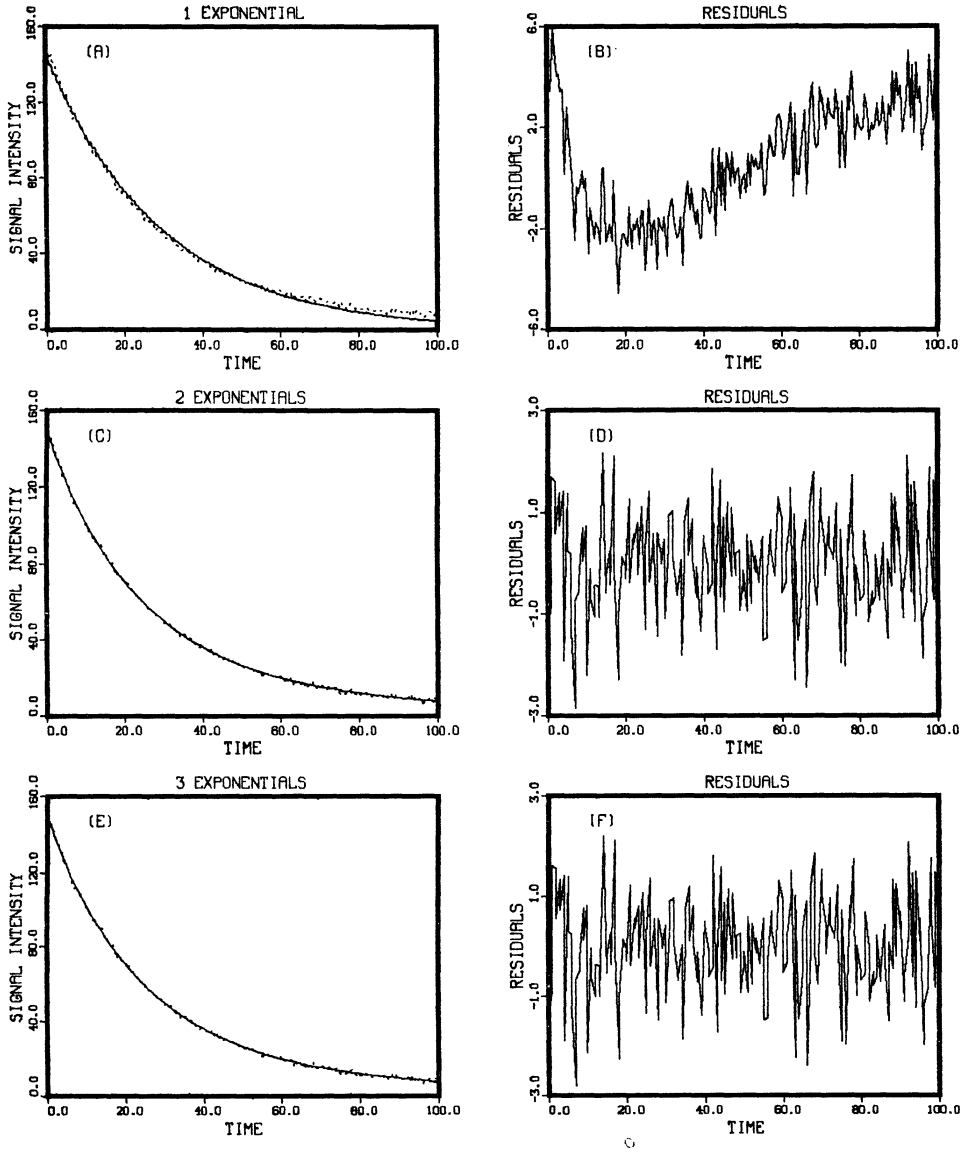
The choice between the two and three exponential model is very interesting. The residuals did not improve substantially for the three exponential model, and Bayes

Figure 1: Multiple Exponential Decays



The computer simulated data (A) contain two decaying exponentials – see text for the details. There are $N = 201$ data values with signal-to-noise ratio of approximately 60. Panel (B) is a plot of the natural logarithm of the data. From the plot there is little evidence for the second decaying exponential.

Figure 2: Multiple Exponential Decays



The data, (A) dotted line, were fit with a one exponential model, (A) solid line. The residuals are shown in (B). There is a systematic effect in the residuals. The probability of this mode is zero to 8 places. The data were then fit with a two exponential model, (C) dotted line. The residuals are shown in (D). The posterior probability of this model is 0.9984. A three exponential model is shown in (E) and the residuals in (F). There is no noticeable improvement. The posterior probability of this model is 0.0016.

theorem then tells us to choose the simpler model. The prior information is effectively doing this. When we did this calculation, we put in the fact that the amplitudes and decay rates should be taken to be zero unless the data clearly indicate otherwise. In the three exponential model, the data indicate the parameters should be nonzero. However, the prior probability of the model depends on the mean-square amplitude and the mean-square decay rate. When these factors are estimated to be nonzero, their prior probability is low. The fit for the three exponential model did not improve substantially: the probability of the data for the two and the three exponential model is about the same. So the posterior probability of the three exponential model (which is related to the product of these two) is low because the prior probability is low. Bayesian model selection is essentially a quantitative statement of Occams razor: when two models fit the evidence equally well, prefer the simpler model.

Example – Multiple Nonstationary Frequencies

Often an experimenter is faced with a data set that looks like Fig. 3 (A). The problem is to determine how many resonances are present. If the resonances were stationary, the experimenter could sample the signal longer. The discrete Fourier transform would then resolve the resonances. Unfortunately, the resonances are nonstationary, i.e. they decay away with time. Taking data for a longer period of time samples the noise, not the resonances, and no improvement is found. The only recourse is to get the information from the data available. In fact, probability theory indicates there is one thing the experimenter can do: sample the data faster [2], thus obtaining more data in the region where the resonances are big. This gives a $\sqrt{\Delta T}$ improvement in the estimates, where ΔT is the sampling time. But, if the experimenter uses the discrete Fourier transform as his analysis tool, instead of probability theory, this procedure will improve the line shape, but it will not help separate the resonances.

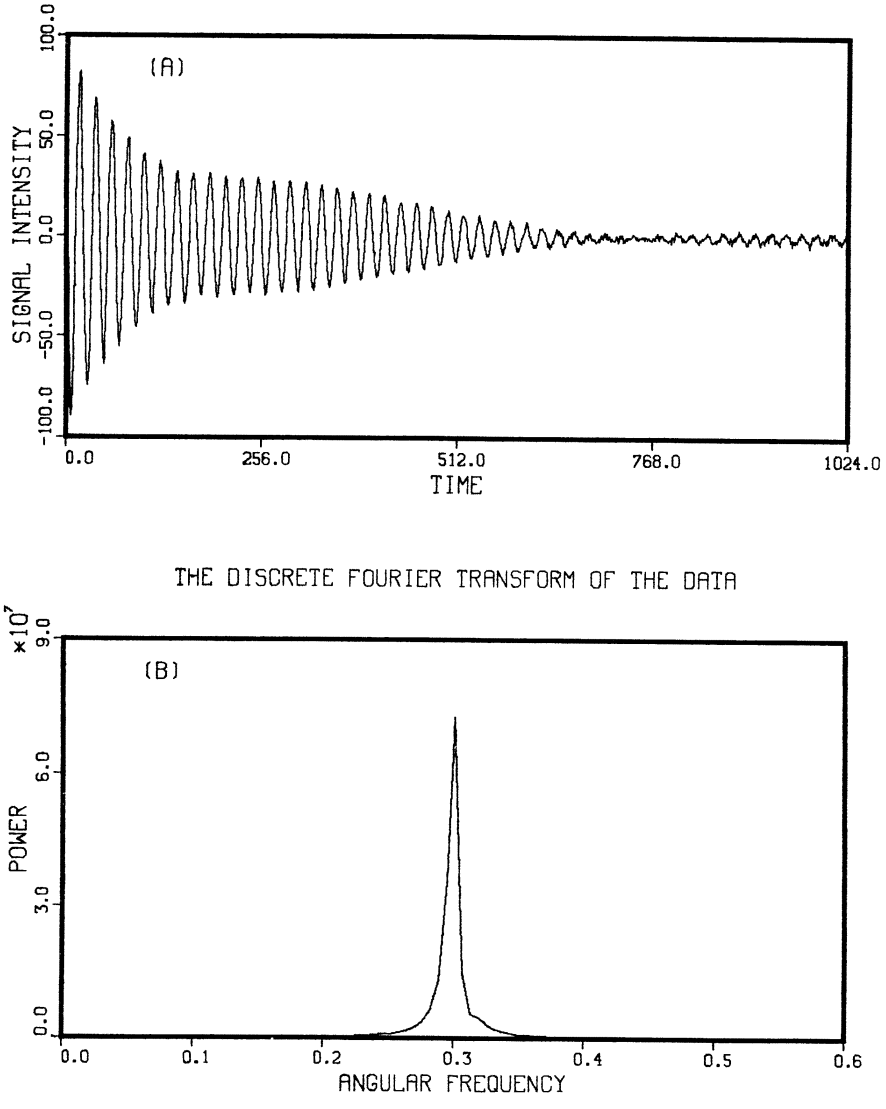
Let us see what probability theory can do on the multiple nonstationary frequency problem. We generated the data, Fig. 3 (A), in such a way that the discrete Fourier transform, Fig. 3 (B), has only a single peak. We then apply the results of the calculation to see if probability theory can determine the number of frequencies present. In this example, we generated the data from

$$d(t_i) = 100\cos(0.3t_i + 1)\mathrm{e}^{-.005t_i} + 25\cos(0.31t_i + 3)\mathrm{e}^{-.003t_i} + n(0,1),$$

with $N = 1024$, signal-to-noise ratio of approximately 20, and $t_i = \{0, 1, \cdots, 1023\}$. The discrete Fourier transform, Fig. 3 (B) has one peak near 0.3. There is no evidence in a discrete Fourier transform for the second frequency. However, we can look at the data and see the beats: the human eye is better at determining the presence of multiple close frequencies than a discrete Fourier transform.

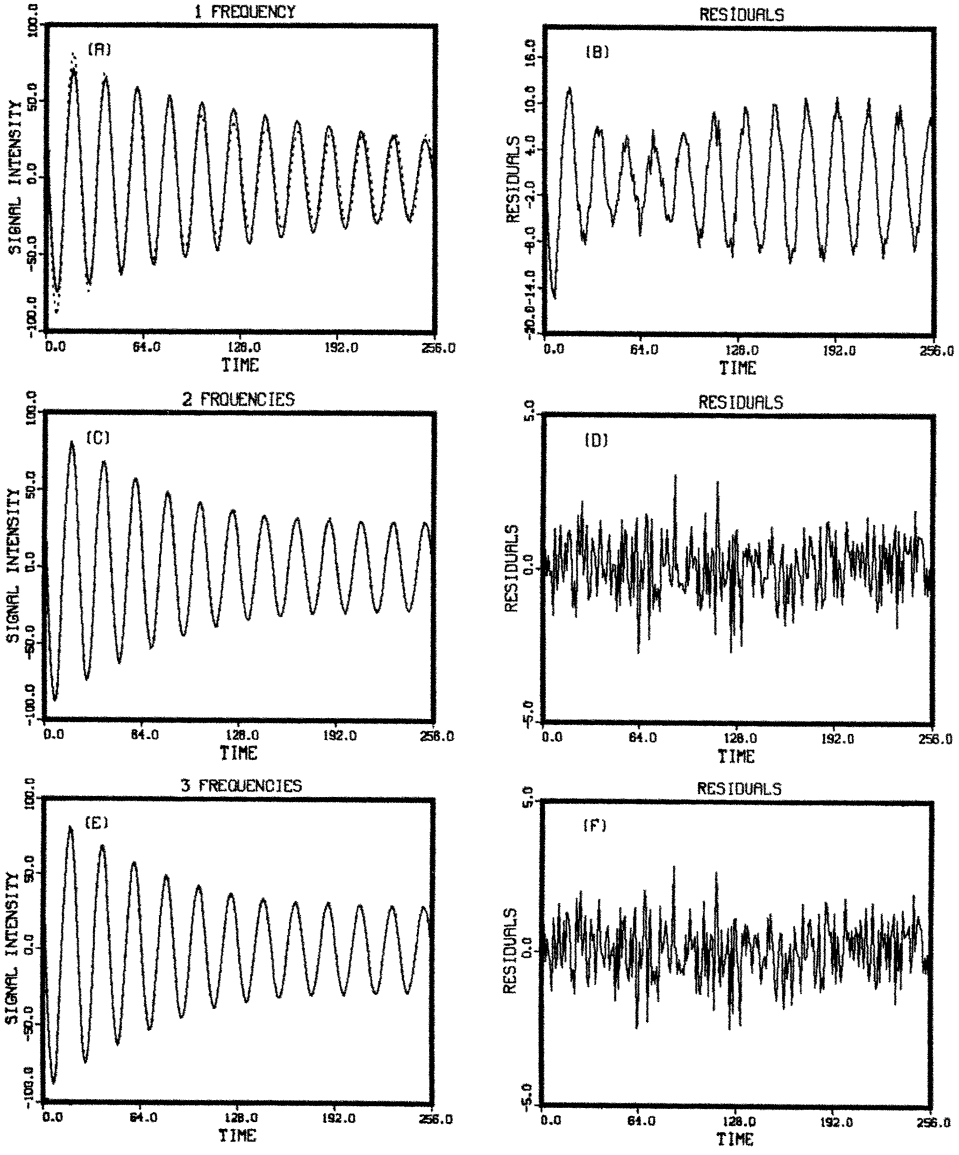
To compute the posterior probability of the model, we must state what set of

Figure 3: Multiple Nonstationary Frequencies



The computer simulated data in (A) contain two nonstationary frequencies. There are $N = 1000$ data values with signal-to-noise ratio of approximately 20. The discrete Fourier transform (B) shows only a single peak.

Figure 4: Multiple Nonstationary Frequencies



The data, (A) dotted line, were fit to a one frequency with decay model, (A) solid line. The residuals are shown in (B). The posterior probability of this model is zero to 690 places! We then fit a two frequency with decay model, (C) solid line. The residual are shown in (D). The posterior probability of this model is 1 to eight places. We then fit a three frequency model, (E) solid line. The residuals are shown in (F). The posterior probability of this model is zero to eight places.

models is to be examined. Here we used

$$f_j(t) = \sum_{k=1}^{j} (B_k \cos \omega t + B_{k+j} \sin \omega t)\, e^{-\alpha_k t}, \qquad (j = 1, 2, \cdots) \tag{8}$$

as the nonorthogonal models. The question we would like to ask is: "What value of j is needed to adequately describe the data?" To do this calculation we again fit a model containing one, two, three, etc. frequencies until the posterior probability of the data had a well defined peak.

We again displaye the results of this calculation as six plots, Fig. 4. The data (dotted line) are shown in (A), (C), and (E). We have included the model as a solid line; a one frequency model is shown in (A), a two frequency model in (C), and a three frequency model in (E). The residuals for the three models are shown in (B), (D), and (F) respectively.

We begin by fitting a one frequency with decay model. We compute the nonorthogonal model, Eq. (8), and then the orthogonal model, Eq. (1). From the orthogonal model we compute the global likelihood of the data, Eq. (7), and last we compute the posterior probability, Eq. (4), using a uniform prior on the models. The posterior probability of the one frequency with decay is zero to 690 decimal places, strong evidence indeed! Note the logarithm of the posterior probability increases like the number of data values. Here we have $N = 1024$ data values. Additionally, for this model, each sampled value significantly misfits the data: the data are very improbable in view of this model; the posterior probability of the model is extremely low.

We then fit a two frequency model to the data, Fig. 4 (C) – the residuals are shown in (D). Now the model and the data are essentially identical: the residuals (D) look like white noise. We then compute the posterior probability of this model and find it to be 1 to thirteen decimal places. Of course, all we knew at this point was that the two frequency model was strongly preferred to the one frequency model, so we proceeded to the three frequency model.

We then computed the three frequency model. The data and the model are displayed in Fig. 4 (E), and the residuals are shown in Fig. 4 (F). The model does not fit the data any better than one would expect from fitting the noise. The posterior probability of this model is zero to thirteen places. In this example not only does probability theory find the correct number of resonances, but also the evidence in these data is overwhelmingly in favor of the two frequency with decay model.

Conclusions

We have demonstrated in these two examples that Bayesian probability theory is capable of giving a quantitative interpretation to Occam's razor. These procedures are readily implemented and work well under conditions where more standard procedures fail. The multiple nonstationary frequency example illustrated that the human eye

is a better tool for determining the presence of multiple resonances than the discrete fourier transform. In data where the human eye outperforms the discrete Fourier transform, the Bayesian calculation gives overwhelming evidence for the frequencies. In the multiple decaying exponential example the human eye is no better than the more traditional techniques. However, the Bayesian analysis works under conditions where the more traditional tests fail and gives strong evidence in support of the correct hypothesis.

Acknowledgments

This work was partially supported by NIH grant GM-30331, J. J. H. Ackerman principal investigator, and by a gift from Nabisco brands. The encouragement of Dr. J. J. H. Ackerman and Professor E. T. Jaynes is greatly appreciated.

References

[1] Bretthorst G. L., (1987), Bayesian Spectrum Analysis and Parameter Estimation, Ph.D. thesis, Washington University, St. Louis, MO.; available from University Microfilms Inc., Ann Arbor, Mich.

[2] Bretthorst, G. L., (1988), Bayesian Spectrum Analysis and Parameter Estimation, in *Lecture Notes in Statistics,* Vol. 48, Springer-Verlag, New York, New York

[3] Jeffreys, H., (1939), Theory of Probability, Oxford University Press, London, (Later editions, 1948, 1961).

[4] Zellner, A., (1980), in Bayesian Statistics, J. M. Bernardo, ed., Valencia University Press, Valencia, Spain.

REGULARIZATION AND INVERSE PROBLEMS

SIBUSISO SIBISI

Department of Computational and Applied Mathematics
University of Witwatersrand
Johannesburg
South Africa

ABSTRACT. We present an overview of the regularization of inverse problems from a Bayesian viewpoint. We derive a Bayesian principle for the optimal choice of regularizing parameter. We apply the principle to zeroth order quadratic regularization and extend it to also determine the optimal derivative order for higher order regularization of convolution problems. We also briefly discuss the Generalized Cross-Validation method for choosing the optimal regularization parameters. We present numerical results for a severely ill-posed problem and for the well-posed Fourier problem.

1. INTRODUCTION

The regularization or smoothing approaches to the treatment of inverse problems, density estimation, ridge regression and spline smoothing bear a lot in common. Titterington [8] provides an extensive review. As practitioners in these fields know, the choice of smoothing or regularizing parameter can be critical, in particular for severely ill-posed problems. The approach most commonly used in the practical treatment of inverse problems (e.g image reconstruction), is the *discrepancy principle* which determines the parameter from the requirement that the reconstructed data misfit must match the noise level. A major drawback is that a good estimate of the noise level σ^2 must be available. Two alternative methods which do not require prior knowledge of the noise level have received considerable attention, particularly in the statistical literature. The one method, generalized cross validation (GCV), has been most extensively applied in non-parametric spline smoothing [1,7] and related areas [5,10]. Its application to large scale image processing problems has also been suggested [6,9]. The other method bears the name of generalized maximum likelihood (GML) [11] and it has, amongst other areas, been applied to deconvolution problems [2,3,4]. In addition to finding the optimal smoothing parameter, both methods give posterior estimates of the noise level. Despite the success of these techniques, they have not become a common tool in regularization practice.

The second section gives a derivation of a Bayesian principle for determining the smoothing parameter and a posterior estimate of the noise level. For the special case of a quadratic smoothing function, the principle is equivalent to GML but our derivation differs from Wahba's [11]. We also discuss the treatment of spatial correlations in the solution through including derivative terms in the smoothing function. In addition to determining the smoothing parameter, the Bayesian principle can be used to determine the optimal derivative order. Details are presented for the convolution problem. The reader is also referred to the work of Davies et al [2,3,4]. The third section gives some details of GCV and the fourth section presents numerical investigations.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 389–396.

2. A BAYESIAN PRINCIPLE

The desired principle can be derived by straghtforward application of Bayes' theorem. Let α be the smoothing parameter, σ the noise level, f the solution and D the available data. The joint *posterior* probability of α, σ, f and D given a *model* m for f is

$$\begin{aligned}\Pr(\alpha,\sigma,f,D|m) &= \Pr(\alpha,\sigma)\Pr(f,D|\alpha,\sigma,m)\\ &= \Pr(\alpha,\sigma)\Pr(f|\alpha,m)\Pr(D|f,\sigma)\end{aligned} \tag{1}$$

Also,

$$\Pr(\alpha,\sigma,f,D|m) = \Pr(D)\Pr(\alpha,\sigma,f|D,m) \tag{2}$$

Then, the posterior of α, σ and f given D and m is

$$\Pr(\alpha,\sigma,f|D,m) = \frac{\Pr(\alpha,\sigma)}{\Pr(D)}\ \Pr(f|\alpha,m)\Pr(D|f,\sigma) \tag{3}$$

Since we have a single fixed dataset, $\Pr(D)$ is a constant and our ignorance of the parameters α and σ is expressed by a flat *prior* $\Pr(\alpha,\sigma)$. Hence the posterior for α and σ, obtained by *marginalizing* eq.(3) over f, is

$$p \equiv \Pr(\alpha,\sigma|D,m) \propto \int d^N f \Pr(f|\alpha,m)\Pr(D|f,\sigma) \tag{4}$$

and it remains to perform the integration over f and then find the optimal values of α and σ by maximizing p over these parameters. To perform the integration, we must provide explicit expressions for the prior $\Pr(f|\alpha,m)$ and the *likelihood* $\Pr(D|f,\sigma)$.

The likelihood of independent data with Gaussion noise is

$$\Pr(D|f,\sigma) = e^{-C(f)/2\sigma^2}/Z_C(\sigma) \tag{5}$$

where $C(f) = (Of - D)^T(Of - D)$, O is the experimental response matrix and

$$Z_C(\sigma) = \int d^n D e^{-C(f)/2\sigma^2} \propto \sigma^n \tag{6}$$

The prior for f, given a smoothing function $S(f,m)$, is

$$\Pr(f|\alpha,m) = e^{\alpha S(f,m)}/Z_S(\alpha)\ ; \quad Z_S(\alpha) = \int d^N f e^{\alpha S(f,m)} \tag{7}$$

$Z_S(\alpha)$ can be explicitly evaluated in the general quadratic case $S = -\frac{1}{2}\,(f-m)\cdot\Omega\cdot(f-m)$ where Ω is a non-negative definite matrix. Ignorance of f may be expressed by a flat model, which, without loss of generality, can be taken to be zero in the quadratic case. Hence

$$Z_S(\alpha) = \int d^N f e^{-\frac{1}{2}\alpha f\cdot\Omega\cdot f} \propto \alpha^{-N/2}(\det\Omega)^{-1/2} \tag{8}$$

Since $\det\Omega$ is independent of α and σ, eq.(4) becomes

$$p \propto \sigma^{-n}\alpha^{N/2}\int d^N f e^{-Q(f)/\sigma^2} \tag{9}$$

where $Q = C/2 - \beta S$ and $\beta = \alpha\sigma^2$. Let $\hat{f}$ maximize Q at given α, σ. Then $Q(f) = Q(\hat{f}) + \frac{1}{2}\delta f \cdot \nabla\nabla Q \cdot \delta f \quad (\delta f = f - \hat{f})$, and

$$\int d^N f e^{-Q(f)/\sigma^2} = e^{-Q(\hat{f})/\sigma^2} \int d^N f e^{-f\cdot(\nabla\nabla Q/\sigma^2)\cdot f}$$

$$\propto (\sigma^2)^{-N/2} e^{-Q(\hat{f})/\sigma^2} (\det \nabla\nabla Q)^{-\frac{1}{2}} \tag{10}$$

Then

$$p = \text{ const. } \sigma^{-n} e^{-Q(\hat{f})/\sigma^2} \beta^{N/2} (\det \nabla\nabla Q)^{-1/2}$$

or

$$\log p = \text{ const. } - n \log \sigma - \frac{Q(\hat{f})}{\sigma^2} + \frac{N}{2} \log \beta - \frac{1}{2} \log \det \nabla\nabla Q \tag{11}$$

Now, Q depends only on β, which may be treated as the smoothing parameter and varied independently of σ^2. So, we can readily maximize $\log p$ with respect to σ to obtain the posterior noise estimate

$$\hat{\sigma}^2 = 2Q(\hat{f})/n \tag{12}$$

Substituting this expression for σ in eq.(11) gives

$$\log p = \text{ const. } - \frac{n}{2} \log Q(\hat{f}) + \frac{N}{2} \log \beta - \frac{1}{2} \log \det \nabla\nabla Q \tag{13}$$

This expression is equivalent to Wahba's GML [11]. $\nabla\nabla Q = O^T O + \beta\Omega$ may have zero eigenvalues, in which case $\det \nabla\nabla Q$ may be taken to be the product of the non-zero eigenvalues only. The optimal value of β, β^*, is then found by maximizing eq.(13) with respect to β. Correspondingly, the optimal noise estimate, σ^*, is found by evaluating $\hat{\sigma}$ at β^* and the optimally regularized solution, f^*, is found by evaluating $\hat{f}$ at β^*.

In the next section, we set Ω to the unit matrix and retain a flat (zero) model. Since there are also no derivative terms in S in the case considered, we refer to it as the zeroth order case. This S contains no spatial correlation structure.

(a) Regularization of Order Zero

We seek explicit expressions in terms of β for $\hat{f}$, $\hat{\sigma}$ and $\log p$. For this purpose, it is convenient to introduce the singular value decomposition of the response matrix O given by $O = \sum \lambda_i u_i v_i^T$ where $\{v_i\}$ and $\{u_i\}$ are respectively the orthonormal eigenvectors of $O^T O$ and OO^T with eigenvalues $\{\lambda_i^2\}$. Then we can readily show that

$$\hat{f} = (\beta I + O^T O)^{-1} O^T D = \sum_{i=1}^{N} \frac{\lambda_i d_i v_i}{\lambda_i^2 + \beta}\,; \quad d_i = u_i \cdot D \tag{14}$$

thus

$$Q(\hat{f}) = \frac{1}{2} \sum_i \frac{\beta d_i^2}{\lambda_i^2 + \beta} \quad \text{and} \quad \det \nabla\nabla Q = \prod_{i=1}^{N} (\lambda_i^2 + \beta) \tag{15}$$

Then $\hat{\sigma}^2$ can be readily calculated from eq.12 and eq.(13) becomes

$$\log p = \text{ const. } - \frac{n}{2} \log \sum_i \frac{\beta d_i^2}{\lambda_i^2 + \beta} + \frac{1}{2} \sum_i \log \left(\frac{\beta}{\lambda_i^2 + \beta} \right) \tag{16}$$

It follows from eq.(16), that $\log p$ does not have a maximum at finite β in the ideally well-posed case of orthogonal O for which all λ_i^2 are the same (e.g. the Fourier problem). This is to be expected since zeroth order quadratic regularization stabilizes (smooths) an ill-posed problem, but does not take into account spatial correlation in the solution. One way of incorporating correlations in quadratic regularization is to allow S to be a function of derivatives of f. The derivative order, s, could, like σ and β, also be determined by the Bayesian method. Only for non-zero s can a quadratic S with a flat model provide meaningful regularization for the ideally well-posed case. But such an approach need by no means be restricted to this case, it may be applied to ill-posed problems where spatial correlation is considered appropriate. In the next section we consider s-order regularization in the convolution case.

(b) Regularization of Order s: the Convolution Problem

Given a blurring function b and noisy data D we seek f such that $D = b * f$ or, equivaletly, $d = BF$ where $d = \mathcal{F}D$, $B = \mathcal{F}b$ and $F = \mathcal{F}f$, $\mathcal{F}$ being the Fourier transform. It is convenient to consider the continuous case first. The Fourier pair $f(\omega)$ and $F(t)$ are related as follows

$$F(t) = \int dw e^{-iwt} f(\omega) \quad \text{and} \quad f(\omega) = \int dt e^{iwt} F(t) \tag{17}$$

(factors of 2π are unimportant). So

$$f^{(s)}(\omega) = \frac{d^s f}{dw^s} = \int dt (it)^s e^{iwt} F(t)$$

Let $S(f)$ be a combination of zero order and s order terms

$$S(f) = -\frac{1}{2} \int d\omega [\delta (f(\omega))^2 + (f^{(s)}(\omega))^2] = -\frac{1}{2} \int dt [\delta + t^{2s}] (F(t))^2 \tag{18}$$

where δ is a (non-negative) number. Choosing $\delta = 1$ and discretizing gives

$$S = -\frac{1}{2} \sum_{j=1}^{n} \left(f_j^2 + \left(f_j^{(s)} \right)^2 \right) = -\frac{1}{2} \sum_{k=1}^{n} \left(1 + t_k^{2s} \right) F_k^2 \tag{19}$$

This S may also be written in the form $S = -\frac{1}{2} f \cdot \Omega \cdot f$. Here, $\Omega = \mathcal{F}^T X \mathcal{F}$ where $X = I + \text{diag}(t_k^{(2s)})$. Then the rest of the analysis is similar to that of the preceding section. The 'partition function' for the prior $\Pr(f|\alpha, s) = e^{\alpha S}/Z_S(\alpha, s)$ is

$$Z_S(\alpha, s) = \int d^n f e^{\alpha S} \propto \alpha^{-n/2} \prod_{i=1}^{n} \left(1 + t_i^{2s} \right)^{-\frac{1}{2}} \tag{20}$$

(The choice $\delta = 0$ in eq.(18) can lead to unbounded $Z_S(\alpha)$ if the sample point $t_i = 0$ is included, resulting in a partially improper prior.) Let $\hat{F} = \mathcal{F}\hat{f}$, where, as before, $\hat{f}$ is the solution at fixed β. Then $\hat{F}$ is given by

$$\hat{F}_k = \frac{B_k d_k}{B_k^2 + \beta(1 + t_k^{2s})} \quad k = 1 \ldots n \tag{21}$$

which leads to

$$Q(\hat{f}) = \frac{1}{2}\sum_i \frac{\beta\left(1+t_i^{2s}\right)d_i^2}{B_i^2+\beta\left(1+t_i^{2s}\right)} \quad \text{and} \quad \det\nabla\nabla Q = \prod_{i=1}^{N}\left(B_i^2+\beta\left(1+t_i^{2s}\right)\right) \tag{22}$$

and

$$\log p = \text{const.} - \frac{n}{2}\log\sum_i \frac{\beta\left(1+t_i^{2s}\right)d_i^2}{B_i^2+\beta\left(1+t_i^{2s}\right)} + \frac{1}{2}\sum_i \log\left(\frac{\beta\left(1+t_i^{2s}\right)}{B_i^2+\beta\left(1+t_i^{2s}\right)}\right) \tag{23}$$

which must be maximized with respect to s and β to find the optimal values of these parameters.

The Fourier problem is simply a special case of the convolution problem with the Fourier domain data d as the raw data and $B_i = 1 \quad \forall_i$. The equations indicate clearly that, in order to achieve any smoothing, s must be non-zero.

3. GENERALIZED CROSS VALIDATION

Let $\hat{f}^{[k]}$ be the regularized solution with the k^{th} datum omitted and let the corresponding mock data be $\tilde{D}^{[k]} = O\hat{f}^{[k]}$. Then the choice of β is good if the mock datum $\tilde{D}_k^{[k]}$ is a good prediction of the true datum D_k. Thus, the optimal β is the minimizer of the weighted mean-squared prediction error

$$V = \frac{1}{n}\sum_{k=1}^{n} w_k(\tilde{D}_k^{[k]} - D_k)^2 \tag{24}$$

The weights are chosen so as to make V invariant under data rotations [5]. Then V becomes

$$V = \frac{n\left((I-A)D\right)^2}{\left(Trace(I-A)\right)^2} \tag{25}$$

and the noise estimate can be shown [7] to be $\hat{\sigma}^2 = VTrace(I-A)/n^2$ where $A = O(\nabla\nabla Q)^{-1}O^T$. In the case of order zero regularization

$$\left((I-A)D\right)^2 = \left(\sum \frac{\beta d_j}{\lambda_j^2+\beta}\right)^2 \quad \text{and} \quad Trace(I-A) = \sum \frac{\beta}{\lambda_j^2+\beta} \tag{26}$$

leading to corresponding expressions for V and $\hat{\sigma}^2$. The expressions for the case of order s regularization for the convolution problem can be obtained by simply replacing λ_i^2 in eq.26 by $B_i^2/\left(1+t_i^{2s}\right)$ (similarly, eq.23 could have been directly obtained from eq.16 by the same replacement).

4. NUMERICAL RESULTS

(a) A Severely Ill-posed Problem (Order zero)

We consider a one-dimensional problem with $n = N = 15$ and the classic ill-conditioned Hilbert transformation matrix $O_{ij} = 1/(i+j-1)$. The true solution is the

simulated top hat shown in fig.1(a). The maximum magnitude of the corresponding datapoints is $O(1)$. Gaussian noise of mean zero and standard deviation $\sigma = 1.0 \times 10^{-5}$ is added to the data.

In figs. 1(b), 1(c) and 1(d), the simulation is displayed as dots and the regularized solution as triangles. The regularized solution of fig.1(b), with $\beta = 1.0 \times 10^{-13}$ and a data misfit of $\tilde{C} \equiv C/n\sigma^2 = 1.022$, oscillates wildly and looks nothing like the simulation. The solutions for smaller β, and hence a better fit to the data, oscillate with an even higher dynamic range. Thus the value of β corresponding to the discrepancy constraint $\tilde{C} = 1.0$ leads to a catastrophic result. In fig.1(c), with $\beta = 1.0 \times 10^{-12}$ and $\tilde{C} = 1.030$, the regularized solution begins to resemble the simulation. The solution of fig.1(d), with $\beta = 1.0 \times 10^{-10}$ and $\tilde{C} = 1.032$ looks visually better but larger values of β lead to too much smoothing. Any further improvement in the solution would require additional prior information such as positivity. A visual judgment suggests that the optimal value of β is of the order of 10^{-10}.

The Bayesian optimal β, which maximizes eq.18, is 1.42×10^{-10} while the GCV optimal value, obtained from minimizing eq.29, is 2.92×10^{-11}. Both these values are in the expected range and the two corresponding solutions (not displayed) are hardly distinguishable visually. The one point of concern is that GCV has a second lower minimum at $\beta = 5.7 \times 10^{-22}$ but the highly oscillatory solution is totally unacceptable.

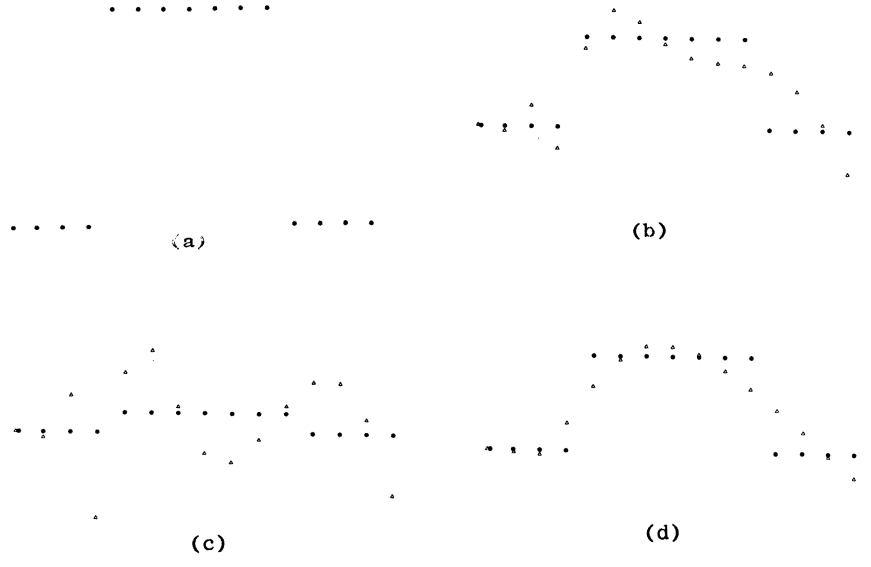
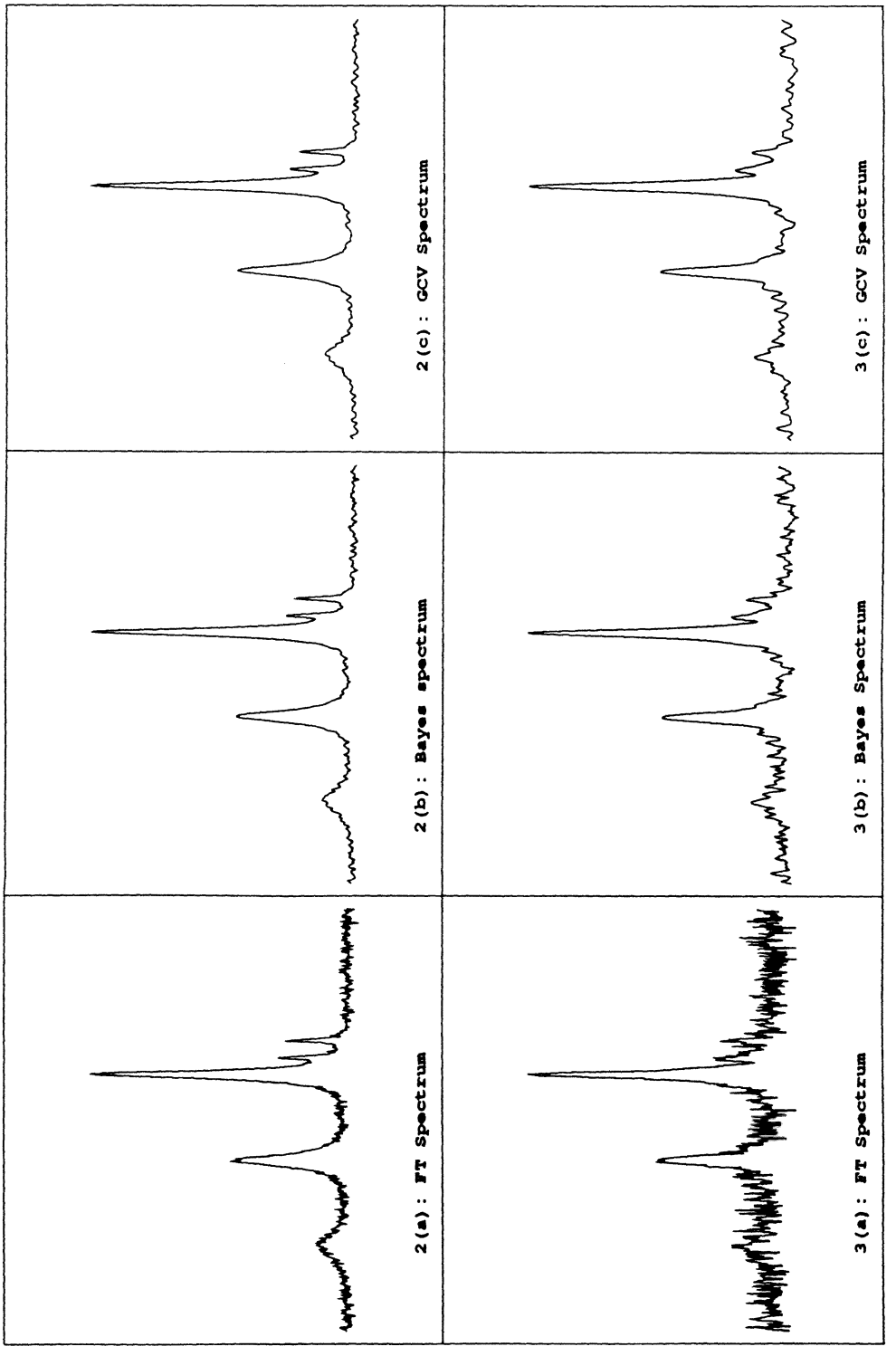


Fig.1 (a) Simulation. Solution with (b)$\beta = 10^{-13}$ (c)$\beta = 10^{-12}$ (d)$\beta = 10^{-10}$.

(b) A Fourier Problem (Order s)

We consider time series data generated by a sum of decaying oscillators of unit total amplitude. The data are corrupted by Gaussian noise with variance σ^2. The Fourier spectrum of Lorentzians is shown in fig.2(a) for $\sigma = 0.005$. Eq.23 (Bayes) has a well-defined maximum at $s = 1.7$ and $\beta = 3.6 \times 10^{-5}$ and the corresponding noise estimate is $\hat{\sigma} = 0.0046$


2(a): FT Spectrum
2(b): Bayes spectrum
2(c): GCV Spectrum
3(a): FT Spectrum
3(b): Bayes Spectrum
3(c): GCV Spectrum

(several simulations were run at different seeds, giving very similar results). V (GCV), on the other hand, is very flat and, correspondingly, the minimum drifts considerably for different seeds. Representative values are $s = 2.5$ and $\beta = 1.0 \times 10^{-6}$. The predicted values of σ lie in the range 0.0002 to 0.00025. The Bayes and GCV spectra generated using the obtained parameters are shown in fig.2(b) and fig.2(c) respectively.

At the higher noise level of 0.02, the Bayes parameters are $s = 1.3$ and $\beta = 1.31 \times 10^{-3}$ with $\hat{\sigma} = 0.019$. GCV gives $s = 1.9$, $\beta = 1.9 \times 10^{-4}$ and $\hat{\sigma} = 0.0009$. The corresponding Fourier, Bayes and GCV spectra are shown in fig.3(a), fig.3(b) and fig.3(c) respectively.

CONCLUSION

In conclusion, we have given a brief but complete derivation of a Bayesian principle for choosing regularizing parameters. We have tested the principle, as well as GCV, on an ill-posed and a well-posed problem. While the results presented do not constitute a thorough comparison of the two methods, we find the Bayesian principle more appealing because of its theoretical justification in Bayesian terms. For a further discussion of the principle, see the paper by Gull in these proceedings.

REFERENCES

[1] Craven, P. and Wahba, G.(1979) Smoothing Noisy Data with Spline Functions: estimating the correct degree of smoothing by the method of generalized cross validation. *Numer. Math.*, **31**, 317-403.

[2] Davies, A.R.(1982) On the Maximum Likelihood Regularization of Fredholm Convolution Equations of the First Kind. in *Treatment of Integral Equations by Numerical Methods (ed. Baker, C.T. and Miller, G.F.)* Academic Press.

[3] Davies, A.R., Iqbal, M., Maleknad, K. and Redshaw, T. C.(1983) A Comparison of Statistical Regularization and Fourier Extrapolation Methods for Numerical Deconvolution. in *Numerical Treatment of Inverse Problems in Differential and Integral Equations (ed. Deulfhard, P. and Hairer, E.)* Birkhauser.

[4] Davies, A.R. and Anderssen, R.S.(1986) Optimization in the Regularization of Ill-posed Problems. *J. Austr. Math. Soc. Ser. B*, **28**, 114-133.

[5] Golub, G.H., Heath, M. and Wahba, G.(1979) Generalized Cross Validation as a Method for Choosing a Good Ridge Parameter. *Technometrics* **21**, 215-224.

[6] Hall, P. and Titterington, D. M.(1986) On Some Smoothing Techniques used in Image Restoration. *J. Roy. Statist. Soc. B*, **48**, 330-343.

[7] Silverman, B.W.(1985) Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting. *J. R. Statist. Soc. B* **47**, 1-52.

[8] Titterington, D.M.(1985a) Common Structure of Smoothing Techniques in Statistics. *Int. Statist. Rev.*, **53**, 141-170.

[9] Titterington, D.M.(1985b) General Structure of Regularization Procedures in Image Reconstruction. *Astron. Astrophys.*, **144**, 381-387.

[10] Wahba, G.(1977) Practical Approximate Solutions to Linear Operator Equations When the Data Are Noisy. *SIAM J. Numer. Anal.* **14**, 651-667.

[11] Wahba, G.(1985) A Comparison of GCV and GML for Choosing the Smoothing Parameter in the Generalized Spline Smoothing Problem. *Ann. Statist.*, **13**, 1378-1402.

Maximum Entropy and Inductive Inference

J.B. Paris & A. Vencovská,
Department of Mathematics,
University of Manchester,
England.

Key Words: Inductive inference, Maximum entropy, Inexact reasoning.

Abstract

The following problem is currently receiving some attention in developing expert systems capable of dealing with uncertainty.

> Given sentences $\theta_1,\ldots,\theta_n$ of the propositional calculus and some linear constraints on the weights of belief attached to those sentences what weight of belief should we give to a new sentence θ from the same language?

Here we are thinking of these weights as some kind of subjective probabilities. We show that if we treat the assignment of a weight to θ as the result of an *inference process* applied to the constraints then there are *logical* requirements of consistency on such a process which force it to agree with the Maximum Entropy Inference Process. This then provides a logical rather than statistical or information theoretic justification for the use of the Maximum Entropy Inference Process.

Introduction

The results presented in this paper were motivated by considering the following problem.

> Given sentences $\theta_1,\ldots,\theta_n$ of the propositional calculus and some linear constraints on the weights of belief attached to those sentences what weight of belief should we give to a new sentence θ from the same language?

In order to make this problem mathematically precise we first introduce some notation.

Let $SL(A_{i_1},\ldots,A_{i_m})$ be all those sentences of the propositional calculus formed using propositional variables $A_{i_1},\ldots,A_{i_m}$ and let $SL = \bigcup_n SL(A_1,\ldots,A_n)$. We shall use θ,φ etc for members of SL.

A function $w:SL \longrightarrow [0,1]$ is said to be a *weight of belief* function if w satisfies

(i) If sentences θ,φ are logically equivalent then $w(\theta) = w(\varphi)$.

(ii) If θ is a theorem of the propositional calculus then

$$w(\theta) = 1 \quad \text{and} \quad w(\neg\theta) = 0.$$

(iii) For sentences θ,φ, if $\theta \wedge \varphi$ is contradictory then

$$w(\theta\vee\varphi) = w(\theta) + w(\varphi).$$

A set of linear constraints is a finite set of identities of the

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 397–403.

form

$$\sum_{j=1}^{n} \alpha_{ij}\, w(\theta_j) = \beta_i,$$

where the α_{ij}, β_i are real, which is consistent in the sense that it is satisfied by some weight of belief function w.

We shall use S, S' etc for sets of linear constraints and $CL(A_{i_1},\ldots,A_{i_n})$ etc for the set of sets of constraints for which each $\theta_j \in SL(A_{i_1},\ldots,A_{i_n})$ etc. We can now restate our problem as

Given $S \in CL$ how should we pick a weight of belief function w to satisfy S?

We wish to point out here that we are not asking the statistical question 'find a best estimate of the true weight of belief function.' Instead we are asking 'what weight of belief function w is a *logical consequence* of the set of constraints S?' A frequent retort to this question is that it is meaningless in that, in general, there will be many such function w satisfying the given constraints. Certainly this is justified when considering a single set of constraints in isolation. However, if we consider the choice of w as an instance of an *inference process* N taking sets of linear constraints to weight of belief functions then it is clear that there are natural requirements of consistency and independence on N which severely limit the possible choices of $N(S)$ for a given set of constraints S.

For example, if S_1 and S_2 are the same up to renaming of propositional variables then 'consistency' requires that $N(S_1)$, $N(S_2)$ should be similarly related. In particular if the propositional variables A_i, A_j are not mentioned in a set of constraints S then $N(S)(A_i)$ and $N(S)(A_j)$ should be equal.

In what follows let N be an inference process i.e. for all $S \in CL$, $w = N(S)$ satisfies the set of constraints S.

The main result of this paper is that there are natural, logical principles which force N to be the Maximum Entropy Inference Process. By 'natural' here we mean principles whose failure would normally by described as an 'inconsistency' in the natural language sense of the word.

The plan of the paper is as follows. We first introduce the principles and derive a straightforward consequence of them. We then state the main result and give a detailed description of the Maximum Entropy Inference Process. We end by pointing out a generalization of the main result.

The Principles

We now present the 'natural' principles referred to above. We have made no effort for economy here preferring instead that each stated principle contains a single idea. We do not know if any of the principles are redundant.

Principle 1 For $S \in CL$, $\theta \in SL$, $N(S)(\theta)$ is continuous in the coefficients α_{ij}, β_i appearing in S.

Justification: Microscopic changes in constraints should not result in macroscopic changes in the beliefs inferred.

Principle 2 If S_1, $S_2 \in CL$ are equivalent on the basis of properties (i), (ii), (iii) of w (i.e. each constraint in S_1 is derivable form constraints in S_2 using properties (i), (ii), (iii) of w and visa-versa) then $N(S_1) = N(S_2)$.

Justification: The way the constraints are expressed should be irrelevant to the inference process.

Principle 3 Let $B(A_{i_1},\ldots,A_{i_n})$ be the Boolean Algebra of equivalence classes $\overline{\theta}$ of elements of $SL(A_{i_1},\ldots,A_{i_n})$ with respect to the relation $\equiv$ of logical equivalence. Let $g:B(A_{i_1},\ldots,A_{i_n}) \cong B(A_{j_1},\ldots,A_{j_n})$ and suppose that $S_1 \in CL(A_{i_1},\ldots,A_{i_n})$ $S_2 \in CL(A_{j_1},\ldots,A_{j_n})$ are such that

$$S_1 = \left\{ \sum_{j=1}^{k} \alpha_{ij}\, w(\theta_j) = \beta_i \mid i = 1,\ldots,n \right\}$$

$$S_2 = \left\{ \sum_{j=1}^{k} \alpha_{ij}\, w(\varphi_j) = \beta_i \mid i = 1,\ldots,n \right\}$$

where $\overline{\varphi}_j = g(\overline{\theta}_j)$ $j = 1,\ldots,k$.

Then $N(S_1)(\theta) = N(S_2)(\varphi)$ whenever $\overline{\varphi} = g(\overline{\theta})$.

Justification: S_2 is simply a renamed version of S_1 and so $N(S_1)$, $N(S_2)$ should agree up to this renaming since the particular names chosen should be irrelevant to the inference process.

Principle 4. Let $S_1 \in CL(A_{i_1},\ldots,A_{i_n})$, $S_2 \in CL(A_{j_1},\ldots,A_{j_m})$ with $\{i_1,\ldots,i_n\} \cap \{j_1,\ldots,j_m\} = \emptyset$. Then for $\theta \in SL(A_{i_1},\ldots,A_{i_n})$

$N(S_1)(\theta) = N(S_1 + S_2)(\theta)$.
Justification: The belief constraints given by S_2 have no effect on S_1 or θ and so should be irrelevant to the weight of belief assigned to θ.

Principle 5 Suppose S_1, S_2 are respectively the sets of constraints

$$\sum_j \alpha_{ij}\, w(\theta_j \wedge \varphi) = \beta_i, \quad w(\varphi) = \gamma, \quad \sum_r \delta_{kr}\, w(\psi_r \wedge \neg\varphi) = \nu_k$$

$$\sum_j \alpha_{ij}\, w(\theta_j \wedge \varphi) = \beta_i, \quad w(\varphi) = \gamma, \quad \sum_q \tau_{sq}\, w(\eta_q \wedge \neg\varphi) = \lambda_s.$$

Then for $\theta \in SL$, $N(S_1)(\theta \wedge \varphi) = N(S_2)(\theta \wedge \varphi)$.

Justification: Under the assumption that φ holds S_1, S_2 are providing identical constraints on belief.

Principle 6: Suppose that S_1, $S_2 \in CL$ and that $N(S_1)$ satisfies the constraints in S_2. Then $N(S_1) = N(S_1 + S_2)$.

Justification: If on the basis of S_1 the inference process gave that the constraints in S_2 held then knowing S_2 should not cause the inference process to change its mind i.e. the inference processes should be self consistent.

Principle 7: Let S_1 be $w(A_1) = \alpha$, $w(A_2) = \beta$ and let S_2 be

$S_1 + w(A_3) = \gamma$, $w(A_4) = \delta$, $w(A_2 \wedge A_3) = w(A_2 \wedge A_4) = w(A_3 \wedge A_4) = 0$.

Then $N(S_1)(A_1 \wedge A_2) = N(S_2)(A_1 \wedge A_2)$.

Justification: Whilst S_2 puts additional constraints on A_2 these appear irrelevant to the value N should give $A_1 \wedge A_2$.

This concludes our list of principles. Before stating our main theorem we show:-

Lemma The following principle is a consequence of principles 1-7:

If S is $w(A_1) = \alpha$, $w(A_1 \wedge A_2) = \beta$, $w(A_1 \wedge A_3) = \gamma$ then

$\alpha N(S)(A_1 \wedge A_2 \wedge A_3) = \beta\gamma$.

Sketch proof. For S_1 being $w(A_1) = \beta$, $w(A_2) = \gamma$ let

$N(S_1)(A_1 \wedge A_2) = h(\beta, \gamma)$. Then h is continuous by principle 1.

Given $0 < \alpha, \beta, \delta < 1$ consider $x(t) = N(S_2(t))(A_1 \wedge A_3)$ where

$S_2(t)$ is $w(A_1) = \alpha + \beta$, $w(A_2) = \delta$, $w(A_3) = t$, $w(A_4) = 1$. Then

$x(0) = 0$, $x(1) = \alpha + \beta$ so for some $0 < t_0 < 1$, $x(t_0) = \beta$. Let

$S_2 = S_2(t_o)$. Define an automorphism g of $B(A_1, A_2, A_3, A_4)$ such that for the corresponding equivalence classes

$$g(A_1 \wedge A_2 \wedge A_3 \wedge A_4) = A_1 \wedge \neg A_2 \wedge A_3 \wedge \neg A_4$$

$$g(A_1 \wedge A_2 \wedge \neg A_3 \wedge A_4) = A_1 \wedge A_2 \wedge \neg A_3 \wedge \neg A_4$$

$$g(A_1 \wedge \neg A_2 \wedge A_3 \wedge A_4) = \neg A_1 \wedge \neg A_2 \wedge A_3 \wedge \neg A_4$$

$$g(A_1 \wedge \neg A_2 \wedge \neg A_3 \wedge A_4) = \neg A_1 \wedge A_2 \wedge \neg A_3 \wedge \neg A_4$$

$$g(\neg A_1 \wedge A_2 \wedge A_3 \wedge A_4) = A_1 \wedge \neg A_2 \wedge \neg A_3 \wedge A_4$$

$$g(\neg A_1 \wedge A_2 \wedge \neg A_3 \wedge A_4) = A_1 \wedge \neg A_2 \wedge \neg A_3 \wedge \neg A_4$$

$$g(\neg A_1 \wedge \neg A_2 \wedge A_3 \wedge A_4) = \neg A_1 \wedge \neg A_2 \wedge \neg A_3 \wedge A_4$$

$$g(\neg A_1 \wedge \neg A_2 \wedge \neg A_3 \wedge A_4) = \neg A_1 \wedge \neg A_2 \wedge \neg A_3 \wedge \neg A_4 \quad \text{etc.}$$

Then, to within a use of principle 2, $g(S_2') = S_3$ where S_3 is $w(A_1) = \delta$, $w(A_2) = \alpha$, $w(A_3) = \beta$, $w(A_4) = t_o - \beta$, $w(A_2 \wedge A_3) = w(A_2 \wedge A_4) = w(A_3 \wedge A_4) = 0$ and $S_2' = S_2 + w(A_1 \wedge A_3) = \beta$.

It follows by principles 3, 4, 6, 7 that

$$h(\alpha, \delta) = N(S_3)(A_1 \wedge A_2) = N(S_2)(A_1 \wedge A_2 \wedge \neg A_3)$$

$$h(\beta, \delta) = N(S_3)(A_1 \wedge A_3) = N(S_2)(A_1 \wedge A_2 \wedge A_3)$$

whilst $h(\alpha + \beta, \delta) = N(S_2)(A_1 \wedge A_2)$. Hence

$h(\alpha + \beta, \delta) = h(\alpha, \delta) + h(\beta, \delta)$ and so $h(x, y) = xy$ by continuity.

Now for S being as above let $N(S)(A_1 \wedge A_2 \wedge A_3) = q(\alpha, \beta, \gamma)$. Let $C_1, \ldots, C_{2^n}$ be all the conjuncts of the $A_4, A_5, \ldots, A_{4+n}$ or their negations and let $E = \bigvee_{i=1}^{m} C_i$. Then by principle 3

$N(S)(A_1 \wedge E) = \frac{\alpha m}{2^n}$, $N(S)(A_1 \wedge A_2 \wedge E) = \frac{\beta m}{2^n}$, $N(S)(A_1 \wedge A_3 \wedge E) = \frac{\gamma m}{2^n}$

$N(S)(A_1 \wedge A_2 \wedge A_3 \wedge E) = \frac{m}{2^n} q(\alpha, \beta, \gamma)$. Also by principles 2, 4, 5 we obtain these same answers from $N(S_4)$ and $N(S_5 + w(E) = 1)$ where S_4 is $w(A_1 \wedge E) = \frac{\alpha m}{2^n}$, $w(A_1 \wedge A_2 \wedge E) = \frac{\beta m}{2^n}$, $w(A_1 \wedge A_3 \wedge E) = \frac{\gamma m}{2^n}$ $w(E \wedge \neg A_1) = 1 - \frac{\alpha m}{2^n}$ and S_5 is $w(A_1) = \frac{\alpha m}{2^n}$, $w(A_1 \wedge A_2) = \frac{\beta m}{2^n}$, $w(A_1 \wedge A_3) = \frac{\gamma m}{2^n}$. But then by principle 4

$$q\left[\frac{\alpha m}{2^n}, \frac{\beta m}{2^n}, \frac{\gamma m}{2^n}\right] = N(S_5)(A_1 \wedge A_2 \wedge A_3) = N(S_4)(A_1 \wedge A_2 \wedge A_3)$$

$$= N(S_4)(A_1 \wedge A_2 \wedge A_3 \wedge E) = N(S)(A_1 \wedge A_2 \wedge A_3 \wedge E)$$

$= \frac{m}{2^n} q(\alpha, \beta, \gamma)$. Hence by continuity for $\alpha \neq 0$

$q(\alpha, \beta, \gamma) = \alpha q(1, \beta/\alpha, \gamma/\alpha) = \alpha h(\beta/\alpha, \gamma/\alpha) = \frac{\beta\gamma}{\alpha}$ as required.

In fact it follows by results in [1], [2] (or directly) that principle 7 follows from principles 1-6 and the lemma. We prefer principle 7 since it appears more intuitively obvious than the lemma although of course the latter is clearly justifiable on probabilistic grounds.

The main theorem now follows by a proof given in [2].

Main Theorem The Maximum Entropy Inference Process is the only inference process satisfying principles 1-7.

We now describe the Maximum Entropy Inference Process, ME. Given $S \in CL$ and $\theta \in SL$ we define $ME(S)(\theta)$ as follows. Let m be such that $\theta \in SL(A_1,\ldots,A_m)$ and $S \in CL(A_1,\ldots,A_m)$. Let $C_1,\ldots,C_{2^m}$ list all sentences

$$A_1^{\epsilon_1} \wedge A_2^{\epsilon_2} \wedge \ldots \wedge A_m^{\epsilon_m} \qquad \epsilon_1, \epsilon_2, \ldots, \epsilon_m \in \{0,1\}$$

where $A^0 = A$, $A^1 = \neg A$. Then by the disjunctive normal form theorem and using properties (i), (ii), (iii) of w S can be expressed as a system of linear equations

$$B(w(C_1),\ldots,w(C_{2^m}))^T = (b_1,\ldots,b_n)^T.$$

By the assumed consistency of S,

$$\vec{x} > 0, \quad B(x_1,\ldots,x_{2^m})^T = (b_1,\ldots,b_n)^T, \quad \sum x_i = 1$$

has a solution. Let $(\rho_1,\ldots,\rho_{2^m})$ be the solution to these constraints for which the function

$$-\sum_{i=1}^{2^m} x_i \log(x_i)$$

is maximal. (For a proof that there is a unique such maximum point we refer the reader to [5]. Let θ be equivalent to the disjunctive normal form

$\bigvee_{k=1}^{s} C_{i_k}$ where the i_k are distinct and set $ME(S)(\theta) = \sum_{k=1}^{s} \rho_{i_k}$.

We remark that this value is independent of m, subject to the above requirements.

Remarks

There are of course many other axiomatizations of the the 'Maximum Entropy Principle' in other areas of mathematics, e.g. [4], [5], [6]. In our defence we wish to remark that our original intention was to list thoses principles of inexact reasoning which appeared to us self evident and to go on to classify the inference processes satisfying them. It was in fact to our great surprise that we discovered that there was only one!

The main theorem can be generalized to wider classes of constraints (for example by including inequalities) providing that the set of solutions $w(C_1),\ldots,w(C_{2^m})$, (in the notation above) form a closed, non-empty convex set. To see this suppose that there was a process N satisfying the principles but not agreeing with the ME solution. Then by principle 6 there would be a line segment $[\alpha, \beta]$ within the set $\mathbb{C}$ of possible $w(C_1),\ldots,w(C_{2^m})$ such that

$N([\alpha, \beta]) = \alpha \neq \beta = ME([\alpha, \beta])$. Let $[\alpha_o, \beta_o]$ be the longest extension of $[\alpha, \beta]$ within $\mathbb{C}$. Then by the main theorem

$N([\alpha_o, \beta_o]) = ME([\alpha_o, \beta_o]) = \beta_1$, say, with β lying between α and β_1. But then by principle 6 $N([\alpha, \beta_1]) = \beta_1$. By continuity $N([\alpha, \gamma]) = \beta$ some γ between β and β_1. But then by principle 6 $N([\alpha, \beta]) = \beta$ which is a contradiction.

The main theorem has both positive and negative aspects. On the postivie side it enables us to develop a limited 'inexact proof theory' by treating the principles as rules of proof (see [3]) which provides a candidate for human inexact reasoning. On the negative side such an inference process appears computationally unfeasible (see [1]).

References

[1] J.B. Paris & A. Vencovská, "On the Applicability of Maximum Entropy to Inexact Reasoning", to appear in the International Journal of Approximate Reasoning.

[2] J.B. Paris & A. Vencovská, "A Note on the Inevitability of Maximum Entropy", submitted to the International Journal of Approximate Reasoning.

[3] J.B. Paris & A. Vencovská, "Inexact and Inductive Reasoning", to appear in the Proceedings of the 8th International Congress of Logic, Methodology and Philosophy of Science, Moscow 1987.

[4] C.F. Shannon & W. Weaver, "The Mathematical Theory of Communication", University of Illinois Press, Urbana, Illinois, (1949).

[5] J.E. Shore & R.W. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy", IEEE, Vol IT-26, No. 1, 1980, pp26-37.

[6] J. Skilling, "The Axioms of Maximum Entropy", presented at the 6th Maximum Entropy Workshop, Seattle, August 1986.

MAXIMUM SPECIFIC ENTROPY, KNOWLEDGE, ORDERING, AND PHYSICAL MEASUREMENTS

L.Preuss,
74 Feldeggstr.
8008 Zürich
Switzerland.

ABSTRACT. Given a set of deterministic or stochastic observations which relate the values of two (possibly multidimensional) variables, a pair of specific entropies is defined, which evaluate the average efficiency of inferences from one variable to the other. Further, a symmetric specific entropy is introduced which gauges the overall effectivity of the representation used for recording the observations; it is complemented by a measure for the span of the representation. The amount of useful knowledge derivable from a set of observations equals the span of the representation, scaled with its symmetric specific entropy. An extremalisation of specific entropies provides both a general solution of the ordering problem, and an essential insight into the mechanism of physical measurements: ordering minimizes local values of the specific entropy, whilst measuring is a search for neighbourhoods where the product of this entropy with the local density of observations is maximal.

1. A DEFINITION OF KNOWLEDGE

This paper is an introduction to the applications of maximum specific entropy methods to the problems of ordering, clustering, and optimizing the representation of knowledge, as well as to the analysis of physical measurements. The definitions and algorithms presented here rest on Shannon's entropy concept, or information rate. In contradistinction to classical usage, however, the central theme is not the flow of information but the static representation of knowledge. Information is a dynamic concept: it resides in the entropy reduction - due to an external event (e.g. the passage of time) - which accompanies the transition from a larger number of *a priori* possibilities to a smaller number still allowed after the event. Knowledge, on the other hand, is embodied in the static connections which past experience has shown to exist between entities belonging to different sets. Such an ensemble of connections [nexi] cannot in itself provide any information, i.e. entropy reduction, but it allows to transfer a reduction of entropy from one set to another. This is of particular importance when the latter set is directly observable whilst the former is not.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 405–414.

The distinction between information and knowledge can be readily visualized at the receiving end of a communication channel: each incoming letter of a message supplies some information to the receiver, whether the corresponding knowledge, which consists in the connection of the letter with a particular position in the message, is saved for later use or not. In order to read, and possibly to save, the message, the receiver must register the order in which the individual letters arrive; in other words he must have at his disposal a device such as a clock which identifies successive instants in time, and which might be called a source of positional entropy, or negentropy for short. Without it, the receiver ends up with an amorphous heap of letters from which only their frequencies can be inferred, not the message itself. Indeed, its static content consists not of letters, but of connections between letters and instants in time, or positions in the message.

Although the negentropy furnished by the clock (or equivalent means) plays no role in classical communication theory, it becomes essential when the message is transformed into knowledge, i.e. recorded for later use, be it on a piece of paper or in the memory of a computer. Without an investment of negentropy, it is impossible to retrieve a single letter located in one amongst N equiprobable positions on the paper, or amongst N addresses in memory, because either selection necessitates an expenditure of negentropy proportional to ln(N).

Incidentally, an identification of knowledge with the awareness of existing connections between entities corresponds to everyday usage: reading the single word "Skilling" on a visiting card casually picked up from the pavement, provides information, but no knowledge. The association of two words on this same card, however, say "Skilling, Carpenter", conveys the knowledge that there (presumably) exists a carpenter by the name of Skilling. If now the card was not picked up from a nondescript location on the pavement, but found in one's letterbox, then the connection of the same two words with that particular location gives evidence that (again presumably) carpenter Skilling has called during one's absence.

More generally, knowledge is embodied in the relative frequencies of the mutual links between values of at least two discrete variables, where by definition each link is attributed a non-negative weight proportional to the frequency with which the corresponding combination of values was observed. This applies not only to variables defined on a nominal scale as in the above examples, but also to quantitative variables, provided that their values are identified on a discrete scale, which may be either arbitrary or made up of intervals, the width of which is determined by the ultimate resolution of the instruments used for the observation. Thus fig.1 could for instance represent the partial pressure of a gas in- and outside an enclosure which stretches from A to B in a one-dimensional space. On a sufficently fine resolution, fluctuations of the local pressure will become noticeable upon repeated observations, so that the pressure in each location must be identified by a distribution, as indicated in fig.1 by spots of various sizes.

By definition, a primitive observation - hereafter called observation - consists in the selection of a value for an independent variable x, followed by the determination of a single value of a dependent variable y

conditioned by the former. Once a full set of observations has been carried out, each link (i.e.pair of values of the variables) is attributed a non-negative number - or weight - proportional to its frequency of occurrence.

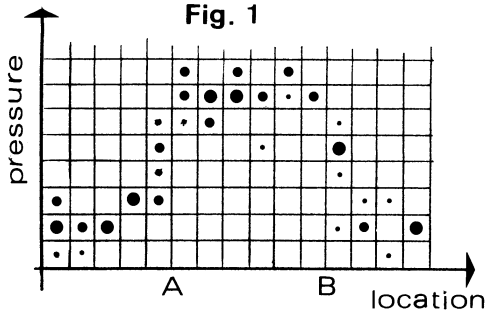
Although it is usual to select each allowed value of x with equal frequency, this is not mandatory; also, the occurrence of several nexi with finite weights for a single value of x need not denote a flawed observation. If, for instance, each x-value identifies a location in an image, and each y-value a colour (more precisely: a small wavelength interval), then all locations in an image which reflect different wavelengths will exhibit several non-vanishing nexi for the same value of x. Hence the multiple occupancy of a value of the independent variable may result from an essential superposition, and need not signal a lack of precision.

When dealing with amounts of knowledge, a distribution which characterizes a multiple occupancy must be handled as an indivisible entity; this requires a measure of the dissimilarity between distributions. A specific entropy will be defined, which provides such a measure, and which - in contradistinction e.g. to Kullback's divergence [1] - does not require the distributions to be mutually absolutely continuous. It is therefore capable to handle problems which are beyond the scope of Kullback's divergence, and of other previous measures.

2. THE EFFICIENCY OF INFERENCE

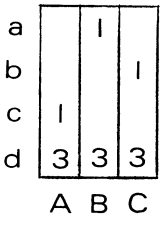
In order to apply notations standard to communication theory to the above representations, the observation of a value of x (i.e. of an item, which is represented by a column of nexi in the tables) will be viewed as a random event, the probability of which is determined by the distribution in the marginal row. Similarly, the values of the dependent variable y (i.e. the features which characterize the items, and are represented by rows in the tables) will be viewed as observables with conditional probabilities represented by the distributions within the

Fig. 1



OBSERVATIONS OF A PRESSURE: SPOT SIZES DENOTE FREQUENCIES OF OBSERVATIONS

Fig. 2



A GROUP WITH A READILY INTUITED SEPARABILITY: $\gamma=1/4$

columns, and with unconditional probabilities which are listed in the marginal column. On average, the amount of information which the observation of a feature provides about the item from which it stems equals the entropy of the marginal row less the average entropy of the rows, i.e. the mutual information:

$$I(x; y) = H(x) - H(x|y) = H(y) - H(y|x) \ .$$

Dividing this classical measure by the entropy reduction H(x) that would be obtained, on average, through a direct identification of the item among the group of all possible ones, yields a dimensionless and normalized coefficient which will be called the separability of the group of items, and designated by γ_X:

$$\gamma_X = [H(x) - H(x|y)] \ / \ H(x) \ . \qquad (1)$$

If one reverses the roles of the variables, making y the independent variable and x the dependent one, one obtains the separability of the features, to wit:

$$\gamma_Y = [H(y) - H(y|x)] \ / \ H(y) \ . \qquad (2)$$

The separability γ_X evaluates - in terms of entropy reduction - the efficiency with which values of the independent variable can be inferred from the observation of the dependent one, that is the efficiency with which the direct observation of an object can be replaced by an observation of its features. A few easily demonstrated properties of the separability are the following:

- It depends only on the relative weights of the nexi.

- It is dimensionless, and invariant with respect to the basis of the logarithms used for the entropies.

- It ranges from 0 to 1, inclusive, and becomes zero when all items in the group are identical. Conversely, $\gamma_X = 1$ when no two items in the group have a feature in common.

- It is not a probability. For instance in fig.2, one has: $\gamma_X = 1/4$, as one would expect of an efficiency, considering that there are 3 chances in 4 to observe feature d, which is completely uninformative about its column of origin whilst in all other cases each item can be unambiguously identified from its observed feature. Note that the probability to successfully guess a column, given a line, is larger than γ_X, namely 1/2.

- It depends on the direction of inference.

- Its definition puts no restriction on the kind of features attributed to the items. Therefore, γ can be used with arbitrarily defined distances, if required.

- It satisfies a fundamental relation which is a weak counterpart of the triangle inequality, and which provides the space spanned by the items with a modicum of structure based on a generalisation of the

concept of convexity. This important relation will be demonstrated for three items A, B, C defined by the links a_i, b_i, c_i, and a compound item AB defined by the links (a_i+b_i), as shown in fig.3.
To simplify typography, define a, b, c, and m as follows:

$$a = \Sigma\, a_i \qquad b = \Sigma\, b_i \qquad c = \Sigma\, c_i \qquad m = (a+b+c) \ .$$

For the demonstration, the arguments in the expressions for the entropy and the mutual information will be augmented by a separating colon and the letter x (if needed), followed by curly brackets which enclose the current range of x-values, in order to indicate explicitely to which of the three sets of items shown in fig.3 an expression refers, e.g.:

$$H(x\{A,B,C\}) = H(x\{AB,C\}) + [(a+b)/m]\, H(x\{A,B\}) \ , \tag{3}$$

$$H(y:x\{AB,C\}) = [(a+b)/m]\, H(y:x\{AB\}) + [c/m]\, H(y:x\{C\}) \quad \text{etc.}$$

Similarly, a curly bracket indicates the range of x-values for γ_X, e.g.:

$$\gamma_X\{AB,C\} = I(y;\, x\{AB,C\}) \,/\, H(x\{AB,C\}) \ .$$

By the definition of the mutual information one has:

$$m\, I(y;\, x\{A,B,C\}) = m\, H(y:x\{A,B,C\}) - a\, H(y:x\{A\}) - b\, H(y:x\{B\}) - c\, H(y:x\{C\}), \tag{4}$$

$$m\, I(y;\, x\{AB,C\}) = m\, H(y:x\{AB,C\}) - (a+b)\, H(y:x\{AB\}) - c\, H(y:x\{C\}) \tag{5}$$

Substracting (5) from (4), dividing by m, and reshuffling yields:

$$I(y;\, x\{A,B,C\}) = I(y;\, x\{AB,C\}) + [(a+b)/m]\, I(y;\, x\{A,B\}). \tag{6}$$

For brevity, define : $p = I(y;\, x\{AB,C\})$, $r = [(a+b)/m]\, I(y;\, x\{A,B\})$,

$$q = H(x\{AB,C\}) \ , \qquad s = [(a+b)/m]\, H(x\{A,B\}) \ .$$

Then, from the definition (1) of γ_X, and from (3) and (6) one obtains:

$$\gamma_X\{A,B,C\} = (p+r) \,/\, (q+s) \ .$$

Finally, this equation, together with the nearly self-evident lemma :

if $p,q,r,s > 0$, and $p/q < r/s$, then : $p/q < (p+r)/(q+s) < r/s$,

leads to the fundamental relation:

$$\gamma_X\{A,B\} \leq \gamma_X\{A,B,C\} \Longleftrightarrow \gamma_X\{A,B,C\} \leq \gamma_X\{AB,C\} \ , \tag{7}$$

where the equality holds simultaneously on both sides.

3. EVALUATING AMOUNTS OF KNOWLEDGE

It has been shown that, given a representation of knowledge through nexi, the separability calculates the efficiency with which direct observations of items can be replaced by observations of their features, on the strength of that knowledge. Obviously, it is also desirable to evaluate the effectivity and the extension of the knowledge itself. By the same token as before, this effectivity is a ratio of entropies, namely the quotient of the entropy reduction v imposed by the existing

connections on allowed combinations of variable values, divided by the (average) negentropy needed to identify a nexus, i.e. the joint entropy u of all nexi. Because a representation of knowledge may have any number of dimensions, its effectivity must be derived in an N-dimensional space. However, in order to avoid a messy use of indexes, and to simplify the phrasing, the formulae for u and v will be derived in a four-dimensional space spanned by the variables w,x,y, and z; the extension to an arbitrary number of dimensions is obvious.

The determination of the joint entropy u is immediate:

$$u = H(wxyz) = H(w) + H(x\,|w) + H(y\,|wx) + H(z\,|wxy) \qquad (8)$$

To determine the entropy reduction v, one may for instance chose the following sequence: first determine the reduction enforced by the connections between w and the subspace xyz, then - within this subspace - the reduction due to the connections between x and the sub-subspace yz, and finally the reduction due to the connections between y and z in the sub-subspace. One then obtains:

$$v = I(w;\, xyz) + I(x;\, yz\,|w) + I(y;\, z\,|wx)\ . \qquad (9)$$

Inserting the three standard identities:

$I(w;xyz) = H(w) - H(w|xyz)$,

$I(x;yz|w) = H(x|w) - H(x|yzw)$, and

$I(y;z|wx) = H(y|wx) - H(y|zwx)$ on the right side of (9) yields :

$$v = H(w) + H(x|w) + H(y|wx) - H(w|xyz) - H(x|yzw) - H(y|zwx). \qquad (10)$$

Adding and substracting the additional term $H(z|wxy)$ on the right side of equation (10), and taking into account (8), one obtains :

$$v = H(wxyz) - [\ H(w|xyz) + H(x|yzw) + H(y|zwx) + H(z|wxy)\ ] \qquad (11)$$

which shows that equation (9) actually is symmetric in the variables, and hence independent of the sequence in which the variables were called up for its derivation. Extending equation (11) to N dimensions $x_1, x_2, \ldots x_N$ yields:

$$v = H(x_1, x_2, \ldots x_N) - \sum H(x_i \mid x_1, \ldots x_{i-1}, x_{i+1}, \ldots x_N) \qquad (12)$$

(here and in what follows all sums range from i = 1 to N)

Equation (12) will be taken as the definition of a generalized mutual information. From (8) and (12) one derives the expression for the effectivity of a representation of knowledge in N dimensions, which will be called its diagonality, and designated by the letter ζ:

$$\zeta \equiv v/u = 1 - \sum H(x_i \mid x_1, \ldots x_{i-1}, x_{i+1}, \ldots x_N) \,/\, H(x_1, x_2, \ldots x_N)\ .$$

Obviously, one has :

$$0 \le \sum H(x_i \mid x_1, \ldots x_{i-1}, x_{i+1}, \ldots x_N) \le H(x_1, x_2, \ldots x_N)\ ,$$

with equality on the left if and only if any N-1 variables uniquely determine the remaining one, and with equality on the right if and only if

the variables are statistically independent. Hence the diagonality is a measure of the dependence of single variables, and:

$0 \leq \zeta \leq 1$.

For completeness, note that the effective extension of a representation can be measured by its span S, which is defined as follows:

$S = \exp [\sum H(x_i)]$.

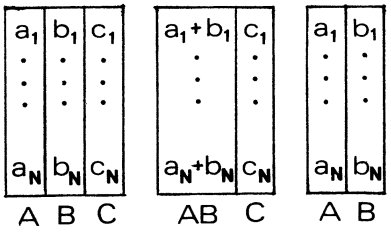
The span is a continuous and consistent variant of the number of degrees of freedom. The product $\zeta \cdot S$ evaluates the serviceable content of a representation of knowledge, a topic which will not be pursued any further here.

4. APPLICATIONS

4.1. CLUSTERING

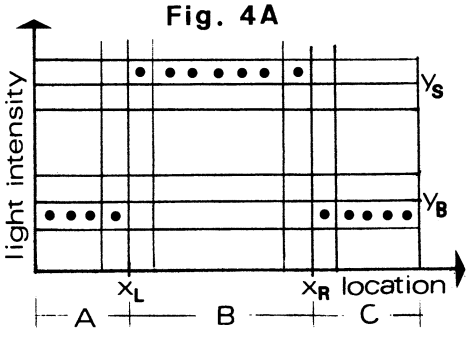
One of the pleasant properties of separability is to provide a contingency coefficient which takes into account the direction of inference and is free of the problems that beset such classics as χ^2, Pearson's C, etc.. [2]. However, the most fruitful uses of γ have hitherto been its applications to clustering (or equivalently: ordering), and mensuration. The algorithm for clustering is simple: given a set of items identified by the discrete variable x (where the items may have different frequencies of occurrence), and a set of features identified by the discrete variable y, and given further that each item is characterized by some features which it exhibits according to a known distribution when observed, then the algorithm must search for the pair of items with the smallest separability γ_X, and fuse them into a new (compound) item, by pairwise addition of the frequencies pertaining to the same feature. The procedure can then be repeated until all items have been fused together. Relation (7) can be used to ensure a monotonous increase of the separability during successive fusions.

Fig. 3



SETS MADE UP OF SIMPLE AND COMPOUND ITEMS

Fig. 4A



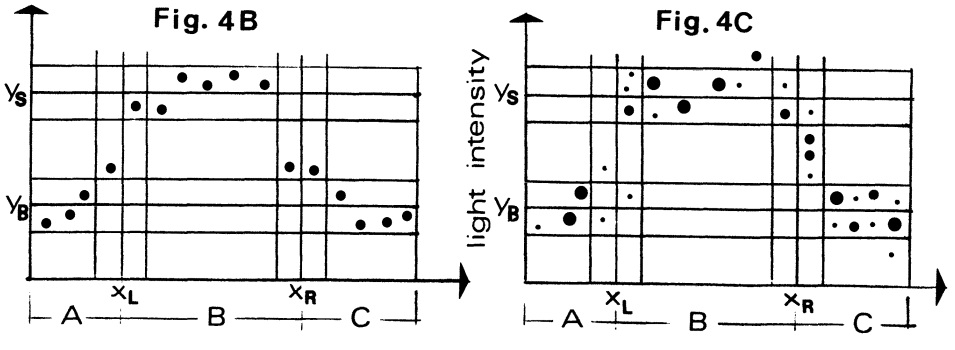
MENSURATION OF A STICK: THE IDEAL CASE

It is possible to keep a check on the effectivity of the clustering by calculating the diagonality of the representation at each step. The increase of ζ over its original value evaluates the increase in effectivity obtained by the performed fusions, and the maximum of ζ identifies the optimal partition, thus providing a desirable and useful stopping rule. An essential property of this clustering algorithm is the possibility to formally exchange the roles of both variables, and then to cluster the features (i.e. the values of the dependent variable) in precisely the same way as the items. This is not too far off the bootstrapping practice which underlies most natural learning processes, and provides a ready-made solution to such problems as the automatized choice of optimal keywords in a database. The program TAXIS exploits some of these possibilities [3].

4.2. MEASURING

Let us now examine the mechanism of a physical mensuration, say the measure of the length of a stick, or - to simplify matters even further - the determination of the position of one of its extremities. The physical aspects of the problem shall also be simplified in the extreme, to the extent that all variables are absolutely continuous, observations can be made with a limitless precision, all wave effects disregarded, etc. Surprisingly, even under these idealized conditions, taking a measurement is not an elementary process, but requires an entire set of observations, followed by the extremalization of a specific entropy.

Suppose the stick is aligned with the x-axis, and is of a lighter shade than the background (fig.4). Suppose further that one has at one's disposal a photometer, the position of which along the axis can be read off with unlimited accuracy. Obviously, a measure such as the following: "an extremity of the stick is located at coordinate 3.6501 cm ..." cannot be derived from a primitive, irreducible observation which consists in a single reading of the photometer, together with its position when this reading was made. Indeed, no single observation of the light inten-



OBSERVATIONS FOR THE MENSURATION OF A STICK
4 B: NOISY, SINGLE-VALUED 4 C: NOISY, AMBIGUOUS

sity excludes the possibility that it has the same value in all other locations along the x-axis, in which case the stick will be invisible to the photometer, and there will be nothing to measure !

Now the question arises: how is a mensuration performed? Suppose that a large number of observations have been made in the region of interest, along the stick and beyond its ends. Depending on the amount of noise, the width of the viewing angle of the photometer etc., a record of these observations will resemble fig.4A, 4B, or 4C. In these figures, the independent variable x represents positions of the photometer, and the dependent variable y light intensities observed in these positions. The stick extends approximately from x_L to x_R, its grey value is about y_S, and that of the background about y_B. Each point in the diagram denotes an observation, and the grid represents the ultimate resolutions, whether actual or supposed, of the photometer and of its positioning device, respectively.

As locations in space (here values of x) are entities which cannot be distinguished *per se*, but only through the values of some dependent variable attached to each of them, the situation in terms of entropy is as follows: the observed range of x comprises three large, homogeneous regions A, B, C with low specific entropies, separated by two narrow strips around X_L and X_R, which form the boundaries between the former, and which have high specific entropies. In essence, the positions of the strips represent the desired "measurements". Their usefulness is immediately apparent: once the homogeneous regions are adequately delimited, the resolution of x within them can be significantly lowered without much loss of knowledge. If, further, the strips are sufficiently narrow, the values of y in them become practically irrelevant. In the ideal case of a perfectly sharp boundary, these values become totally uncertain, which is why (in contradistinction to observations, which connect two quantities) physical measurements state only the value of a single quantity (e.g. the location of a boundary), and associate it with an undefined qualitative entity such as "the end of a stick", or "the volume of a vessel" etc.

Hence, whilst one can assign a probability to an observation, i.e. to the occurrence of a pair of values (or rather intervals) of the variables, this is inadequate for a mensuration, because by itself the frequency of observations in an interval provides no valid estimation of its importance as a boundary. Rather, this frequency must be weighted with the effectivity of an exact location of the interval or, equivalently, of each value of the independent variable in it. This requirement entails that boundaries must be intervals of the independent variable in which a high local variation of the dependent one warrants a high precision in the determination of the former.

Obviously, this is another way to say that boundaries must have a large separability: given a small interval of the independent variable, the separability determines how effectively its values can be inferred from observations of the dependent one. This determines quantitatively how fully the resolution of the former is exploited, as this resolution will only be functional if distinct values of the independent variable entail different values of the dependent one. Hence γ_X provides a specific measure of the degree to which an interval qualifies as a boundary;

to obtain the absolute "measuring power" of an interval, its separability must be multiplied with the number of observations in it.

The reasoning is symmetric in the variables, so that after exchanging their roles in the above example, one obtains y_S and y_B as the mensurations of y. In the idealized case of fig.4A, the value of their measuring power is proportional to the number of observations performed on the stick and on the background, respectively, because $\gamma_y = 1$. If the values of y do stray, but remain unambiguous, as shown in fig.4B, then γ_y retains the same value, but the observational weight of the intervals will decrease according to the reduced number of observations therein, provided their width remains constant. If individual observations become ambiguous, as shown in fig.4C, then γ_y gets smaller, and it will eventually become necessary to use strips of a greater width, such as Q for instance. These results are a definite improvement on *ad hoc* probability interpretations, even though the question of an optimal choice of the strip width remains open yet.

5. CONCLUSION

In conclusion, ordering (or clustering) and mensuration are two complementary aspects of the same quest for an efficient representation of the knowledge derived from a set of observations. Ideally, this representation should consist of large domains with a small specific entropy - in which a comparatively low resolution of the independent variable is adequate - separated by boundary strips with a high specific entropy and with such a small width that only the position of a strip is of the essence, not individual values of the dependent variable therein.

The extremalization of the specific entropy provides the necessary tool to build such a representation, or to recognize boundaries in an existing one. An ordering process structures the knowledge through the creation of domains with a low specific entropy, in which the difference between individual locations may be disregarded, and treats the intervals between the clusters as ideal boundaries. A mensuration, on the other hand, must respect the preexisting order enforced by the underlying continuity of the variables. It therefore searches for regions with a high specific entropy, and declares them as boundaries. The search will only be worthwhile if the distribution of the specific entropy is sufficiently non-uniform, unless such a distribution can be created by an appropriate transformation of the variables, but this - as Kipling said - is another story.

REFERENCES

[1] Kullback S. *Information Theory and Statistics* Wiley, New-York 1959.
[2] Preuss L. 'A Class of Statistics ... ' *Comm.in Stat.A9 15*, 1980.
[3] Preuss L. *TAXIS: A Programme for handling imperfectly defined data* (PC-diskette available from the author).

The Metrics Induced by the Kullback Number

Carlos C. Rodriguez
State University of New York at Albany
Department of Mathematics and Statistics
Albany, New York 12222

This research was supported by PHS grant number 1-R01-CA41171-01A1 awarded by the National Cancer Institute, DHHS.

Abstract. Within the universe of the gaussian distributions: How far away is the $N(\mu,\sigma^2)$ from the standard $N(0,1)$? Clearly, the concept of distance between elements of a statistical manifold is central to the theory of inference. It is broadly accepted that the meaningful distances should not depend on the coordinate systems used to label elements in the sample space and in the parameter space. We show in this paper that the Kullback number (entropy) induces a large class (possibly all of them) of invariant metrics on the statistical manifold.

Introduction.

Let $(X,\mathcal{B})$ be an euclidean measurable space. The separation between probabilities P and Q on $(X,\mathcal{B})$ is usually measured by the Kullback number I(P:Q) where

$$I(P:Q) = \begin{cases} \int \log \frac{dP}{dQ} \, P(dx) & \text{if } P \ll Q \\ \infty & \text{otherwise} \end{cases} .$$

As it is well known I(P:Q) does not satisfy the triangular inequality and it is clearly non symmetric. Hence, it is not a metric in the classical sense. However, we show in this paper that there exists a strong connection between the Kullback number and the metrics (possibly all of them) that are compatible with the statistical structure. i.e. that are invariant under reparametrizations and invaraint under monotone transformations of the sample space X. The language of differential geometry of statistical models provides the right framework to our results. We use the Kullback number to define a family of length functions on the tangent bundle of the model. The distance between two probability measures in the model is then given by the length of the geodesic joining them.

I. Regular Models as Riemannian Manifolds

We denote by $\mathcal{M}$ the set of all probability measures on $(X,\mathcal{B})$. A subset $P \subset \mathcal{M}$ is called a model (also known as the hypothesis space) for a given statistical problem if we assume that the

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 415–422.

(true) probability measure that generated the observations belongs to **P**. A k-dimensional regular model can be written as $\mathbf{P} = \{P_\theta : \theta \in \Theta \subset \mathbb{R}^k\}$ and in the language of differential geometry **P** is a differentiable manifold of dimension k with an atlas containing a single global chart (see Amari, 1985 and Thirring, 1978). The parameter θ plays the role of the coordinates of the measure P_θ and thus, reparametrization corresponds to changes of the coordinate system.

The Tangent Space at $P_0 \in \mathbf{P}$: $T_{P_0}(\mathbf{P})$

The tangent space at a point of an abstract manifold is the natural generalization of the concept of a tangent plane at a point on a smooth surface in euclidean 3-space. Roughly speaking, the tangent space at P_0 is the linear space of all the velocity vectors at P_0 of a point moving on the manifold. This is easy to visualize when the manifold is embedded in the familiar three dimensional space but some mental acrobatics are needed when the outside is not so familiar or when there is no outside at all! (e.g. in the space-time manifold of general relativity or in our hypothesis space **P**). It is therefore necessary to be able to define the tangent space at P_0 without reference to the "exterior" of the model **P**. We can do this (see e.g. Thirring, 1978) but here we motivate the results with an intuitive analysis by embedding **P** in the set of all the signed measures on X.

A curve passing through $P_0 \in \mathbf{P}$ is a mapping $\gamma : [-\varepsilon, \varepsilon] \to \mathbf{P}$ (for some $\varepsilon > 0$)

$$t \to \gamma(t) = P_{\theta(t)}$$

where $\gamma(0) = P_{\theta(0)} = P_0$. The elements of the tangent space $T_{P_0}(\mathbf{P})$ are all of the form

$$\lim_{h \to 0} \frac{P_{\theta(h)} - P_{\theta(0)}}{h} = \lim_{h \to 0} Q_h = Q_0 \ .$$

Notice that $\forall h \in [-\varepsilon, \varepsilon]$, Q_h is a signed measure on $(X, \mathcal{B})$. Moreover, since $P_{\theta(h)}$ and $P_{\theta(0)}$ are probability measures $Q_h(X) = 0$ $\forall h \in [-\varepsilon, \varepsilon]$. Thus, this first look immediately reveals two important characteristics of the tangent elements Q_0. They are signed measures and they put null mass on the whole space X.

Let us denote by $p(\theta)$ the density of P_θ with respect to the dominating measure μ. The densities are in fact functions of $x \in$ i.e. $p(x|\theta)$ but to simplify the notation and (more important) to emphasize the dependence on the coordinates θ we keep the variable x only implicit. We have

$$\frac{dP_{\theta(h)}}{dP_{\theta(0)}} = \frac{p(\theta(h))}{p(\theta(0))} \ .$$

Thus,

$$\frac{dQ_0}{dP_{\theta(0)}} = \lim_{h\to 0}\left\{\frac{p(\theta(h))/p(\theta(0))-1}{h}\right\} \tag{1.1}$$

and if $p(\theta)$ is differentiable at $\theta(0)$ we can write

$$p(\theta(h)) = p(\theta(0)) + \sum_i \frac{\partial p}{\partial \theta^i}\,\dot{\theta}^i(0)\,h + 0(h)$$

where $\dot{\theta}^i(0) = \left.\frac{d\theta^i(t)}{dt}\right|_{t=0} \equiv v^i$. Replacing in (1.1) we obtain

$$\frac{dQ_0}{dP_0} = \sum_{i=1}^{k} v^i \frac{1}{p(\theta(0))}\left.\frac{\partial p}{\partial \theta^i}\right|_{\theta(0)}$$

and writing $\log p(\theta)=\ell(\theta)$ (the log likelihood) we have

$$\dot{\ell}_i(P_\theta) \equiv \partial_i \ell(\theta) = \frac{1}{p(\theta)}\left.\frac{\partial p}{\partial \theta^i}\right|_\theta .$$

From where we can write

$$\frac{dQ_0}{dP_0} = \sum_{i=1}^{k} v^i\, \dot{\ell}_i(P_0) . \tag{1.2}$$

Equation (1.2) provides the natural representation of the tangent elements at P_0. They are signed measures with total null mass and their densitites with respect to P_0 are linear combinations of the derivatives of the log likelihood. Thus, since the densitites characterize the measures we can think of the tangent space at P_0 as the k-dim linear space spanned by $\dot{\ell}_1(P_0)$, $\dot{\ell}_2(P_0),\ldots,\dot{\ell}_k(P_0)$. By the usual regularity assumptions $\dot{\ell}_i(P_0)\in L^2(P_0)$ and therefore, the thangent space at P_0 is a subspace of $L^2(P_0)$. This representation has the extra virtue of being a Hilbert space which makes $\mathbf{P}$ a Riemannian manifold with metric tensor $g_{ij}(\theta)$ equal to the Fisher information matrix $g_{ij}(\theta) = \langle \partial_i \ell(\theta),\ \partial_j \ell(\theta)\rangle = \int \dot{\ell}_i(\theta)\dot{\ell}_j(\theta) P_\theta(dx)$.

II. The Entropic Length Function

Let $|\varepsilon|\le 1$. We define on the tangent bundle of $\mathbf{P}$ a family of length functions H_ε in the following way:

For $v\in T_P(\mathbf{P})$

$$H_\varepsilon(v) = \begin{cases} \|v/\varepsilon\|_\infty \, I^{1/2}(P:P_\varepsilon^{v/\|v\|_\infty}) & \text{if } v\in L^\infty(P),\ v\neq 0,\ \varepsilon\neq 0, \\ 0 \quad \text{if} \quad v = 0 & \\ \frac{1}{\sqrt{2}}\|v\|_2 \quad \text{if} \quad v\notin L^\infty(P) \quad \text{or} \quad \varepsilon = 0 & \end{cases} \tag{2.1}$$

where, if $|\varepsilon|\leq 1$, $v\in T_P(P)$ and $\|v\|_\infty=1$

$$P_\varepsilon^v(dx) = P(dx)[1 + \varepsilon v(x)] \tag{2.2}$$

is a well defined probability measure on $(X,\mathcal{B})$. We call P_ε^v the ε-deviation from P in the tangent direction v.

From the definition (2.1) and the convexity inequality for the Kullback number we have for $|\varepsilon|\leq 1$

1) $H_\varepsilon(v)\geq 0 \quad \forall v\in T(P)$ and $H_\varepsilon(v)=0$ if and only if $v=0$.
2) $\forall c\geq 0 \ H_\varepsilon(cv) = cH_\varepsilon(v)$
3) H_ε is continuous (see (2.4), (2.5) and (2.6)).

From the definition of the Kullback number we can write

$$H_\varepsilon^2(v) = \frac{\|v\|_\infty^2}{\varepsilon^2} \int - \log\left[1 + \varepsilon \frac{v(x)}{\|v\|_\infty}\right] P(dx) \ .$$

Using the expression

$$- \log(1+z) = \sum_{n=0}^{\infty} \frac{(-1)^{n+1}}{n+1} z^{n+1} \quad \text{valid for } |z|<1$$

and the fact that $E_P(v)=0$ we obtain

$$H_\varepsilon^2(v) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n+2} \left(\frac{\varepsilon}{\|v\|_\infty}\right)^n E_P\left(v^{n+2}\right) \tag{2.3}$$

$$= E_P\left\{v^2(x) \sum_{n=0}^{\infty} \frac{(-1)^n}{n+2} \left(\frac{\varepsilon v(x)}{\|v\|_\infty}\right)^n\right\} \ .$$

The last equation shows that

$$\lim_{\|v\|_\infty \to \infty} H_\varepsilon^2(v) = \frac{1}{2} E_P(v^2(x)) \tag{2.4}$$

and that

$$H_0^2(v) = \lim_{\varepsilon \to 0} H_\varepsilon^2(v) = \frac{1}{2} E_P(v^2(x)) \tag{2.5}$$

Hence, the Riemannian length is the special case $\varepsilon=0$. Moreover, from (2.3) we have

$$H_\varepsilon^2(v) \leq E_P\left\{v^2(x) \sum_{n=0}^{\infty} \frac{|\varepsilon|^n}{n+2}\right\}$$

$$= \varepsilon^{-2}\left\{\sum_{n\geq 2} \frac{|\varepsilon|^n}{n}\right\} E_P(v^2) = C_\varepsilon H_0^2(v) \quad . \tag{2.6}$$

We now associate in the classical way distances to the length functions H_ε. Let $\gamma:[a,b]\to P$ be a path (i.e. Piece-wise C^1) joining P and Q in P i.e. $\gamma(a)=P$ and $\gamma(b)=Q$. We define the ε-length of γ by

$$L_\varepsilon(\gamma) = \int_a^b H_\varepsilon(\gamma'(\tau))d\tau \tag{2.7}$$

where $$\gamma'(\tau) \in T_{\gamma(\tau)}(P) \quad .$$

the length function H_ε induces a distance on P given by

$$d_\varepsilon(P,Q) = \inf_\gamma L_\varepsilon(\gamma) \tag{2.8}$$

where the inf is taken over all the paths from P to Q or from Q to P. Clearly $\forall P, Q, R \in P$ we have:

i) $d_\varepsilon(P,Q) \geq 0$

ii) $d_\varepsilon(P,Q) \leq d_\varepsilon(P,R) + d_\varepsilon(R,Q)$

iii) $d_\varepsilon(P,Q) = d_\varepsilon(Q,P)$

iv) $d_\varepsilon(P,Q) = 0$ if and only if $P=Q$ a.e.-μ .

Part iv) follows from Theorem 1.1 in S. Lang, 1987. Moreover, all these distances are invariant under reparametrizations and invariant under monotone transformations of the sample space X. This follows trivially from the invariance of the Kullback number. Furthermore, Theorem 1.2 in S Lang, 1987 assures the equivalence of the metrics d_ε. They generate the topology of the

Hellinger distance since this is the underlying Hausdorff topology of the manifold **P**.

Extension: We can increase the family of invariant length functions (2.1) by taking averages over ε. To each distribution G on [-1,1] we associate the length function

$$H_G(v) = \int_{-1}^{1} H_\varepsilon(v)\, dG(\varepsilon) \quad . \tag{2.9}$$

Invariant metrics are obtained by replacing ε by G in (2.7) and (2.8).

III. The length of the Entropic Prior Model

Example: Let **P** be the manifold of discrete distributions on $X=\{0,1,\ldots,M\}$. We have

$$\theta = \{\theta = (f_0,\ldots,f_M): f_i>0, \sum_0^M f_i = 1\}.$$

Thus, the densities $p(x|\theta)$ with respect to the counting measure on X are

$$p(x|\theta) = (1 - \sum_1^M f_i)\ \delta(x) + \sum_1^M f_i\ \delta(x-i)$$

the metric tensor is

$$g_{ij}(\theta) = \begin{cases} \frac{1}{f_i} & \text{if} \quad i=j \\ \\ \frac{1}{f_o} & \text{if} \quad i\neq j \end{cases} \qquad i,j=1,2,\ldots,M \ . \tag{3.1}$$

Notice, that in analogous way we can show that the metric tensor in the extended manifold **P** of the positive measures on X (i.e. we do not impose $\Sigma f_i=1$) is

$$\tilde{g}_{ij}(\theta) = \begin{cases} \frac{1}{f_i} & \text{if} \quad i=j \\ \\ 0 & \text{if} \quad i\neq j \end{cases} \qquad i,j=0,1,\ldots,M \tag{3.2}$$

and one can readily check that

$$ds^2 = \sum_{i,j\geq 1} g_{ij}\, df_i df_i = d\tilde{s}^2 = \sum_{i,j\geq 0} \tilde{g}_{ij}\, df_i df_j \quad .$$

(I thank Dr. J. Skilling for pointing out this last equation to me.) It is interesting to note that (3.2) appears in the derivation of the remarkable entropic Bayesian prior "exp{αS}" (see the paper by J. Skilling in this same volume). Therefore, when the underlying hypothesis space is not the manifold of discrete distributions but an arbitrary regular space $\mathbf{P}=\{P_\theta:\theta\in\Theta\}$ the entropic prior model changes. It becomes the one parameter exponential family with sufficient statistic $I(P_\theta:m)$, generated by the invariant measure in $\mathbf{P}$ i.e. $\eta(d\theta) = \sqrt{\det g(\theta)}\, d\theta$. Where $g(\theta)$ is the Fisher information matrix in $\mathbf{P}$ and m is an initial probability measure. This is itself a one dimensional Riemannian manifold dominated by η. The η-densities on this "line" are

$$\frac{d\pi_\alpha}{d\eta}(\theta) = \pi(\theta|\alpha,m) = \exp\{-\alpha I(P_\theta:m) - \beta(\alpha,\eta)\} \qquad (3.3)$$

for $\alpha > \alpha_{Min} \geq 0$ where $\alpha_{Min} = \inf\{\alpha>0: \int \pi(\theta|\alpha,m)\eta(d\theta)<\infty\}$. Thus, the tangent (space) at α is generated by

$$\dot{\ell}(\alpha) = -I(P_\theta : m) - \dot{\beta}(\alpha,\eta)$$

which is a function of θ in $L^2(\pi_\alpha)$. The Fisher information amount is given by

$$g_{11}(\alpha) = E_\alpha(\dot{\ell}^2) = var_\alpha(I(P_\theta:m)) \quad . \qquad (3.4)$$

Therefore, the Riemannian length of the entropic prior model $\Pi_{\mathbf{P}} = \{\pi_\alpha:\alpha\geq\alpha_{min}\}$ becomes

$$L_0(\Pi_{\mathbf{P}}) = \int_{\alpha_{Min}}^{\infty} \sqrt{g_{11}(\alpha)}\, d\alpha = \int_{\alpha_{Min}}^{\infty} \sigma_\alpha(I(P_\theta:m))d\alpha \qquad (3.5)$$

where $\sigma_\alpha(I(P_\theta:m))$ denotes the standard deviation of the random variable $I(P_\theta:m)$ when $\theta \to \pi_\alpha$. Equation (3.5) provides an information-theoretic measure of the size of $\mathbf{P}$ as seeing from m.

References:

Amari, S., 1985. Differential-Geometrical Methods in Statistics. Lecture Notes in Statistics, 28, Springer-Verlag.

Lang, S., 1987. Introduction to Complex Hyperbolic Spaces. Springer-Verlag, New York.

Skilling, J., 1988. Classical Max Ent Data Analysis. These proceedings.

Thirring, W., 1978. A Course in Mathematical Physics I. Springer-Verlag, New York.

THE PARADOX OF THE MONEY PUMP: A RESOLUTION

Randall Barron
Laboratory Technologies Corp.
400 Research Drive
Wilmington, MA 01887
USA

ABSTRACT. The Paradox of the Money Pump, introduced by Carlos Rodriguez at the 1986 MAXENT Workshop, can be quietly resolved by a Bayesian analysis of the Envelope Game. Everything derives from the prior probability distribution for the total purse, that is, the sum of the checks in both envelopes. It turns out that the Money Pump can indeed operate -- but only when the expectation-value of the total purse is infinite. In that case, the expected cost to the sponsor of the game is infinite, and there is no fair price that can be charged for admission to the game as a player.

THE ENVELOPE GAME

The Paradox of the Money Pump was presented at an earlier meeting of this Workshop by Carlos Rodriguez of SUNY [1]. It is formulated in the context of a game of chance, known for convenience as the Envelope Game. You, the player, are to select one of two envelopes, knowing that each contains a check made out to "Bearer". You are informed that one of the checks has a face value twice that of the other, but you do not know the absolute values. After making a selection, you open the envelope and learn the face value of the check therein, say \$A . You are now offered the option of exchanging the check you hold for the one in the remaining envelope. Carlos argued -- more or less plausibly -- that, if you exchange envelopes, you will, with even odds, either double your initial winnings or suffer a loss of half. Your expected gain on the exchange is therefore $(1/2)(2A) - (1/2)(A/2) = A/4$, which is strictly positive. Given the symmetry inherent in the initial choice of envelopes, this seems absurd.

Carlos went on to make it seem even more absurd by introducing a secondary player who is assigned, by default, the envelope remaining after your initial selection. Reasoning as above, you should both be willing to exchange your envelopes and, moreover, to pay some token amount to a regulatory third party for the privilege of making the exchange.

Of course, there is no mystery about the source of the money. The game has a Sponsor, namely, the agency on whose account the checks are drawn. A Bayesian analysis of the Envelope Game is based on the prior probability distribution for the total purse, that is, the sum of the checks in both envelopes. The naive analysis given above tacitly assumes that the prior distribution for the total purse is flat -- all possible values are equally likely. It seems to me that the naive analysis entails an unnormalizable prior distribution but, as I will show, the operation of the Money Pump does not depend on a flat or

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 423–428.

unnormalizable prior. There exist proper, normalized probability distributions for the total purse such that the expected gain on exchange of envelopes is strictly positive for all values of A. This latter condition, I take it, defines the essence of the Money Pump.

STATISTICAL INDEPENDENCE AND SYMMETRY

There are two random processes involved here: (1) the sponsor somehow decides on the total purse, X; and (2) you, the player, blindly select one of the two envelopes.

Let H be the proposition that your envelope contains the smaller share, $X/3$, and H', the complementary proposition that it contains the larger share, $2X/3$. Drawing on your common sense, you observe that processes (1) and (2) are completely independent of each other, and you conclude that the joint probabilities factor,

$$P(H,X=C\,|\,I) = P(H\,|\,I) \bullet P(X=C\,|\,I)\,,$$

$$P(H',X=C\,|\,I) = P(H'\,|\,I) \bullet P(X=C\,|\,I)\,.$$

As usual, the probabilities are conditional on your prior information I. Under the conditions of the game, the envelopes are indistinguishable before opening and, by symmetry, you assign equal values to the prior probabilities of H and H',

$$P(H\,|\,I) = P(H'\,|\,I) = 1/2\,.$$

PHASE 1: SELECT AN ENVELOPE

In Phase 1 of the game, you select an envelope. The face value of the check within is a random variable, Y, with a prior probability distribution given by

$$P(Y=A\,|\,I) = P(Y=A,H\,|\,I) + P(Y=A,H'\,|\,I)$$

$$= P(X=3A,H\,|\,I) + P(X=3A/2,H'\,|\,I)$$

$$= (1/2)\bullet P(X=3A\,|\,I) + (1/2)\bullet P(X=3A/2\,|\,I)\,.$$

Your expected winnings in Phase 1 can be computed as

$$\langle Y\,|\,I\rangle = \Sigma_A\, A \bullet P(Y=A\,|\,I)$$

$$= \Sigma_A\, (1/6)\bullet(3A)\bullet P(X=3A\,|\,I) + (1/3)\bullet(3A/2)\bullet P(X=3A/2\,|\,I)$$

$$= (1/6)\bullet\langle X\,|\,I\rangle + (1/3)\bullet\langle X\,|\,I\rangle = (1/2)\bullet\langle X\,|\,I\rangle\,,$$

which is half the expected value of the total purse. We can be free with the infinite summations because of positivity. We have to assume, for definiteness, that Y takes values in some discrete set, e.g., the integer powers of 2.

PHASE 2: OPTIONAL EXCHANGE

Suppose that you open your envelope and discover that $Y=A$. In Phase 2 of the game you consider exchanging the envelope you hold for the one remaining. The posterior probability distribution for (H,H') can now be computed using Bayes' Rule,

$$p = P(H \mid Y=A,I) = P(H,Y=A \mid I) \;/\; P(Y=A \mid I)$$

$$= P(X=3A,H \mid I) \;/\; \{P(X=3A,H \mid I) + P(X=3A/2,H' \mid I)\}$$

$$= P(X=3A \mid I) \;/\; \{P(X=3A \mid I) + P(X=3A/2 \mid I)\} ,$$

$$p' = P(H' \mid Y=A,I) = 1 - p .$$

Your expected gain on exchange of envelopes is therefore

$$\langle G \mid Y=A,I \rangle = p \bullet A + (1-p) \bullet (-A/2)$$

$$= A \bullet \frac{P(X=3A \mid I) - (1/2)P(X=3A/2 \mid I)}{P(X=3A \mid I) + P(X=3A/2 \mid I)} .$$

Whether this is positive, negative or zero depends on local details of the prior probability distribution for the total purse. The following example shows that this quantity may be strictly positive for all possible values of A , i.e., the Money Pump can operate.

EXAMPLE

Suppose the face values on the checks (expressed in pennies, say, or pence) are restricted to be non-negative integer powers of two,

$$y(k) = 2^k , \qquad k = 0, 1, 2, \ldots ,$$

so that the possible values of the total purse are the numbers,

$$x(k) = 3 \bullet 2^k , \qquad k = 0, 1, 2, \ldots .$$

Suppose that the prior distribution for the total purse is

$$p(k) = (1-z)z^k , \qquad k = 0, 1, 2, \ldots ,$$

for some suitable positive $z < 1$. This is a proper, normalized probability distribution, although the expectation-value,

$$\langle X \rangle = \Sigma_k \, x(k)p(k) = \Sigma_k \, 3(1-z)(2z)^k ,$$

will be infinite whenever $z \geq 1/2$. The expected gain on exchange of envelopes,

$$g(k) = 2^k \bullet [p(k) - (1/2)p(k-1)] \; / \; [p(k) + p(k-1)]$$

$$= \begin{cases} 2^k \bullet (z - 1/2)/(z+1) & k = 1, 2, \ldots \\ 1 & k = 0 , \end{cases}$$

is strictly positive for all $k = 0, 1, 2,\ldots$ provided that $z > 1/2$.

GENERAL CASE

Averaging the conditional expected gain $\langle G \mid Y=A,I \rangle$ over the prior probability distribution $P(Y=A \mid I)$ yields a useful auxiliary quantity,

$$\langle\langle G \mid [Y] I \rangle\rangle = \Sigma_A \langle G \mid Y=A,I \rangle \bullet P(Y=A,I)$$

$$= \Sigma_A \, (A/2) \bullet \{ P(X=3A \mid I) - (1/2)P(X=3A/2 \mid I) \}$$

$$= (1/6) \, \Sigma_A \{ (3A) \bullet P(X=3A \mid I) - (3A/2) \bullet P(X=3A/2 \mid I) \} .$$

The individual terms in the sum are each expressed as the difference of two non-negative quantities. We can split each term into positive and negative components, then regroup and sum the components of like sign separately, provided that at least one of the definite partial sums is finite. That is,

$$\langle\langle G \mid [Y] I \rangle\rangle = (1/6) \, \Sigma_A \{ (3A) \bullet P(X=3A \mid I) - (3A/2) \bullet P(X=3A/2 \mid I) \}$$

$$= (1/6) \, \Sigma_A (3A) \bullet P(X=3A \mid I) - (1/6) \, \Sigma_A (3A/2) \bullet P(X=3A/2 \mid I)$$

$$= (1/6) \bullet \langle X \mid I \rangle - (1/6) \bullet \langle X \mid I \rangle = 0 ,$$

provided $\langle X \mid I \rangle$ is finite. If $\langle X \mid I \rangle$ is infinite, no such conclusion can be drawn; the example presented earlier illustrates this point. Thus,

$$\langle\langle G \mid [Y] \rangle\rangle = \begin{cases} 0 & \text{if } \langle X \mid I \rangle \text{ is finite} , \\ \text{unconstrained} & \text{otherwise} . \end{cases}$$

The gain on exchange of envelopes is a well-defined random variable,

$$G = S(H) \bullet (X/3) ,$$

where $S(H) = H - H'$ is 1 if H is TRUE, and -1 if H is FALSE. If you adopt a policy of always exchanging envelopes (whether or not you look at the contents of the first), your expected gain in Phase 2 will be

$$\langle G \mid I \rangle = \Sigma_C (C/3) \bullet P(X=C,H \mid I) + \Sigma_C (-C/3) \bullet P(X=C,H' \mid I) .$$

It is appropriate to group the positive and negative terms and sum them separately, because they represent disjoint events in the product space of joint possibilities. Hence,

$$\langle G \mid I \rangle = P(H \mid I) \bullet (1/3) \bullet \langle X \mid I \rangle - P(H' \mid I) \bullet (1/3) \bullet \langle X \mid I \rangle$$

$$= (1/6) \bullet \langle X \mid I \rangle - (1/6) \bullet \langle X \mid I \rangle \, .$$

If $\langle X \mid I \rangle$ is finite, the sum is zero. If $\langle X \mid I \rangle$ is infinite, the sum is $\infty - \infty$, or indefinite; in the customary parlance, it "does not exist".

$$\langle G \mid I \rangle = \begin{cases} 0 & \text{if } \langle X \mid I \rangle \text{ is finite} \, , \\ \text{indefinite} & \text{otherwise} \, . \end{cases}$$

EXTENDED REAL NUMBERS

We are already familiar with the Extended Real Numbers, comprising the Real Line $\mathbf{R}$ and the Definite Infinities, $+\infty$ and $-\infty$. Suppose that we extend the set further by means of the Real Indefinite quantity, $\sim = \infty - \infty$. Formally Real sums and integrals which "do not exist" in the usual (Lebesgue) sense may be considered, in a consistent sense, to be Real Indefinite [2]. In particular, one can show that the expectation-value of a Real random variable, not necessarily bounded, always exists as an element of the enlarged set of extended real numbers consisting of $\mathbf{R}$ and $\{+\infty, \sim, -\infty\}$. Of course, the laws of arithmetic are different for expressions involving $\sim$, just as they are for expressions involving $\pm\infty$.

The IEEE Standard for Binary Floating Point Arithmetic (to be performed by machines) requires reserved bit patterns encoding the definite infinities, $\pm\infty$. These are examples of what is called a NaN (short for Not-a-Number). The Intel 80X87 implementation of the IEEE Standard includes a reserved NaN encoding the Real Indefinite, $\sim = \infty - \infty$. Single-Precision (4-byte) bit patterns for these NaN's are tabulated below, along with some representative real numbers. For those of you who are programmers, I have also given the Hex representation.

TABLE: Single-Precision Bit Patterns from the Intel 80X87 Implementation

ASCII	BINARY	HEXADECIMAL
0.0	00000000000000000000000000000000	00000000
1.0	00111111100000000000000000000000	3F800000
2.0	01000000000000000000000000000000	40000000
$+\infty$	01111111100000000000000000000000	7F800000
$-\infty$	11111111100000000000000000000000	FF800000
$\sim$	11111111110000000000000000000000	FFC00000

CONCLUSIONS

If $\langle X \mid I \rangle$ is finite, the net flux of cash from the Money Pump is zero, and our intuition is vindicated. If $\langle X \mid I \rangle$ is infinite, then intuition fails us, and the Money Pump can operate.

The technical resolution of the paradox can be stated as a theorem.

Theorem: If $\langle X \mid I \rangle$ is finite, then $\langle\langle G \mid [Y] I \rangle\rangle = \langle G \mid I \rangle = 0$.
If $\langle X \mid I \rangle$ is infinite, then $\langle\langle G \mid [Y] I \rangle\rangle$ is unconstrained and $\langle G \mid I \rangle = \sim$.

Corollary of The Money Pump: If $\langle G \mid Y=A, I \rangle$ is strictly positive whenever $P(Y=A \mid I) > 0$, then $\langle\langle G \mid [Y] I \rangle\rangle$ is positive, $\langle X \mid I \rangle = +\infty$, and $\langle G \mid I \rangle = \sim$.

The classical interpretation of the notion of expectation-value is provided by the Strong Law of Large Numbers. According to this theorem, the average over a long series of independent samples of a random variable G,

$$S_N = (1/N) \sum_{n=1}^{N} G_n .$$

converges as $N \longrightarrow \infty$ to $\langle G \rangle$ (with probability 1), provided $\langle G \rangle$ is finite. Special cases of the Strong Law have been known for more than two centuries. Recent investigations have uncovered an extension of the Strong Law to the case where $\langle G \rangle$ is infinite or indefinite [3]. If $\langle G \rangle = \pm\infty$, then S_N diverges definitely to $\langle G \rangle$ (w.p.1). If $\langle G \rangle = \sim$, there are three possibilities, depending on finer details [4] of the probability distribution: (1) S_N diverges definitely to $+\infty$ (w.p.1); (2) S_N diverges definitely to $-\infty$ (w.p.1); or (3) S_N has a non-trivial set of limit points that (w.p.1) includes both $+\infty$ and $-\infty$.

In the context of the Envelope Game, we have seen that the Money Pump can operate only when $\langle G \mid I \rangle = \sim$. By symmetry in (H,H') we argue that case (3) then obtains. In a long series of independent instances of the game, the average gain on exchange of envelopes will (w.p.1) diverge indefinitely to both $+\infty$ and $-\infty$. However, if we select only those instances in which $Y_n = A$, the average gain on exchange of envelopes will (w.p.1) converge to the conditional expectation, $\langle G \mid Y=A, I \rangle$, which is, of course, finite -- and may even be strictly positive. There is no contradiction here, although the result is a bit surprising.

The Strong Law of Large Numbers provides a rationale for setting the admission fee to games of chance where the expected gain is finite. This rationale breaks down when the expected gain is infinite or indefinite, as in the classical St. Petersburg Paradox, and, also, in the case at hand, when the Money Pump is running.

REFERENCES

[1] Rodriguez, C.C., 'Understanding Ignorance', *Maximum-Entropy and Bayesian Methods in Science and Engineering*, proceedings of the 6th Maximum Entropy Workshop, Seattle, August 1986, 2 vols., edited by G.J. Erickson and C.R. Smith, published by Kluwer Academic Publishers, Dordrecht, 1988.

[2] Barron, A.R., *Integrals, Expectation-Values and Entropy*, PhD Thesis (Physics), Brandeis University, 1981. *Dissertation Abstracts* **42B**, 4828 (1982).

[3] Kesten, Harry, 'The limit points of a normalized random walk', *Ann. Math. Statist.* **41**, 1173-1205 (1970). See especially Corr. 3, p. 1195.

[4] Erickson, K.B., 'The Strong Law of Large Numbers when the mean is undefined', *Trans. Amer. Math. Soc.* **185**, 371-381 (1973). See especially Th. 2, Corr. 2, p. 372.

Constrained Maximum Entropy Methods in an Image Reconstruction Problem

Andrew Gelman*
Department of Statistics
Harvard University
Cambridge, Massachusetts 02138
U.S.A.

Abstract

Maximum entropy and maximum likelihood methods are compared for a simplified version of a medical imaging problem. Iterative reconstructions are tracked by plotting successive values of log-likelihood and entropy, and we find a tradeoff between these two measures of fit. Maximum likelihood is found to fit the data more closely, but maximum entropy creates more reasonable images. We conclude that the former uses the data efficiently, but the latter gives a better choice of image. This reasoning leads to a somewhat Bayesian version of the constrained maximum entropy method of Gull and Daniell (1978). The constraint of that method is interpreted from a Bayesian perspective.

1 Background and setting up the problem

This paper discusses image reconstruction from incomplete, noisy data. Our main example is a simplified model of positron emission tomography (Vardi et al. (1985)). We consider this problem on a theoretical level only, and the brief description which follows may be thought of as motivation for our study. In emission tomography, the image x of interest is an intensity function of radioactive emissions from a two-dimensional region in the human brain. We cannot directly observe x on a live person, but we can count the emissions that leave the brain, and observe their direction. These indirect observations come in the form of a finite set of counts, labeled $y = (y_1, \ldots, y_n)$, in n pairs of radiation detectors outside the brain. (Note: all vectors in this paper are column vectors.) The assumed probability model is:

$$\begin{aligned} y_i &\sim \text{independent Poisson } (M_i),\ i = 1, \ldots, n, \\ M &= (M_1, \ldots, M_n) \\ &= Ax. \end{aligned}$$

*Much of this paper derives from helpful comments by Donald Rubin and Stephen Ansolabehere. This research was supported by a U.S. National Science Foundation graduate fellowship.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 429–435.

The expectations M_i (the 'mock data' of Skilling (1986)) are derived from x by a linear transformation A of conditional probabilities. To make the problem tractable, the image x is defined on the discrete space of a grid of N picture elements or 'pixels'. The image is then a vector $x = (x_1, \ldots, x_N)$ of nonnegative elements, and the linear transformation A can be identified with an matrix of rank n, each of whose columns sum to 1. (None of the entries of A will be negative.) The parameter N is chosen by the analyst; to avoid major discretization errors, we will typically assume $N > n$. Note, however, that care is then required in picking reasonable images from a large N-dimensional space.

For this problem, the likelihood is $f(y|M) \propto \prod_i M_i^{y_i} e^{-M_i}$, and we define

$$\begin{aligned} \text{-2LL}(M|y) &= -2\log f(y|M) + \text{ arbitrary constant} \\ &= -2\sum_i \left[y_i \log\left(\frac{M_i}{y_i}\right) + y_i - M_i \right]. \end{aligned}$$

(This corresponds to the 'chisquared' statistic of Skilling (1986).) In passing from the first line to the second, we have set the constant so that $\text{-2LL}(M|y) = 0$ at the maximum, when $M = y$.

In general, the entropy of a vector $a = (a_1, \ldots, a_K)$, relative to a measure $b = (b_1, \ldots, b_K)$, is defined as:

$$S(a|b) = -\sum_k \left(\frac{a_k}{\sum a_j}\right) \log\left(\frac{a_k/\sum a_j}{b_k/\sum b_j}\right).$$

2 Comparing maximum entropy and maximum likelihood estimators

We will consider two simple estimators $\hat{x}$ of x. In both cases, we define the estimated sampling expectations $\hat{M} = A\hat{x}$. First, the constrained maximum entropy estimate of Gull and Daniell (1978) and Skilling (1986) is the $\hat{x}$ that maximizes $S(\hat{x}|m)$, subject to the constraint: $\text{-2LL}(\hat{M}|y) \leq n$. (We will define the entropy relative to the uniform measure: $m_j = 1$, for all j.) If the constraint on -2LL cannot be satisfied, then the maximum likelihood estimate (defined below) will be labeled as 'constrained maximum entropy', too.

Second, the maximum likelihood estimate is an nonnegative image $\hat{x}$ that minimizes $\text{-2LL}(\hat{M}|y)$. The estimate $\hat{x}$ will be unique, except when the absolute minimum, $\text{-2LL}(\hat{M}|y) = 0$ (that is, $A\hat{x} = y$) can be achieved. In this case, we choose, as a unique 'maximum likelihood estimate', the $\hat{x}$ that maximizes the entropy $S(\hat{x}|m)$, subject to the constraint: $\text{-2LL}(\hat{M}|y) = 0$.

Conditional on the true image x, an estimate $\hat{x}$ is a function of the random variable y. Rather than examine an $\hat{x}$ directly, we look at its fit to the prior measure m, observations y, true image x, and true sampling expectations M. These four summary comparisons are: $S(\hat{x}|m)$, $\text{-2LL}(\hat{M}|y)$, $S(\hat{x}|x)$, and $S(\hat{M}|M)$, respectively. We are interested in the expectations of these quantities, averaged over the sampling distribution of y. For fixed dimensions n and N, a fixed transition matrix A, and a fixed true image x, we can simulate independent data sets y (from the appropriate Poisson distributions). Given n, N, A, and y, a computer program finds the constrained maximum entropy and maximum likelihood estimates of x. For each estimator, the program then calculates the average values of the

Table 1: Approximate sampling expectations of various functions of two estimators $\hat{x}$ of the image x

True image	Dimension of data vector	Reconstruction grid	Estimator	Fit to prior measure: $-S(\hat{x}\|m)$	Fit to data: - 2LL($\hat{M}\|y$)	Fit to true image: $-S(\hat{x}\|x)$	Fit to truth in data space: $-S(\hat{M}\|M)$
	n=6	4 × 4	max-ent	.13	6.0	.10	.0064
			m.l.e.	.38	0.0	.15	.0050
		8 × 8	max-ent	.10	6.0	.16	.0061
			m.l.e.	.29	0.0	.19	.0050
20 20 20 20	n=12	4 × 4	max-ent	.08	12.0	.12	.0050
20 100 100 20			m.l.e.	.89	4.1	.51	.0034
20 100 100 20		8 × 8	max-ent	.06	12.0	.17	.0053
20 20 20 20			m.l.e.	1.43	3.4	1.11	.0037
	n=24	4 × 4	max-ent	.16	24.0	.07	.0115
			m.l.e.	.57	9.3	.28	.0099
		8 × 8	max-ent	.13	24.0	.12	.0117
			m.l.e.	1.60	2.2	1.35	.0132
	n=6	4 × 4	max-ent	.07	6.0	.28	.0067
			m.l.e.	.30	0.0	.38	.0048
		8 × 8	max-ent	.06	6.0	.29	.0066
			m.l.e.	.23	0.0	.38	.0048
20 20 20 20	n=12	4 × 4	max-ent	.19	12.1	.36	.0064
20 200 20 20			m.l.e.	1.21	5.9	.89	.0055
20 20 20 20		8 × 8	max-ent	.27	12.1	.49	.0066
20 20 20 20			m.l.e.	2.37	5.3	2.18	.0057
	n=24	4 × 4	max-ent	.21	24.0	.19	.0299
			m.l.e.	.82	11.6	.69	.0411
		8 × 8	max-ent	.14	24.0	.21	.0133
			m.l.e.	1.69	2.2	1.24	.0152

four comparisons described above, over 20 simulations of y. For this paper, we did the above computation for 12 cases: 2 true images $\hat{x}$ (each defined on a 4 × 4 grid); 3 sets of n and A; and 2 reconstruction grids (4 × 4 and 8 × 8; that is, $N = 16$ and $N = 64$). The results are shown in Table 1.

3 Tradeoff between likelihood and entropy

Table 1 shows that maximum likelihood better fits the true M, as well as, of course, the data y. Constrained maximum entropy better fits the true x, as well as, of course, the prior measure m. These results imply a tradeoff between fit in data space and fit in image space, with constrained maximum entropy performing better in the key measure of fit to the true image. Looking at the results more closely, we also find that maximum likelihood does reasonably well when it fits the data exactly, and worse when it cannot.

Both methods of course fit the data or prior model better when they estimate over a finer grid; at the same time, they fit the truth less well. This makes sense in our example, because we defined the true image over the coarse grid. The constrained maximum entropy reconstructions are only slightly worse in the fine grid, however, while some maximum likelihood images fit far worse when allowed these extra degrees of freedom.

The maximum likelihood estimate (when there is no perfect fit) is found by EM (Vardi et al. (1985)). This is an iterative algorithm, each step of which increases the likelihood of the estimate (Dempster et al. (1977)). We can track the entropy and likelihood of the EM iterates, starting at a uniform image (thus moving from maximum entropy to maximum likelihood). We have examined two such plots: one that converges to an image for which $\hat{M} = y$ (and so -2LL$(\hat{M}|y) = 0$) and one for which no such image exists. Interestingly, $S(\hat{x}|m)$ decreases in each step of the algorithm, in both cases. These plots imply a tradeoff, in models, between entropy and likelihood, especially in the region near maximum likelihood, where entropy shows a great decrease. For these same iterative estimates, we have also plotted their fit $S(\hat{x}|x)$ to the true image and the corresponding fit $S(\hat{M}|M)$ to the truth in data space. Here we find that in the first few iterations, both measures of fit improve. However, as the algorithm approaches the maximum likelihood estimate, the fit in data space gets slightly worse, and the fit in image space gets much worse. This is apparently due to the spiky character of the maximum likelihood estimates and holds even in a case of a very spiky true image.[1]

4 Rationale for constrained maximum entropy

To understand this apparent tradeoff, we must explore the link between a model in image space and the data in their space. We are interested in the image x, but the data tell us only about the sampling expectations M, and nothing about x, given M. To get an image, we must estimate M from the information provided by the data, and then choose an $\hat{x}$ consistent with our estimate $\hat{M}$. We need models on data space and on image space, given data. If we do not formalize our models, we are using implicit models. Perhaps these can explain the behavior of the methods presented above.

We will embed the parameter M in a Bayesian model, and hence determine its probable values, given the data. Then we will use maximum entropy to select one image among all those consistent with M, for each value in the posterior distribution of M. We do not extend our Bayesian model to image space because, given M, inference on x would depend solely on the prior distribution. It may be more desirable to choose our image-picking criterion as such, rather than to model in the vast space of images. This is the rationale of Skilling (1986).

As mentioned above, the fineness of the reconstruction grid is specified by the analyst; in fact, there is no logical upper bound for the number N of pixels. Aside from computational difficulties, a Bayesian modeler on x may wish to keep N low to moderate the task of specifying a plausible distribution over the space of all images x in N-dimensional space. Maximum entropy appears to solve this problem easily, however. Entropy is invariant under continuous reparameterization; thus, if an image is left unchanged but is pixellized more finely, its entropy (relative to a locally uniform measure) will not change. Furthermore, this identical image has the highest entropy of all images, on the fine grid, that are consistent with the original coarse image. The simulation results presented in Table 1 imply that this invariance works to our advantage, in that the maximum entropy solution performs relatively well over a too-fine grid.

[1] Graphs are available on request.

5 Bayesian maximum entropy methods

This section discusses a maximum entropy reconstruction method based on Bayesian estimation of parameters in data space, and connects it on a theoretical level to the original approach of Gull and Daniell (1978). Our goal is to suggest an improved method, and to clarify the hidden assumptions in the old method. Assume we have a posterior distribution on $(M|y)$. Assign, to every M, the maximum entropy image *max-ent* $[x(M)]$, satisfying $Ax = M$. This yields a probability distribution of images. If we want to pick just one image, we might take $\hat{M}$ to be the posterior mean $E(M|y)$, and pick the image *max-ent* $[x(\hat{M})]$.

Gull and Daniell perform the more (computationally) difficult task of maximizing $S(x|m)$ subject to the nonlinear constraint: $\text{-2LL}(M|y) \leq C$. If we wish to follow this route, we might set C to the posterior mean of -2LL, given y. Asymptotically (that is, with A and n fixed, but with more Poisson data), $\text{-2LL}(M|y) \sim \chi^2_n$, with mean n. This gives some justification for the usual constraint value $C = n$. From the Bayesian perspective, however, we should consider the posterior distribution of -2LL, conditional on the data y. In a small sample, we would certainly prefer to set $C = E(\text{-2LL}(M|y))$, rather than $C = n$, for the constraint: $\text{-2LL}(M|y) \leq C$. In fact, one may observe data y such that $\text{-2LL}(M|y) > n$ for all positive images x.

6 Illustrative examples

This section shows the use of the methods described above as applied to three situations. We start with a simple, straightforward example and move to an approximation of the main example of this paper. The examples in this section will be based on the Normal model:

$$y_i \sim N(M_i, \sigma^2), \; i = 1, \ldots, n.$$

The sampling expectations M will again be expressed as an all-positive linear transformation of an all-positive image:

$$M = Ax, \text{ with } N \text{ pixels in } x, \; N \geq n. \; A \text{ has rank } n.$$

The fit to the data is then measured by a sum of squares:

$$\text{-2LL}(M|y) = \sum_i (M_i - y_i)^2.$$

The range of the transformation A, applied to the set of nonnegative images x, is a convex region in data space that we will call P. If $y \in P$, then there is an image (in general, an $(N - n)$-dimensional space of images) that fits the data perfectly. We put a uniform prior distribution on M, for all $M \in P$.

In our first example, we set $N = n$ and A to the identity, so $M = x$. The posterior distribution of M is truncated Normal:

$$(M_i|y_i) \sim N(y_i, \sigma^2), \text{ constrained to } M_i > 0, \text{ for } i = 1, \ldots, n.$$

If all the observations y_i are appreciably greater than σ, the truncation will be unimportant. In this case,

$$\text{-2LL}(M|y) \sim \chi^2_n.$$

For any specific $\hat{M}$, the only possible image is $\hat{x} = \hat{M}$, and the posterior distribution of maximum entropy estimates is the truncated n-dimensional Normal, centered at y.

Our second example is the same, but with the additional prior restriction that all the M_i's be equal. Thus restricted, the n-dimensional Normal posterior distribution becomes a univariate Normal on the common parameter M_i:

$$\begin{aligned} (M_i|y) &\sim N(\bar{y}, \frac{\sigma^2}{n}), \text{ constrained to } M_i > 0, \\ \text{-2LL}(M|y) &= \frac{n(M_i - \bar{y})^2}{\sigma^2} + \sum_i \frac{(y_i - \bar{y})^2}{\sigma^2}. \end{aligned}$$

Assuming $\frac{y}{\sigma}$ is sufficiently far from 0, the conditional distribution of $(\frac{n(M_i-\bar{y})^2}{\sigma^2}|y)$ is χ^2_1, and -2LL$(M|y)$ is just that random variable plus a constant that is known, given y. Unconditional on the data, this constant has expectation $E(\frac{\sum_i (y_i-\bar{y})^2}{\sigma^2}) = n - 1$.

In our third example, $N > n$ and A is a complicated matrix. P is now a convex region in data space bounded by m hyperplanes that intersect the origin. The posterior distribution of M, given y, is truncated Normal once again, but this time the truncation matters. The data y might not lie within P. Also, -2LL$(M|y)$ will no longer be approximately distributed as χ^2_n, and its expectation, given y, will most likely not be close to n. As σ^2 decreases, however, y becomes closer to the true M and less likely to be near the boundary of P. Thus, the truncation becomes less important, and as $\sigma^2 \to 0$, we return to the geometry and distribution of $(M|y)$ of the first example of this section. Of course, for any $\hat{M}$, we must still choose a maximum-entropy image $\hat{x}$ from an $(N-n)$-dimensional space satisfying $A\hat{x} = \hat{M}$. This third example is very similar to the main example of this paper, inasmuch as the Normal distribution approximates the Poisson. The asymptotic case of infinite data corresponds to $\sigma^2 \to 0$.

7 Discussion

This last example allows us to understand the problems of the maximum likelihood reconstruction. If $y \notin P$, this reconstruction will lie on the boundary of P, yielding an image with zeroes in many cells. The chance of this happening depends on how close A is to being singular, as well as on the amount of Poisson data and on the true image; it can happen even with a smooth true image. On the other hand, if $y \in P$, then the maximum likelihood estimate gives $\hat{M} = y$. This may overfit the data in those dimensions in which M is constrained to a narrow region. In such directions, the posterior density of M may be nearly constant. An estimator that fits too closely to the position of y will be subject to the latter's great sampling variability. These problems disappear asymptotically, of course. As the number of counts increases, maximum likelihood becomes as precise as any estimator of M.

Our theoretical study also explains the entropy-likelihood tradeoff in two ways. First, as -2LL is allowed to increase, a larger region in data-space, and thus in image-space,

becomes available in which to search for high-entropy images. Second, if the data y are a small sample, parameters M with likelihoods near the maximum will generally be peculiar points, probably nearer to the boundary of P than any reasonably smooth true image. This is apparently common in practice, to judge from reports of unrealistic maximum likelihood reconstructions for hypothetical and real tomography data (Vardi et al. (1984), Fox (1987)). Such behavior can make the tradeoff more extreme near the maximum of likelihood.

In conclusion, maximum entropy and maximum likelihood estimates for our problem differ; the former better fits the true image and the latter better fits the data. When fit on an overly fine grid of pixels, maximum entropy produces reasonable images; maximum likelihood does not. In light of these results, we suggest attacking our image reconstruction problem with separate analyses on data space and image space. We can first estimate our knowledge of the sampling expectations M, from the noisy data y. For any point in the posterior distribution of M, we can then choose the maximum entropy image x consistent with this incomplete information. A simple example shows the connection between this method and that of Gull and Daniell (1978) and Skilling (1986). We interpret the 'hard constraints' of the latter methods as an approximation to our Bayesian approach.

Bibliography

Dempster, A. P., Laird, N. M., and Rubin, D. B. 'Maximum Likelihood from Incomplete Data via the EM Algorithm'. *J. Royal Stat. Soc.* **B 39**, 1 (1977).

Fox, P. Private communication (1987).

Good, I. J. *The Estimation of Probabilities.* M.I.T. Press (1965).

Gull, S. F., and Daniell, G. J. 'Image Reconstruction from Incomplete and Noisy Data'. *Nature* **272**, 686 (1978).

Skilling, J. 'Theory of Maximum Entropy Image Reconstruction'. In *Maximum Entropy and Bayesian Methods in Applied Statistics*, J. H. Justice, ed. Cambridge U. P. (1986).

Vardi, Y., Shepp, L. A., and Kaufman, L. 'A Statistical Model for Positron Emission Tomography'. *J. Amer. Stat. Assoc.* **80**, 8 (1985).

ENTROPY + RAIN = FLOODS

P W JOWITT
Heriot-Watt University
Department of Civil Engineering
Edinburgh EH14 4AS, UK

ABSTRACT. The paper discusses the application of the maximum entropy formalism to catchment behaviour and flood frequency. In the former, the distribution of water within a catchment is treated as a problem of statistical inference and speculations on the consequent behaviour of catchments are addressed. In the problem of flood frequency, appropriate statistics are deduced which would, on application of the entropy formalism, recover the well-known Gumbel distribution.

1. THE FIRST PART: RAINFALL RUNOFF PROCESSES

1.1 CATCHMENT MODELS

Approaches to the dynamic modelling of hydrological catchments are diverse. They range from black box (time series) models which make no pretence to mirror the physical states and processes of the system, to mechanistic models which seek to represent such finescale processes of water movement. There are innumerable models of an intermediate and conceptual nature. Perhaps the simplest of these conceptual models is the linear reservoir, in which the *state* of the system is taken to be the volume $V = V(t)$ of water stored in the catchment, and runoff q from the catchment is assumed proportional to V such that $q = k\,V$. If inflow (rainfall) is denoted p, then mass continuity gives the familiar first order linear ordinary differential equation:

$$\frac{dV}{dt} = p - q = p - kV$$

A novel departure from the general pattern outlined above was suggested by Clarke and Moore (1981) who sought to represent the catchment as a statistical population of storages of individual size s, characterised by a probability density function $f(s)$. The present paper examines the consequences of applying the Maximum Entropy Principle to such a description at the micro level and its effect on dynamic behaviour at the macro level. Prior knowledge of the catchment is confined to just two aspects, namely the mean catchment capacity (which is assumed time-invariant), and the mean catchment wetness.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 437–445.

1.2 MAXIMUM ENTROPY FORMS FOR CATCHMENT STORAGE AND CATCHMENT WETNESS

1.2.1 *Available Catchment Storage f(s)* The form of f(s) is determined from the solution of the program:

$$\text{Maximise } S_s = -\int_0^\infty f(s) \log_e \frac{f(s)}{m(s)}\, ds$$

$$\text{subject to } \int_0^\infty f(s)ds = 1$$

$$\int_0^\infty s\, f(s)ds = \bar{s}$$

and where m(s) is an invariant measure function as defined by Jaynes (1963)
Correspondingly,

$$f(s) = m(s)\exp\,[-\lambda_o - \lambda s]$$

Scale invariance suggests m(s) = constant. Incorporating this within λ_o leads to the simplified result:

$$f(s) = \lambda\, e^{-\lambda s} \quad , \quad \lambda = 1/\bar{s}$$

1.2.2 *Water Storage f(v|s)* The form of f(v|s) is required for each value of s, subject to the requirement that the overall mean catchment water content is $\bar{v}$. Hence the mathematical program:

$$\text{Maximise } S_v = -\int_0^\infty f(s) \int_0^s f(v|s) \log_e \frac{f(v|s)}{m(v|s)}\, dv\, ds$$

$$\text{subject to } \int_0^\infty f(s) \int_0^s f(v|s)\, dv\, ds = 1$$

$$\int_0^\infty f(s) \int_0^s v\, f(v|s)dv\, ds = \bar{v}$$

and leading to $f(v|s) = m(v|s)\exp\,[-\phi_o - \phi\, v]$

Scale invariance again suggests that m(v|s) is independent of v, which, as far as f(v|s) is concerned, may be incorporated within ϕ_o.

Thus, $$f(v|s) = \frac{\phi\ e^{-\phi\ v}}{1 - e^{-\phi\ s}} \qquad \text{for all } s$$

The corresponding moment generation function M(p|s) is

$$M(p|s) = \int_o^s f(v|s)\ e^{-pv}\ dv = \frac{\phi}{(p + \phi)} \cdot \frac{[1 - e^{-(p + \phi)s}]}{[1 - e^{-\phi s}]}$$

The conditional water storage $\bar{v}|s$ is thus given by

$$\bar{v}|s = -\frac{dM}{dp}(p|s)\Big|_{p=o} = \frac{1 - e^{-\phi s} - \phi\ s\ e^{-\phi s}}{\phi\ (1 - e^{-\phi s}\ )}$$

The single parameter ϕ, which governs f(v|s) for all s, is related to the overall mean catchment water content $\bar{v}$ through

$$\int_o^\infty f(s) \int_o^s v\ f(v|s)dv\ ds \equiv \int_o^\infty f(s)\ \bar{v}|s\ ds = \bar{v}$$

$$\text{or} \quad \bar{v} = \int_o^\infty \lambda e^{-\lambda s} \left\{ \frac{1 - e^{-\phi s} - \phi\ se^{-\phi s}}{\phi(1 - e^{-\phi s})} \right\} ds$$

$$= \frac{1}{\phi} - \int_o^\infty \frac{\lambda s\ e^{-(\lambda+\phi)s}}{(1 - e^{-\phi s})}\ ds$$

1.3 CATCHMENT DYNAMICS

In the simple linear model referred to earlier, runoff was taken to be proportional to catchment storage; in the development below runoff q will again be taken to be proportional to catchment content v, *but only in those elements of storage which are full.* Thus

$$\frac{d\bar{v}}{dt} = p - q = p - k \int_o^\infty f(s)\ s\ f(v = s|s)ds$$

The catchment dynamics may also be expressed in terms of the parameter ϕ as follows:

$$\frac{d\phi}{dt} \int_0^\infty \lambda e^{-\lambda s} \left\{ \frac{\phi^2 s^2 e^{-\phi s} - (1 - e^{-\phi s})^2}{\phi^2 (1 - e^{-\phi s})^2} \right\} ds$$

$$= p(t) - k \int_0^\infty \frac{\lambda \phi \, s e^{-(\lambda + \phi)s}}{(1 - e^{-\phi s})} \, ds$$

The integrals in the above may be expressed in terms of Riemann's two parameter Zeta function $\zeta(z,q)$ and the related three-parameter Phi function $\Phi(s,z,q)$. If the dimensionless quantity $y \equiv \frac{\phi}{\lambda}$ is introduced then:

$$\frac{dy}{dt} = \frac{\lambda [p y^2 - k y \zeta(2, 1 + \frac{1}{y})}{\frac{2}{y}[\zeta(2, \frac{1}{y}) - \frac{1}{y}\zeta(3, \frac{1}{y})] - 1}, \quad y > 0$$

with $$\bar{v} = \frac{1}{\lambda y} - \frac{1}{\lambda y^2} \zeta(2, 1 + \frac{1}{y}), \quad y > 0$$

and $$q = -\frac{k}{y} \zeta(2, 1 + \frac{1}{y}) \qquad y > 0$$

Catchment storage, $\bar{v}$, and runoff, q, are related through

$$q = k(1 - \lambda y \bar{v}), \quad \text{for all } y.$$

Note also that

$$\bar{v}(-y) = 1 - \bar{v}(y), \quad \text{for all } y.$$

1.4 SOLUTIONS AND THE PROSPECT FOR MODEL VALIDATION?

The differential equation describing the variation of y with t has (at least, to this author) no obvious solution other than a numerical one, and is not attempted here. The implied variation of q with $\bar{v}$ is given in Figure 1 for $k = \lambda = 1$. The entropy model has the property that runoff increases with catchment wetness. Additionally, the available catchment storage is bounded above (at $\bar{s} = 1/\lambda$) as well as at zero (a condition which is not enjoyed by the simple linear reservoir assumption $q = kv$).

2 THE SECOND PART: FLOODS

2.1 UNCERTAINTY, EXTREMITY AND OBJECTIVITY

The Gumbel (Extreme Value Type I) distribution is just one of many used for modelling flood frequency from catchment flood data, and whilst it has impeccable credentials for the modelling of extremes, actual catchments do not appear to follow such an ideal. The Gumbel model is often rejected because its coefficient of skewness is invariant at the level 1.14 and not mirrored by catchment data.

The maximum entropy principle for generating least-biased distributions from limited information is accepted as a lesson in objectivity, at least when

a) the random events are discrete.

b) the totality of the given information can be objectively and suitably expressed, and refers to known ('population') information rather than sample information.

Condition (a) can be relaxed so long as an invariant measure function can be identified (Jaynes, 1963). Condition (b) can be relaxed to include sample information, by the application of Bayes Theorem to integrate out parameter uncertainty.

But, when the prior information comprises a set of sample data (as opposed to merely sample statistics of such data), then no obvious objective way forward exists; any set of sample moments/transformations could be calculated, each leading to a different maximum entropy distribution. Integrating out any nuisance parameters via Bayes Theorem will not lead to convergence to a unique maximum entropy form. The practitioner is thus faced with the exercise of judgement; the choice of sample moments will, as a result of applying the entropy formalism, in turn determine the statistical model. It is therefore of some interest to discover which forms of prior information lead, via Entropy Maximisation, to particular density functions.

2.2 THE EV1 DISTRIBUTION AND MAXIMUM ENTROPY

Suppose that for some values v and b, the following distribution moments μ_y, ε_y for a random variable x are known

$$\mu_y \triangleq \int_{-\infty}^{\infty} \frac{x - v}{b} f(x)dx$$

$$\varepsilon_y \triangleq \int_{-\infty}^{\infty} \exp\left[-\frac{x - v}{b}\right] f(x)dx$$

Maximising the entropy function

$$S = -k \int_{-\infty}^{\infty} f(x) \log_e \frac{f(x)}{m(x)} \, dx$$

subject to the above constraints (together with the normality condition), and setting m(x) to unity yields

$$f(x) = \frac{\lambda_2^{\lambda_1}}{b\,\Gamma(\lambda_1)} \exp\left[-\lambda_1\left(\frac{x-v}{b}\right) - \lambda_2 \exp\left[-\left(\frac{x-v}{b}\right)\right]\right]$$

where λ_1 and λ_2 are the Lagrangian multipliers associated with the constraints on μ_y and ε_y respectively.

Expressions for μ_y and ε_y in terms of λ_1 and λ_2 are obtained from the partition function $\exp[\lambda_o]$ through $-\frac{\partial\lambda o}{\partial\lambda 1}$ and $-\frac{\partial\lambda o}{\partial\lambda 2}$ yielding:

$$\mu_y = \log_e \lambda_2 - \Psi(\lambda_1) \equiv E\left[\frac{x-v}{b}\right]$$

and

$$\varepsilon_y = \frac{\lambda_1}{\lambda_2} \equiv E\left[\exp\left[-\left\{\frac{x-v}{b}\right\}\right]\right]$$

where $\Psi(.)$ is the digamma function.

The expression for f(x) reduces to the EV1 distribution when $\lambda_1 = \lambda_2 = 1$. Thus, the EV1 distribution is seen to be a maximum entropy distribution when v and b have values u and α which produce the moments

$$\mu_y = 0.5772$$

and

$$\varepsilon_y = 1$$

2.3 PARAMETER ESTIMATION METHODS

The most common ways of estimating u and α from a set of sample data are the methods of moments (using sample estimates of the mean and variance) and maximum likelihood. In 1946 Kimball noted that the EV1 does not possess a pair of sufficient statistics and suggested an alternative scheme, employing what he termed "sufficient statistical estimation functions", X and Y, which when set to zero yield estimates of the unknown parameters u and α (Kimball, 1946).

viz:

$$X = \sqrt{\frac{1}{n}}\left[\frac{(\bar{x} - u)}{\alpha} - 0.5772\right]$$

$$Y = \sqrt{\frac{1}{n}}\left[\bar{\varepsilon}_y - 1\right]$$

where $\bar{x}$ is the sample mean

and $$\bar{\varepsilon}_y = \frac{1}{n}\sum_{i=1}^{n} \exp\left[-\frac{(x_i - u)}{\alpha}\right]$$

Kimball shows that if x is EV1 with parameters u and α, then X, Y are asymptotically Gaussian with covariance matrix

$$\sigma_{xy} = \begin{bmatrix} \frac{\pi^2}{6} & , & -1 \\ -1 & , & 1 \end{bmatrix}$$

2.4 ENTROPY-BASED PARAMETER ESTIMATES

The moments μ_y and ε_y have been shown in Section 2.2 to yield the EV1 distribution after entropy maximisation; it therefore seems natural to use sample estimates of these same two moments for parameter estimation:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{\bar{x} - u}{\alpha}\right]$$

$$\bar{\varepsilon}_y = \frac{1}{n}\sum_{i=1}^{n} \exp\left[-\left[\frac{x_i - u}{\alpha}\right]\right]$$

and seeking values u and α to yield $\bar{y} = 0.5772$ and $\bar{\varepsilon} = 1$.

This pair of estimation equations are seen to be entirely equivalent to Kimball's sufficient statistical estimation functions X and Y. To paraphrase Jaynes (1976), "Kimball was right after all." Practical estimation of u and α is straightforward (Jowitt (1978)).

2.5 THE EV1 DISTRIBUTION

The Gumbel (EV1) distribution has a skewness coefficient of 1.14 for all u and α. The related EV2 and EV3 distributions admit variations of skewness. All three distributions are regenerative, in that if x is EVn, then the extreme value of N samples of x is also EVn.

The generalised form of the maximum entropy distribution outlined earlie can be described in terms of the Moment Generating Function:

$$M(s) = \int_0^\infty e^{sy} f(y)\, dy = \frac{\lambda_2^{\lambda_1}}{\Gamma(\lambda_1)} \cdot \lambda_2^{(\lambda_1 - s)} \cdot \Gamma(\lambda_1 - s)$$

The resulting skewness coefficient is

$$\gamma = \frac{-\Psi^{(2)}(\lambda_1)}{\left[\Psi^{(1)}(\lambda_1)\right]^{3/2}} \qquad 0 \leq \gamma \leq 2$$

where $\Psi^{(1)}(.)$, $\Psi^{(2)}(.)$ are the Trigamma and Tetragamma functions respectively. The usual Gumbel (EV1) distribution ($\gamma = 1.14$) results in the case $\lambda_1 = 1$.

SUMMARY REMARKS ON THE GUMBEL AND THE EV1 DISTRIBUTION

The EVλ distribution has stemmed from a study of the Gumbel (EV1) distribution and its connection with entropy maximisation. The distribution, whilst not generally regenerative, does offer variable skewness. The parameter estimations scheme for the specific case of the EV1 distribution is seen to be in complete correspondence to the estimation functions suggested by Kimball in 1946.

Whenever statistical estimation is based on a set of sample data, there is no unique choice of (sample) moments to constrain the entropy function. In effect, the practitioner's choice of moments is equivalent to choosing the density function. This being so it is informative to determine that set of prior constraints which, via entropy maximisation, would lead to a particular and commonly used-distribution.

REFERENCES

Clark R T and Moore R J (1981)
'A distribution function approach to rainfall runoff modelling'
Water Resources Research, **17**, 5, pp 1367-1382

Jaynes E T (1963)
'Information theory and statistical mechanics'
In: *Statistical Physics*, Vol 3, K W Ford (Ed), W A Benjamin Inc, New York,Ch 4, pp 182-218

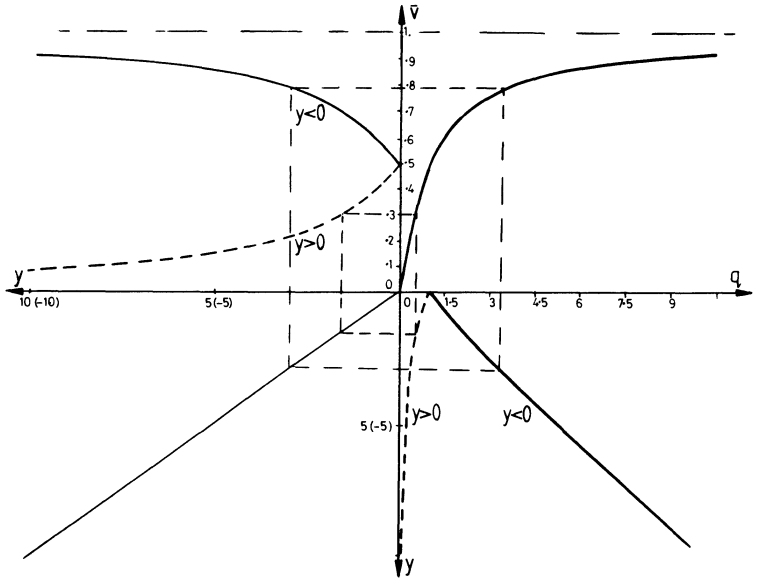
Jaynes E T (1976)
'Jaynes' reply to Kempthornes' comments'
'Confidence intervals vs Bayesian Intervals'
Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science, W L Harper and C A Hooker (Eds), D Reidel Publishing Co, Dordrecht, Holland, p 243

Jowitt P W (1979)
'The extreme value Type 1 distribution and the Principle of Maximum Entropy' *Journal of Hydrology*, **42**, pp 23-38

Kimball B F (1946)
'Sufficient statistical estimation function for the parameters of the distribution of maximum values'
Annals of Mathematical Statistics, **13**, pp 318-325

Mantzos D M (1979)
'The EVλ distribution: an extension of the EV1 distribution'
MSc thesis, Univ of London

FIGURE 1: Co-axial diagram relating runoff q, catchment wetness $\bar{v}$, and the dimensionless parameter y



MAXIMUM ENTROPY AND CONSTRAINED OPTIMIZATION

A.B. TEMPLEMAN and LI XINGSI
Department of Civil Engineering
University of Liverpool
P. O. Box 147
Liverpool L69 3BX

ABSTRACT. Previous work by the present authors [1,2] has introduced the idea that problems of constrained non-linear programming, which have hitherto been treated entirely deterministically in respect of the development of solution methods, may be interpreted probabilistically and solved by appropriate methods employing entropy maximization. This paper gives formal proofs by entirely deterministic mathematical means of the results contained in the earlier work and removes the need for any probabilistic interpretation. This consequently establishes the research on a much firmer base and also implies that the probabilistic interpretation and use of entropy maximization, though no longer strictly essential, is nonetheless valid.

1. INTRODUCTION

The constrained optimization problem studied in this paper is that of finding a local solution $\underset{\sim}{x}^*$ of Problem P:

$$\text{Minimize:} \quad \underset{\underset{\sim}{x}}{} \; f(\underset{\sim}{x}) \tag{1}$$

$$\text{Subject to:} \quad g_j(\underset{\sim}{x}) \leqslant 0 \qquad j=1,..,m \tag{2}$$

where $f(\underset{\sim}{x})$ and $g_j(\underset{\sim}{x})$, $j=1,..,m$ denote real-valued smooth functions of a vector $\underset{\sim}{x} \equiv x_i$, $i=1,..n$. The approach used herein centres around solving problem P indirectly through the use of the surrogate problem, Problem S:

$$\text{Minimize:} \quad \underset{\underset{\sim}{x}}{} \; f(\underset{\sim}{x})$$

$$\text{Subject to:} \quad \sum_{j=1}^{m} \lambda_j g_j(\underset{\sim}{x}) \leqslant 0 \tag{3}$$

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 447–454.

The single surrogate constraint (3) is defined as a positive linear combination of the original constraints (2). $\underset{\sim}{\lambda}$ is an m-vector of non-negative weights, named surrogate multipliers, which may be normalised without loss of generality and which therefore satisfy:

$$\sum_{j=1}^{m} \lambda_j = 1 \quad ; \quad \lambda_j \geq 0 \qquad\qquad j=1,\ldots,m \qquad (4)$$

The idea underlying constraint surrogation is that the single constraint (3) stands in place of the m constraints (2). The various activity levels and interactions among the constraints (2) can be captured and represented by the surrogate multipliers $\underset{\sim}{\lambda}$ and if values for these are chosen correctly then the solutions to problems P and S will be equivalent. Many previous authors have explored the nature of this equivalence [3-5].

This paper is concerned with methods for finding the correct values for the surrogate multipliers, $\underset{\sim}{\lambda}^*$, such that $\underset{\sim}{x}^*$ which solves problem S with $\underset{\sim}{\lambda}^*$ also solves problem P. Previous work by the present authors [1,2] has used a maximum entropy approach to find $\underset{\sim}{\lambda}^*$. The validity of this method depends upon a probabilistic interpretation of the surrogate multipliers. From this probabilistic interpretation several useful results and an iterative solution method for problem P can be derived.

In this paper a brief survey is made of the work of Refs. [1,2] establishing the probabilistic, entropy-based context of problem S. Then some new proofs of the main results are presented which are entirely deterministic in nature and remove the need for any intermediate assumptions about the probabilistic nature of the surrogate multipliers.

2. A MAXIMUM ENTROPY APPROACH

The Lagrangean of problem S is

$$L_s(\underset{\sim}{x},\alpha,\underset{\sim}{\lambda}) = f(\underset{\sim}{x}) + \alpha \sum_{j=1}^{m} \lambda_j g_j(\underset{\sim}{x}) \qquad (5)$$

in which α is the Lagrange multiplier for constraint (3). An essential condition for the equivalence of problems P and S is that L_s must satisfy the Lagrangean saddle-point condition:

$$L_s(\underset{\sim}{x},\alpha^*,\underset{\sim}{\lambda}^*) \geq L_s(\underset{\sim}{x}^*,\alpha^*,\underset{\sim}{\lambda}^*) \geq L_s(\underset{\sim}{x}^*,\alpha,\underset{\sim}{\lambda}) \qquad (6)$$

This saddle-point condition implies that a two-phase iterative approach can be used to solve problem P via S. A typical scheme involves choosing an initial set of surrogate multipliers $\underset{\sim}{\lambda}^{[o]}$ and solving problem S to give a corresponding set of values $\underset{\sim}{x}^{[o]}$ (by minimization

over $\underset{\sim}{x}$, corresponding to the left-hand inequality in (6)). The multipliers are then updated to $\underset{\sim}{\lambda}^{[1]}$ (by some maximization process, corresponding to the right-hand inequality in (6)) and problem S is solved again to give $\underset{\sim}{x}^{[1]}$. The process is repeated until the sequence $(\underset{\sim}{\lambda}^{[0]}, \underset{\sim}{x}^{[0]})$, $(\underset{\sim}{\lambda}^{[1]}, \underset{\sim}{x}^{[1]}),\ldots$ converges upon a solution of S and hence also of P, at $(\underset{\sim}{\lambda}^{*}, \underset{\sim}{x}^{*})$. The main difficulty in this scheme resides in finding an updating scheme to generate the sequence of surrogate multipliers $\underset{\sim}{\lambda}^{[0]}, \underset{\sim}{\lambda}^{[1]},\ldots$

In Ref. [1] entropy maximization was used to generate $\underset{\sim}{\lambda}$ updates. It was noted that if the m components of $\underset{\sim}{\lambda}$ are interpreted as discrete probabilities then (4) represents the axiomatic normality and non-negativity conditions of such probabilities. For the case of problem P in which at least one of the constraints (2) is active at the optimum -the case most frequently occurring in optimum engineering design applications and that under consideration throughout this work - it can easily be proved that constraint (3) must be an equality. Consequently, constraint (3) has the form of an expected value constraint. With this probabilistic view of problem S least biased estimates for $\underset{\sim}{\lambda}^{[k]}$ at the k-th iteration can be found by using the maximum entropy formalism [6], effectively by solving the problem:

$$\underset{\underset{\sim}{\lambda}^{[k]}}{\text{Maximize:}} \quad S = -K \sum_{j=1}^{m} \lambda_j^{[k]} \ell n \lambda_j^{[k]} \tag{7}$$

$$\text{Subject to:} \quad \sum_{j=1}^{m} \lambda_j^{[k]} = 1 \tag{8}$$

$$\sum_{j=1}^{m} \lambda_j^{[k]} g_j(\underset{\sim}{x}^{[k-1]}) = \epsilon \tag{9}$$

in which S is the Shannon entropy [7] and K a positive constant. ϵ in Eq. (9) is an error term reflecting the fact that constraint function values $g^{[k-1]}$ have been used in place of $g^{[k]}$ which are not yet available. ϵ should be small, positive and decrease towards zero as iterations proceed. Values of $\underset{\sim}{\lambda}^{[k]}$ which solve (7)-(9) are given by:

$$\lambda_j^{[k]} = \exp\left[\beta g_j(\underset{\sim}{x}^{[k-1]})/K\right] \Big/ \sum_{j=1}^{m} \exp\left[\beta g_j(\underset{\sim}{x}^{[k-1]})/K\right] \quad j=1,..m \tag{10}$$

in which β is the Lagrange multiplier for Eq. (9). Since ϵ is not uniquely known and K is a positive constant, β/K may be considered as a control parameter. For ϵ to display the desired convergence characteristics β/K must be positive and increase towards infinity with successive iterations.

Eq. (10) represents the entropy-based updating formula for the surrogate multipliers $\underset{\sim}{\lambda}$ in the two-phase iterative solution scheme described above. Full details of the work are given in Refs. [1,2].

An alternative way of incorporating entropy is directly to augment $f(\underset{\sim}{x})$ in problem S with a multiplier entropy term and treat the resulting problem as a minimax optimization over both sets of variables $\underset{\sim}{x}$ and $\underset{\sim}{\lambda}$. Problem S then becomes Problem SA:

$$\underset{\underset{\sim}{x}}{\text{Minimize}}\ \underset{\underset{\sim}{\lambda}}{\text{Maximize:}}\quad f(\underset{\sim}{x}) - (1/p)\sum_{j=1}^{m} \lambda_j \ell n\, \lambda_j \tag{11}$$

$$\text{Subject to:}\quad \sum_{j=1}^{m} \lambda_j = 1$$

$$\sum_{j=1}^{m} \lambda_j g_j(\underset{\sim}{x}) = 0$$

In (11) p is a positive constant. The Lagrangean of problem SA is

$$L_{SA} = f(\underset{\sim}{x}) + \alpha \sum_{j=1}^{m} \lambda_j g_j(\underset{\sim}{x}) + \gamma \left[\sum_{j=1}^{m} \lambda_j - 1\right] - (1/p)\sum_{j=1}^{m} \lambda_j \ell n \lambda_j \tag{12}$$

which is seen to be the same as L_s (Eq. (5)) augmented by multiplier normality and entropy expressions. The minimax nature of problem SA imposes the necessary saddle-point condition. Stationarity of L_{SA} with respect to λ_j and γ gives

$$\lambda_j = \exp[p\alpha g_j(\underset{\sim}{x})] / \sum_{j=1}^{m} \exp[p\alpha g_j(x)] \qquad j=1,..,m \tag{13}$$

Eq. (13) is seen to be similar to Eq. (10), the $\underset{\sim}{\lambda}$ update formulae in the two-phase method, when β/K is replaced by $p\alpha$. Substituting (13) into L_{SA} gives, after algebraic manipulation,

$$L^{*}_{SA} = f(\underset{\sim}{x}) + (1/p)\ell n \sum_{j=1}^{m} \exp[p\alpha g_j(\underset{\sim}{x})] \tag{14}$$

Minimization of the above L^{*}_{SA} over variables $\underset{\sim}{x}$ and with $p\alpha$ taking an increasing positive sequence of values tending towards infinity yields the solution of problem P.

This entropy augmentation approach has effectively collapsed the two-phase iterative method into a single phase. Eq. (14) has the appearance of a penalty function formulation of problem P in which the penalty term has been derived on the basis of maximizing the entropy of the surrogate multipliers.

All the above results have hitherto rested upon and flowed from the validity of the assumption that the surrogate multipliers $\underset{\sim}{\lambda}$ in problem S can be treated as probabilities. This probabilistic view alone justifies the introduction of the Shannon entropy and its maximization in this problem context. The next section of the paper presents new

proofs of the results derived earlier, proofs which are entirely deterministic and do not require any probabilistic interpretations.

3. A DETERMINISTIC APPROACH

The following theorem is essential to the development of alternative proofs of the main results of section 2.

Theorem

For any set of positive numbers, $U_1,..,U_j,..,U_m$ and weights $\lambda_1,..,\lambda_j,..\lambda_m$ satisfying

$$\sum_{j=1}^{m} \lambda_j = 1, \qquad \lambda_j \geq 0 \ \forall j:$$

$$\ell n \left[\sum_{j=1}^{m} U_j \right] \geq \sum_{j=1}^{m} \lambda_j \ell n U_j - \sum_{j=1}^{m} \lambda_j \ell n \lambda_j \tag{15}$$

with equality when the right-hand side of the inequality is maximized over λ_j, $j=1,..,m$.

Proof

The theorem and its proof are consequences of Cauchy's inequality (the arithmetic-geometric mean inequality) [8] which states that for U_j, λ_j, $j=1,..,m$ as defined in the theorem

$$\sum_{j=1}^{m} U_j \geq \prod_{j=1}^{m} (U_j/\lambda_j)^{\lambda_j} \tag{16}$$

Taking natural logarithms of (16) gives

$$\ell n \left[\sum_{j=1}^{m} U_j \right] \geq \sum_{j=1}^{m} \ell n U_j^{\lambda_j} + \sum_{j=1}^{m} \ell n \lambda_j^{-\lambda_j}$$

Hence

$$\ell n \left[\sum_{j=1}^{m} U_j \right] \geq \sum_{j=1}^{m} \lambda_j \ell n U_j - \sum_{j=1}^{m} \lambda_j \ell n \lambda_j$$

and the first part of the theorem is proved.

The second part requires the maximization over λ_j, $j=1,..,m$ of the right-hand side of the inequality subject to normality and non-negativity of the weights. The Lagrangean of this problem is

$$L(\underset{\sim}{\lambda}, \delta) = \sum_{j=1}^{m} \lambda_j \ell n U_j - \sum_{j=1}^{m} \lambda_j \ell n \lambda_j + \delta \left[\sum_{j=1}^{m} \lambda_j - 1 \right] \qquad (17)$$

There is no need to include the non-negativity conditions explicitly as the middle term of L imposes this. Stationarity of L with respect to λ_j, j=1,..,m and δ leads to

$$\lambda_j = U_j / \sum_{j=1}^{m} U_j \qquad j=1,..,m \qquad (18)$$

Result (18) can be shown to be a maximizing point of the right-hand side of (15) by examining the second derivative matrix of L which is negative definite.

Substituting (18) into the right-hand side of (15) gives

$$\underset{\lambda_j, j=1,..,m}{\text{Maximum}} \left[\sum_{j=1}^{m} \lambda_j \ell n U_j - \sum_{j=1}^{m} \lambda_j \ell n \lambda_j \right] = \ell n \left[\sum_{j=1}^{m} U_j \right]$$

after algebraic simplification, and the theorem is proved.

Interesting features of the theorem are the appearance of the Shannon entropy function of the weights $\underset{\sim}{\lambda}$ in the right-hand side of (15), and the maximization over $\underset{\sim}{\lambda}$ of that right-hand side function containing entropy. The context is entirely deterministic, the $\underset{\sim}{\lambda}$ are weights not probabilities.

The theorem is now used to establish some of the results of section 2. Let

$$U_j = \exp(p[f(\underset{\sim}{x}) + \alpha g_j(\underset{\sim}{x})]) \qquad j=1,..,m \qquad (19)$$

where $f(\underset{\sim}{x})$ and $g_j(\underset{\sim}{x})$, j=1,..,m are as defined for problem P, α is defined as in (5) and p is a positive constant. Defining U_j by (19) imposes restrictions upon the dimensionalities of $f(\underset{\sim}{x})$ and $\underset{\sim}{g}(\underset{\sim}{x})$ which may represent different physical quantities. Clearly $f(\underset{\sim}{x}) + \alpha g_j(\underset{\sim}{x})$ must be dimensionally homogeneous for all j and, since α is a constant in all such expressions, the only way in which this homogeneity can be generally assumed is for $f(\underset{\sim}{x})$ and each of the functions $g_j(\underset{\sim}{x})$ in problem P to be dimensionless. It is assumed hereafter that functions $f(\underset{\sim}{x})$ and $\underset{\sim}{g}(\underset{\sim}{x})$ in problem P, and hence in (19) have been appropriately non-dimensionalised. All U_j defined by (19) are then positive numbers as required by the theorem.

For U_j as defined by (19) the theorem states that

$$\ell n \sum_{j=1}^{m} \exp(p[f(\underset{\sim}{x}) + \alpha g_j(\underset{\sim}{x})]) \geq \sum_{j=1}^{m} \lambda_j p[f(\underset{\sim}{x}) + \alpha g_j(\underset{\sim}{x})] - \sum_{j=1}^{m} \lambda_j \ell n \lambda_j$$

$$\therefore pf(\underset{\sim}{x}) + \ell n \sum_{j=1}^{m} \exp[p\alpha g_j(\underset{\sim}{x})] \geq pf(\underset{\sim}{x}) \sum_{j=1}^{m} \lambda_j + \alpha p \sum_{j=1}^{m} \lambda_j g_j(\underset{\sim}{x}) - \sum_{j=1}^{m} \lambda_j \ell n \lambda_j$$

Dividing through by p and using the normality of $\underset{\sim}{\lambda}$ gives

$$f(\underset{\sim}{x}) + (1/p)\ell n \sum_{j=1}^{m} \exp[p\alpha g_j(\underset{\sim}{x})] \geq f(\underset{\sim}{x}) + \alpha \sum_{j=1}^{m} \lambda_j g_j(\underset{\sim}{x}) - (1/p) \sum_{j=1}^{m} \lambda_j \ell n \lambda_j \quad (20)$$

The left-hand side is identical to L^*_{SA}, Eq. (14), and the right-hand side is identical to L_{SA}, Eq. (12), since the weight normalization term is zero. Thus, for any set of weights $\lambda_j, j=1,..,m$:

$$L^*_{SA} \geq L_{SA} \quad (21)$$

The theorem also states that (21) becomes an equality when L_{SA} is maximized over the weights. It may also be noted that values of the weights corresponding to this equality are given by Eq. (18) with the $U_j, j=1,..,m$ defined by (19). This substitution yields results (13) for the optimal weights as in section 2.

Inequality (21) may be further extended to

$$L^*_{SA} \geq L_{SA} \geq L_S \quad (22)$$

where L_S is the Lagrangean function (5) of problem S. L_{SA} is simply L_S augmented by a weight normalising term (which is always zero) and by a weight entropy term which must always be non-negative.

It is evident from (22) that L_{SA} and L^*_{SA} are upper bounds to the surrogate Lagrangean L_S for any $\underset{\sim}{\lambda}$ and p. Examining L_{SA}, Eq. (12), more closely it can be seen that L_{SA} decreases as p increases until at $p = \infty$ $L_{SA} = L_S$. Furthermore, the weights corresponding to a maximum value of L_{SA} yield L^*_{SA}, Eq. (14). Consequently the three functions in inequalities (22) all become equal as p tends to infinity and as the weights take their optimal values. All the results of section 2 are therefore encapsulated in (22).

4. DISCUSSION

The work described in section 2 of this paper was based upon assuming that the surrogate multipliers $\underset{\sim}{\lambda}$ in problem S can be interpreted as probabilities and that optimal values for them, corresponding to a solution of problem P via S, can be obtained by entropy maximization. The Shannon entropy function was introduced artificially to achieve these optimal values of $\underset{\sim}{\lambda}$.

Section 3 has shown that all the formulations of section 2 can be generated entirely logically and deterministically through the use of Cauchy's inequality. No probabilistic interpretations need be made; section 3 deals only with normalised, non-negative weights. Nevertheless, the Shannon entropy function has emerged from the mathematics as

an integral part of the proofs and not as an artificially introduced element. The results of section 3 are identical to those of section 2 and they therefore justify the probabilistic and entropic assumptions made in earlier work [1].

It is important to stress that Refs [1,2] were very exploratory in nature. No previous work existed to suggest that classical deterministic non-linear constrained optimization problems might be connected in any way to entropy maximization. No previous work known to the authors had suggested that normalised non-negative weights might be interpreted as probabilities and exploited through probabilistic methods. In Refs [1,2] and in this paper those links between entropy maximization and constrained optimization have now been established and it is clear that normalised surrogate multipliers can be interpreted as probabilities and their optimum values can be inferred by probabilistic methods. Algorithmic aspects of the work, i.e. development of robust methods for the computational solution of optimization problems based upon the probabilistic approach, has thus far taken second place to establishing the validity of the approach itself but has now commenced.

In conclusion it is noted that instances of problem P in which there are very many variables x_i, very many constraints g_j, and in which all the functions $f(\underset{\sim}{x})$ and $g(\underset{\sim}{x})$ are highly nonlinear are common throughout engineering design. Such problems are very hard to solve by conventional optimization algorithms. Perhaps the most interesting aspect of this research is that it has established a new way of approaching a difficult classical problem. Entropy maximization and general constrained optimization problems are closely linked and the nature of those links is worth exploring further.

REFERENCES

1. A.B.Templeman and Li Xingsi (1987) 'A maximum entropy approach to constrained nonlinear programming'. Engineering Optimization, Vol. **12**, No. 2, pp. 191-205.
2. Li Xingsi (1987) 'Entropy and optimization'. Ph.D. Thesis, University of Liverpool.
3. F. Glover (1968) 'Surrogate constraints'. Operations Research, Vol. **16**, pp. 741-749.
4. F.J. Gould (1969) 'Extensions of Lagrange multipliers in nonlinear programming'. SIAM Journal of Applied Mathematics, Vol. **13**, No. 6, pp. 1280-1297.
5. H.J. Greenberg and W.P. Pierskalla (1970) 'Surrogate mathematical programming'. Operations Research, Vol. **18**, pp. 924-939.
6. E.T. Jaynes (1957) 'Information theory and statistical mechanics'. The Physical Review, Vol. **106**, pp. 620-630 and Vol. **108**, pp. 171-190.
7. C.E. Shannon (1948) 'A mathematical theory of communication', Bell System Technical Journal, Vol. **27**, No. 3, pp. 379-428.
8. G.H. Hardy, J.E. Littlewood and G. Polya (1934) Inequalities, Cambridge University Press.

The Eigenvalues of Mega-dimensional Matrices

JOHN SKILLING
ST JOHN'S COLLEGE, CAMBRIDGE CB2 1TP, ENGLAND

ABSTRACT

Often, we need to know some integral property of the eigenvalues $\{x\}$ of a large $N \times N$ symmetric matrix $\mathbf{A}$. For example, determinants $det(A) = exp(\sum log(x))$ play a role in the classic maximum entropy algorithm [Gull, 1988]. Likewise in physics, the specific heat of a system is a temperature- -dependent sum over the eigenvalues of the Hamiltonian matrix. However, the matrix may be so large that direct $O(N^3)$ calculation of all N eigenvalues is prohibited. Indeed, if $\mathbf{A}$ is coded as a "fast" procedure, then $O(N^2)$ operations may also be prohibited.

Then the *only* permitted use of $\mathbf{A}$ is to apply it to one or a few vectors $\mathbf{v}_0, \mathbf{v}_1, \mathbf{v}_2, \ldots$. We use the resulting vectors in an entropic Bayesian algorithm to estimate the eigenvalue spectrum of $\mathbf{A}$, and thence its integral properties.

A million–by–million matrix is used as an example.

Introductory ideas.

We start with the premise that the symmetric matrix $\mathbf{A}$ is so large that it can only be used to generate some fairly small set of vectors $\mathbf{Av}_0, \mathbf{Av}_1, \ldots, \mathbf{Av}_\mathbf{m}$. The eigenvalue spectrum of $\mathbf{A}$ is invariant under orthogonal coordinate rotations. Hence any information on eigenvalues which resides in the vector set $\{\mathbf{Av}\}$ must share this invariance, so must be contained in scalar products. At least one externally supplied vector is needed to "seed" the vector set, and consistency demands that our state of knowledge of this seed should be rotationally invariant. Hence a seed vector, $\mathbf{r}$ say, must be drawn from a spherically symmetric probability distribution: we may conveniently set

$$pr(r) = (2\pi)^{-\frac{1}{2}N} \; exp\left(-\frac{1}{2}\sum r_i^2\right)$$

with each component r_i of $\mathbf{r}$ drawn independently from the unit normal distribution. Each new seed vector introduces more randomness, so that for any fixed size of vector set, fewer of the scalar products are informative. We shall treat the case in which just one seed vector is used, the extension to more being straightforward.

The vector set we must then use consists of $m + 1$ vectors

$$\mathbf{v_j} = \mathbf{A}^j\mathbf{r} \;,\; j = 0, 1, ..., m$$

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 455–466.

and the only quantities relevant to eigenvalues are their scalar products $\mathbf{v_i}^T\mathbf{v_j}$, which define $2m+1$ data

$$D_k = \mathbf{r}^T\,\mathbf{A}^k\mathbf{r} \qquad , \qquad k = 0, 1, ..., 2m$$

In terms of the eigenvalues $x_1, x_2, \ldots, x_N$,

$$D_k = \sum_{i=1}^{N} r_i^2\; x_i^k \qquad , \qquad k = 0, 1, \ldots, 2m$$

where r_i is the ith component of $\mathbf{r}$ in the diagonal frame of $\mathbf{A}$, itself a random sample from the unit normal distribution. Thus the data are probabilistic estimates of successive moments

$$M_k = \int dx f(x)\, x^k$$

of the eigenvalue spectrum f, $f(x)dx$ being the number of eigenvalues lying in range dx. Our task is to infer f from the data. With complete knowledge of all N eigenvalues, our estimate of f would become a sum of N delta functions, but of course we will have to settle for less than this certainty.

Formalism.

For both algebraic and arithmetical reasons, it is better to work with suitable polynomials $P_k(x)$ of order k $(k = 0, 1, 2, \ldots)$ instead of with raw powers x^k. The vector set becomes

$$\mathbf{v}_j = P_j(A)\;\mathbf{r} \qquad , \qquad j = 0, 1, \ldots, m$$

generated as are the P_j. For example Tchebyshev polynomials $\mathcal{T}_j$ are generated by

$$\mathcal{T}_0(x) = 1 \qquad , \qquad \mathcal{T}_1(x) = x \qquad , \qquad \mathcal{T}_{j+1}(x) = 2x\mathcal{T}_j(x) - \mathcal{T}_{j-1}(x)$$

so that in this case

$$\mathbf{v}_0 = \mathbf{r} \qquad , \qquad \mathbf{v}_1 = \mathbf{A}\mathbf{v}_0 \qquad , \qquad \mathbf{v}_{j+1} = 2\mathbf{A}\mathbf{v}_j - \mathbf{v}_{j-1}$$

The scalar products can be manipulated to give

$$D_k = \mathbf{r}^T P_k(A)\;\mathbf{r} \qquad , \qquad k = 0, 1, \ldots, 2m$$

For example, Tchebyshev polynomials have

$$\mathcal{T}_{2j+1}(x) = 2\mathcal{T}_j(x)^2 - 1 \qquad , \qquad \mathcal{T}_{2j+1}(x) = 2\mathcal{T}_j(x)\,\mathcal{T}_{j+1}(x) - \mathcal{T}_1(x)$$

so that in this case

$$D_{2j} = 2\mathbf{v}_j^T \mathbf{v}_j - \mathbf{v}_0^T \mathbf{v}_0 \quad , \quad D_{2j+1} = 2\mathbf{v}_j^T \mathbf{v}_{j+1} - \mathbf{v}_0^T \mathbf{v}_1$$

Whatever polynomials are chosen, D_k estimates

$$M_k = \int dx \; f(x) \; P_k(x)$$

It is algebraically awkward to work with the full form of

$$pr(D|\{x\}) = \int d^N r \; pr(r) \; pr(D|r, \{x\})$$

but we can use the mean and covariance

$$\langle D_k \rangle = M_k \quad , \quad < (D_j - \langle D_j \rangle) \; (D_k - \langle D_k \rangle) >= 2K_{jk}$$

where

$$K_{jk} = \int dx \; f(x) \; P_j(x) \; P_k(x)$$

to set

$$pr(D|f) = (4\pi)^{-\frac{1}{2}N} (det \; K)^{-\frac{1}{2}} \; exp(-L(f, D))$$
$$L(f, D) = \frac{1}{4} (D - M)^T K^{-1} (D - M)$$

in the usual Gaussian approximation (derived from maximum entropy). With Tchebyshev polynomials, K becomes bi–Toeplitz

$$K_{jk} = \frac{1}{2} \left(M_{j+k} + M_{|j-k|} \right)$$

so that its inversion can be performed in only $O\left(m^2\right)$ operations. Alternatively, with polynomials Q which are orthonormal over f,

$$\int dx \; f(x) \; Q_j(x) Q_k(x) = \delta_{jk}$$

K becomes unit and its inversion is trivial. (Regrettably Q can not be defined until f is known!)

Classic entropic spectrum

As usual in data analysis, we need to invert $pr\,(D|f)$ to estimate $pr\,(f|D)$, and to do this we need the prior $pr\,(f)$. Fortunately, this problem is already solved. Like any other number density, f is positive and additive. Hence the classic maximum entropy analysis [Gull, 1988] applies. The prior on $f\,(x)$ is entropic

$$pr\,(f|\alpha) \;\propto\; exp(\alpha S\,(f))$$

where α is an initially unknown constant and S is the entropy

$$S\,(f,m) = \int dx\;\; (f\,(x) - m\,(x) - log\,(f\,(x)\,/m\,(x))\;)$$

relative to a pre–assigned measure or "model" $m\,(x)$. Normally, one would most naturally use a flat model

$$m\,(x) = \text{constant} = \frac{N}{(x_{max} - x_{min})}$$

within some range (x_{min}, x_{max}) large enough to cover all plausible eigenvalues. Given this model, we have the classic posterior

$$pr\,(f,\alpha|D) \;\propto\; \prod_i (\alpha/(\alpha+\lambda_i))^{\frac{1}{2}} \;\cdot\; exp(\alpha S\,(f) - L\,(f))$$

where the λ_i are the eigenvalues of $(-\nabla\nabla S)^{-\frac{1}{2}}\;\nabla\nabla L\,(-\nabla\nabla S)^{-\frac{1}{2}}$. Here, the prior factor $pr\,(\alpha)$ has been ignored, because any reasonable prior would be overwhelmed by the other terms.

To maximise this posterior over f, it is simpler to work with the polynomial expansion u of $log\,(f)$,

$$log\;f\,(x) = log\;m\,(x) + \sum_{j=0}^{\infty} u_j P_j\,(x)$$

Working within the Gaussian approximation and not differentiating K, the gradients and curvatures are

$$\partial S/\partial u_i = K_{ij}u_j \qquad , \qquad \partial^2 S/\partial u_i \partial u_j = -K_{ij}$$

$$\partial L/\partial u_i = \begin{cases} \frac{1}{2}\,(M_i - D_i)\;, & i \le 2m \\ 0, & \text{otherwise} \end{cases}$$

$$\partial^2 L/\partial u_i \partial u_j = \begin{cases} \frac{1}{2}K_{ij}\;, & i,j \le 2m \\ 0 & \text{otherwise} \end{cases}$$

Accordingly, the algorithm

$$u_i \longleftarrow \begin{cases} u_i - \left\{\alpha u_i + \frac{1}{2} K_{ij}^{-1} (M_j - D_j)\right\} / \left(\alpha + \frac{1}{2}\right) , & i \leq 2m \\ 0 & \text{otherwise} \end{cases}$$

will maximise the posterior at given α, though as a detail it is advisable to restrict δu to lie within a trust region defined by

$$\delta u^T K \, \delta u \leq O(N)$$

Furthermore, the polynomials Q_i orthonormal over f form a set of orthonormal eigenvectors of $(-\nabla\nabla S)^{-\frac{1}{2}} \nabla\nabla L (\nabla\nabla S)^{-\frac{1}{2}}$ with eigenvalues

$$\lambda_i = \begin{cases} \frac{1}{2} , & i \leq 2m \\ 0 & \text{otherwise} \end{cases}$$

This means that the posterior simplifies to

$$pr(f, \alpha | D) \propto \left(\frac{\alpha}{(\alpha + \frac{1}{2})} \right)^{(m+\frac{1}{2})} exp(\alpha S(f) - L(f))$$

Hence the most probable choice of α (around which the posterior will be sharply peaked) for any given f is

$$\alpha = \frac{m}{\sqrt{S^2 - 4mS} - S}$$

This final step completes the iteration scheme, because (whatever polynomials are used), all of f, M, K, S, α are now defined by u. Hence the most probable coefficients $\hat{u}$ and the corresponding most probable spectrum $\hat{f}$ and its multiplier $\hat{\alpha}$ can be found numerically (Tchebyshev polynomials being convenient because of their finite norms and addition properties).

Fluctuations.

Near the most probable spectrum $\hat{f}$, the posterior is

$$pr(f|D) \propto exp\left(-\frac{1}{2} \left(f - \hat{f}\right)^T (-\hat{\alpha}\nabla\nabla S + \nabla\nabla L) \left(f - \hat{f}\right) \right)$$

so that the Bayesian fluctuations obey

$$< \delta f \; \delta f^T >= (-\hat{\alpha}\nabla\nabla S + \nabla\nabla L)^{-1}$$

In terms of those coefficients u which correspond to the polynomials $\hat{Q}$ orthonormal over $\hat{f}$,

$$\partial^2 S / \partial u_i \partial u_j = \begin{cases} -1 & i = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\partial^2 L/\partial u_i \partial u_j = \begin{cases} \frac{1}{2}, & i = j \leq 2m \\ 0 & \text{otherwise} \end{cases}$$

so that

$$< \delta u_i \delta u_j >= \begin{cases} 1/\left(\hat{\alpha} + \frac{1}{2}\right), & i = j \leq 2m \\ 1/\hat{\alpha}, & i = j \geq 2m \\ 0 & \text{otherwise} \end{cases}$$

and

$$\begin{aligned} < \delta f(x) \delta f(y) >= \hat{f}(x) \hat{f}(y) \sum_{i,j} \hat{Q}_i(x) \hat{Q}_j(y) < \delta u_i \delta u_j > \\ = \hat{f}(x) \hat{f}(y) \left\{ \sum_{i=0}^{2m} \hat{Q}_i(x) \hat{Q}_i(y) / \left(\hat{\alpha} + \frac{1}{2}\right) \right. \\ \left. + \sum_{i=2m+1}^{\infty} \hat{Q}_i(x) \hat{Q}_i(y) / \hat{\alpha} \right\} \end{aligned}$$

The fluctuation splits cleanly into a "measured" part ($i \leq 2m$) restricted by both likelihood and entropic prior, and an "unmeasured" part ($i > 2m$) restricted by the prior alone. As a matter of practicality, the infinite sum can be circumvented by using orthonormality in the form

$$\sum_{j=0}^{\infty} \hat{Q}_i(x) \hat{Q}_i(y) = \delta(x - y)/f(x)$$

to give

$$\begin{aligned} < \delta f(x) \delta f(y) > = \hat{f}(x) \ \delta(x - y)/\hat{\alpha} \\ - \hat{f}(x) \hat{f}(y) \sum_{i=0}^{2m} \hat{Q}_i(x) \hat{Q}_i(y) / \hat{\alpha}(2\hat{\alpha} + 1) \end{aligned}$$

The proportional errors on any individual point x of the spectrum are arbitrarily large, because of the delta function, so that it is meaningless to assign pointwise errors to the spectrum itself. Indeed, one should expect this from finite data, because the actual spectrum might (and does!) exhibit large point–to–point variations. However, an integral

$$\Phi = \int dx \ f(x) \phi(x)$$

over the spectrum can be well–determined. Its mean is

$$< \Phi >= \int dx \ \hat{f}(x) \phi(x)$$

and its variance is

$$< (\delta\Phi)^2 > = \iint dxdy\ \phi(x)\,\phi(y)\ < \delta f(x)\,\delta f(y) >$$
$$= \sum_{i=0}^{2m} \phi_i^2 / \left(\hat{\alpha} + \frac{1}{2}\right) + \sum_{i=2m+1}^{\infty} \phi_i^2/\hat{\alpha}$$

where ϕ_i is the coefficient in the polynomial expansion of ϕ,

$$\phi(x) = \sum_{i=0}^{\infty} \phi_i \hat{Q}_i(x)\ ,\ \phi_i = \int dx \hat{f}(x)\,\hat{Q}_i(s)\,\phi_i(x)$$

If the first "measured" part of the variance of ϕ dominates the second "unmeasured" part, the estimate of ϕ is unlikely to be much improved by acquiring more moments. It will be better to drive the noise down by using extra random seed vectors. Conversely, if the variance is dominated by the unmeasured part, it will usually be better to acquire higher moments from the same seed(s).

Again, the infinite sum can be circumvented by Bessel's equality

$$\sum_{i=0}^{\infty} \phi_i^2 = \int dx \hat{f}(x)\,\phi(x)^2$$

Results.

Consider a 1000 × 1000 periodic square lattice of point masses

$$M_{i,j} = (4 \text{ if i+j even , } 2 \text{ if i+j odd})$$

undergoing perpendicular vibrations with equation of motion

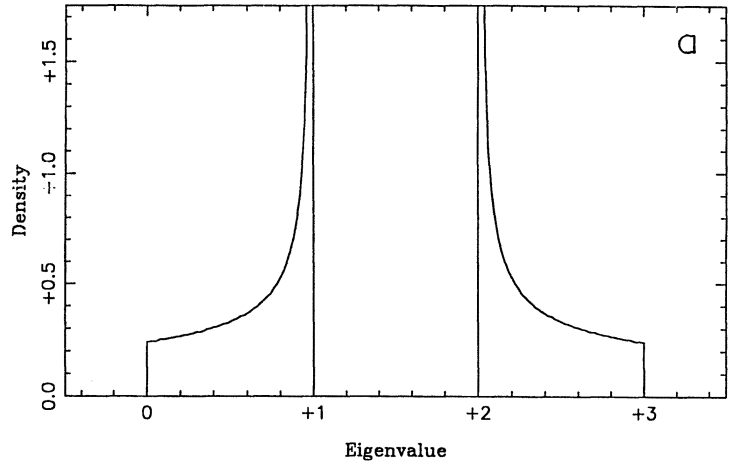
$$M_{i,j}\ d^2y_{i,j}/dt^2 = y_{i,j+1} + y_{i,j-1} + y_{i+1,j} + y_{i-1,j} - 4y_{i,j}$$

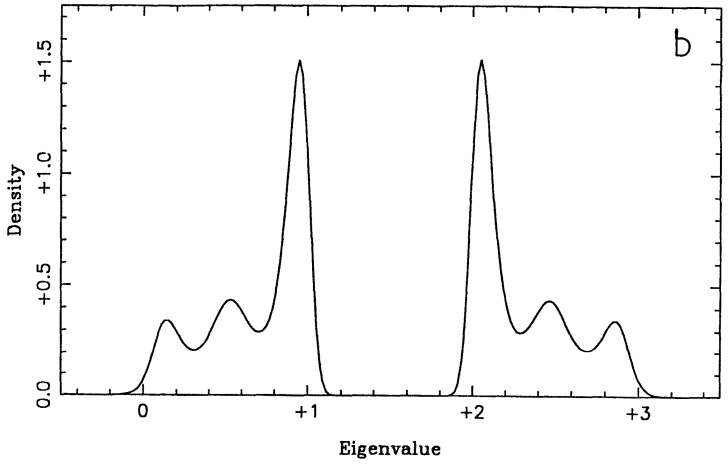
Scaling to $z_{i,j} = y_{i,j}\sqrt{M_{i,j}}$, these equations define the symmetric 1000000 × 1000000 response matrix **A**

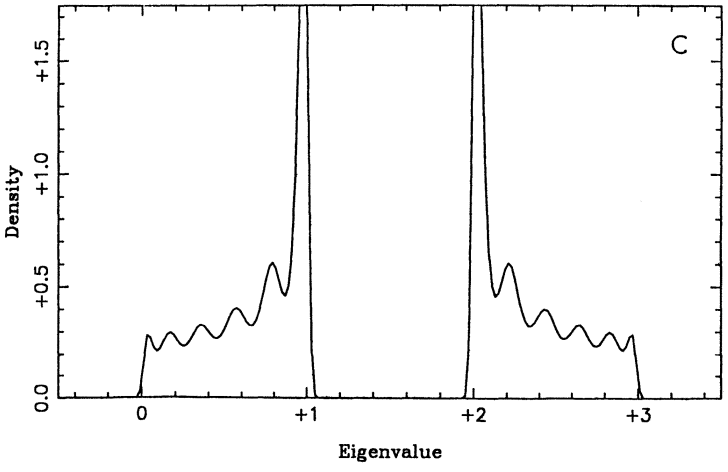
$$-d^2z/dt^2 = \mathbf{A}z$$

whose eigenvalues x are the squared angular frequencies of the normal modes. The true eigenvalues can be found by Fourier analysis (Fig. 1a): they lie in two bands $0 \le x \le 1$ and $2 \le x \le 3$.

Using successively $m = 2, 10,$ and 30 matrix applications, the classic maximum entropy spectra $\hat{f}$, reconstructed on the range (-0.5,3.5), are shown in Figs. 1b,c,d. The two bands are clearly reproduced, the major artefact being "ringing" with amplitude around 20% at a spatial frequency related to the highest measured order


a
Density
0.0
+0.5
+1.0
+1.5
0
+1
+2
+3
Eigenvalue


b
Density
0.0
+0.5
+1.0
+1.5
0
+1
+2
+3
Eigenvalue


c
Density
0.0
+0.5
+1.0
+1.5
0
+1
+2
+3
Eigenvalue

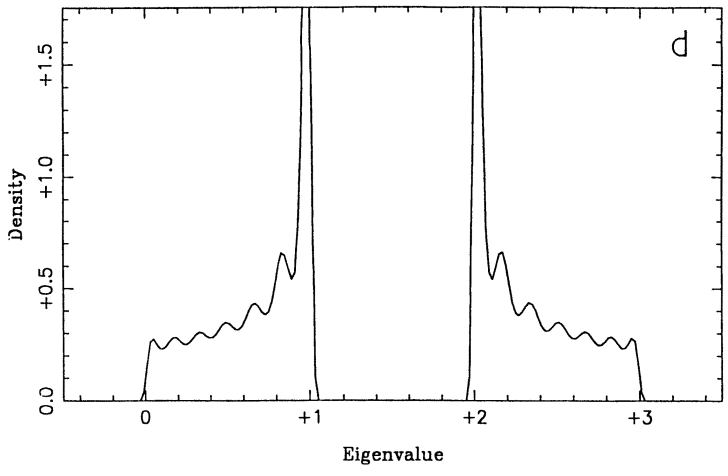


Figure 1. Eigenvalue spectrum (normalised to unit total) of 1000000 × 1000000 matrix describing vibrations on a 1000 × 1000 lattice. (a) True spectrum, (b,c,d) Reconstructed from 2, 10, 30 matrix applications.

of moments. This is a corollary of the attempt to reproduce isolated discontinuities in an otherwise smooth distribution.

Integrals over the spectrum are reproduced to the full accuracy of about 1 in 1000 which one might expect from a technique which uses 1000000 random numbers. Thus the values of

$$log\ det\,(A+I) = \int dx f\,(x)\ log\,(x+1)$$

are shown in the following table (m = number matrix applications) where, if anything, the results are better than the quoted errors.

m	log det (**A**+**I**)
1	844000 ± 11000*
2	849200 ± 1400
3	849100 ± 1300
5	849100 ± 1300
10	849000 ± 1300
20	852600 ± 1300
30	850000 ± 1300

True value 849543

(* : deviation dominated by "unmeasured" components)

As another example, the correct specific heat of the lattice at temperature T is, in appropriate units, (Fig. 2a)

$$C_V(T) = \int_0^\infty dx\, f(x)/T^2 \left(exp\left(x^{\frac{1}{2}}/T\right) - 1\right)^2$$

The reconstruction $\hat{f}$ from just 2 matrix applications yields the band of values ($\pm 1\ \sigma$) shown in Figure 2b, and reconstructions from more applications are visually indistinguishable from the truth.

The following table gives results at two representative temperatures, both for spectra naively reconstructed on the range (-0.5,3.5) and for spectra restricted to the positive range (0,3.5).

	Range = (-0.5,3.5)		Range = (0,3.5)	
m	CV(0.5)	CV(5.0)	CV(0.5)	CV(5.0)
1	85000±13200*	749300±2100*	106800±11200*	790800±1900*
2	102200±5300*	786500±1200	105300±4500*	790800±1200
3	99300±2300*	783400± 1100	108200±3100*	791100±1200
5	106500±900*	789800±1100	107400±900*	790900±1100
10	106000±300	788400±1100	108600±300	791300± 1100
20	108000±200	790700±1100	108400±200	791400± 1100
30	107600±200	790200±1100	108500±200	791600± 1100
Truth	107498	791586	107498	791586

(* : deviation dominated by "unmeasured" components)

Here, some of the quoted deviations are somewhat too small. Any such mistake can presumably be traced to the most important assumption in the analysis — that of a flat entropic model $m(x)$. Indeed, changing the model by zeroing it in $x < 0$ has demonstrably altered some of the answers by more than one or two quoted deviations. It seems likely that future improvements in this method of determining eigenvalues may be found by refining the treatment of this prior eigenvalue model.

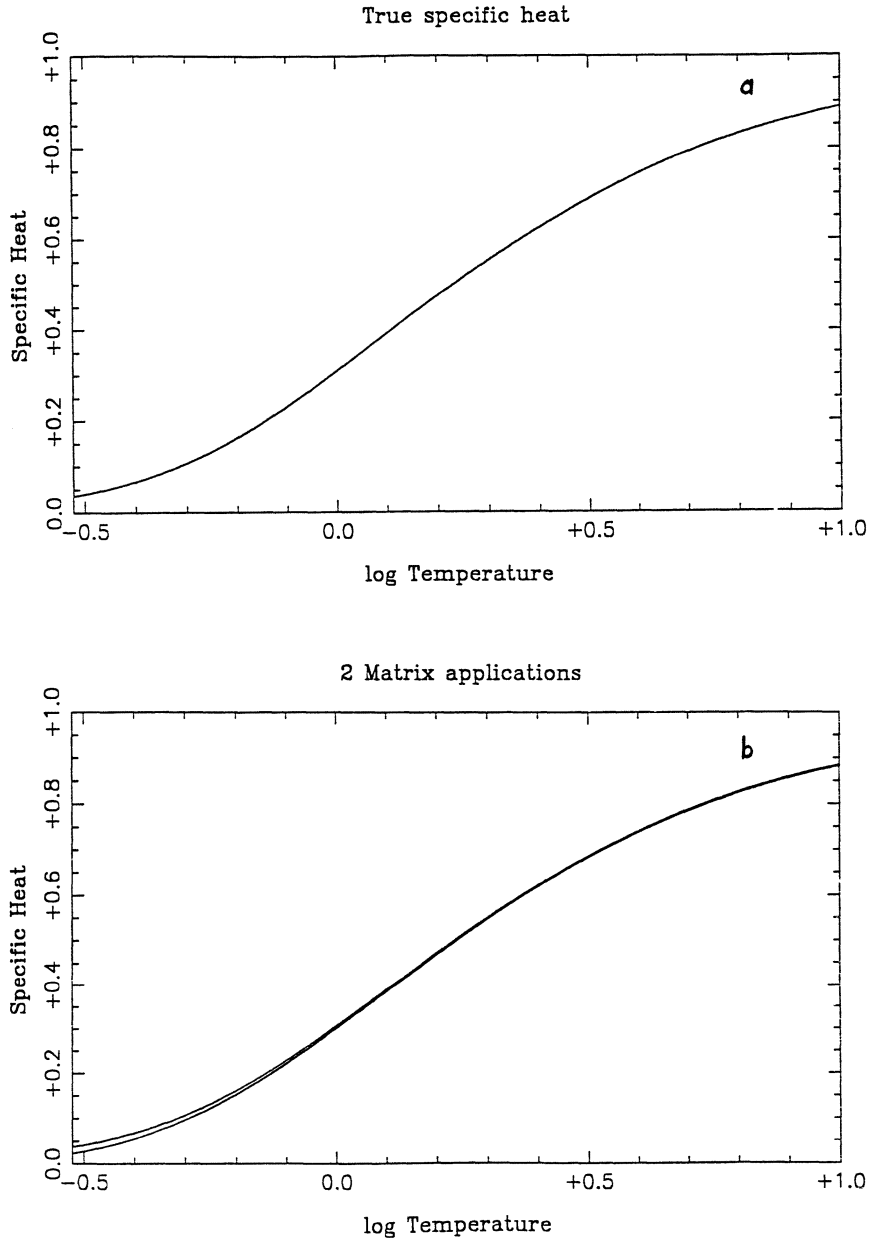


Figure 2. Specific heat per mode of 1000000 × 1000000 matrix describing vibrations on a 1000 × 1000 lattice.

(a) True specific heat of lattice, for temperatures T between 0.3 and 100.

(b) Reconstructed from 2 matrix applications: the two curves are the formal $\pm 1\sigma$ bounds.

Conclusions.

The eigenvalue spectra of symmetric matrices can be obtained by using a Bayesian algorithm with entropic prior. If a matrix is so large that it can be numerically applied to a vector, but nothing more, then this procedure appears to be the most logical way of proceeding. It produces moment data, and as has been noticed by Mead and Papanicolau (1986), good spectra can be obtained from remarkably few moments, so the procedure is efficient.

Integrals over the eigenvalue spectrum can also be obtained, together with error estimates. Although the effects are relatively minor, visible artefacts in the spectra and mistakes in the error estimates point to the possibility of future improvements, presumably in the treatment of the entropic model which underlies the definition of the eigenvalues' entropy.

The million–by–million matrix used as an example was programmed on the author's IBM AT. Although the matrix was very sparse, it took 10 hours CPU to apply it 30 times. By contrast, the twenty iterates which were needed for effective convergence to the classic solution took 10 minutes, and needed less than 1000 words of declared storage.

References.

Gull, S.F. 1988 "Recent advances in maximum entropy", these proceedings.
Mead, L.R. and Papanicolau, N. 1986 "Maximum entropy in the problem of moments" J. Math. Phys. **27** 2903–2907

BAYESIAN EVALUATION OF DISCREPANT EXPERIMENTAL DATA

F. H. Fröhner
Kernforschungszentrum Karlsruhe
Institut für Neutronenphysik und Reaktortechnik
Postfach 3640
D-75 Karlsruhe, West Germany

ABSTRACT. One of the thorniest problems in data evaluation is that of discrepant experimental results, for instance the case of a physical quantity for which two independent measurements yielded values that do not agree "within error bars", indicating unrecognised errors in one or both measurements. The general case of $N \geq 2$ partially discrepant measurements of the same physical quantity is treated by means of a two-stage Bayesian approach, with normally distributed recognised and unrecognised errors as the statistical model. A noninformative prior for the spread of the unrecognised errors is not sufficient for definite predictions under quadratic loss, but already the simplest conjugate maximum-entropy prior yields best estimates for the physical quantity and its uncertainty in a straightforward way. From the corresponding joint posterior of the unrecognised errors one obtains, in saddle point approximation, estimates that resemble James-Stein estimators. Unlike those, however, the Bayesian estimates of unrecognised errors have easily calculated uncertainties and correlations associated with them, and they are free of the pathological discontinuities of many "improved" (e. g. "plus-rule") James-Stein estimators. A practical example from the evaluation of nuclear data is given.

1. INTRODUCTION

Data evaluation in the modern sense began in the early 'fifties with the effort of DuMond and Cohen to determine a recommended set of fundamental physical constants (light velocity, Planck's constant, electron charge etc.), and to establish their uncertainties, from all relevant experimental data [1]. At about the same time the rapidly growing nuclear technology began to develop a voracious appetite for accurate nuclear data, especially for cross sections (i.e. probabilities) of neutron-induced nuclear reactions (scattering, fission, radiative capture etc.), but also for nuclear structure and decay data (energies, spins and half-lives of compound-nuclear states, transition strengths etc.), and nuclear data evaluation has become a veritable industry. Modern evaluated neutron data files contain millions of cross section values covering the whole energy range from 10 μeV to 20 MeV for hundreds of isotopes, and computers are indispensable for their maintenance and utilization [2]. Similar files of evaluated data have been established in elementary particle physics, materials research, aerospace technology and other fields.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 467–474.

2. EVALUATION OF INCONSISTENT DATA

One of the thorniest problems in data evaluation is that of inconsistent data. Suppose we are given the results of n completely independent and experimentally different measurements of the same physical quantity, μ, in the form $x_i \pm \sigma_i$, $i = 1, 2, \ldots n$. If the separation of any two values, $|x_i - x_j|$, is smaller or at least not much larger than the sum of their uncertainties, $\sigma_i + \sigma_j$, the data are said to be consistent or to agree "within error bars". (The probability is only erfc 1 = 15.7 % that two equally precise experiments yield a separation larger than $\sigma_i + \sigma_j = 2\sigma$, for Gaussian sampling distributions with standard deviation σ). If some or all separations are much larger, the data are not consistent with the stated uncertainties. Inconsistencies are caused by unrecognized or inadequately corrected experimental effects such as backgrounds, dead time of the counting electronics, instrumental resolution, sample impurities, calibration errors, etc.

What can we say about unrecognized errors? If we have no other information but the data, and know nothing about the experiments that yielded them, positive and negative errors are equally probable, hence the probability distribution for the unrecognized error ε_i of the i-th experiment should be symmetric about zero, and the same distribution should apply to all experiments. Let us therefore assume, in the spirit of the maximum entropy principle, Gaussian distributions for all ε_i,

$$p(\varepsilon_i|\tau_i)d\varepsilon_i = \frac{1}{\sqrt{2\pi\tau_i^2}} \exp\left[-\frac{\varepsilon_i^2}{2\tau_i^2}\right] d\varepsilon_i \quad , \quad -\infty < \varepsilon_i < \infty \; . \tag{1}$$

The probability to measure the value x_i, given the true value μ, the unrecognized error ε_i and the uncertainty σ_i due to all recognized error sources, is then given by

$$p(x_i|\mu,\varepsilon_i,\sigma_i)dx_i = \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left[-\frac{(x_i-\mu-\varepsilon_i)^2}{2\sigma_i^2}\right]dx_i \quad , \quad -\infty < x_i < \infty \; . \tag{2}$$

If the dispersions τ_i of unrecognized errors are known, the joint posterior distribution for μ and the ε_i is

$$p(\mu,\varepsilon|x,\sigma,\tau)d\mu d^n\varepsilon \propto d\mu \prod_{i=1}^{n} d\varepsilon_i \exp\left[-\frac{(x_i-\mu-\varepsilon_i)^2}{2\sigma_i^2} - \frac{\varepsilon_i^2}{2\tau_i^2}\right] . \tag{3}$$

(ε, σ, τ are to be understood as vectors with coordinates ε_i, σ_i, τ_i). Completing squares in the exponent we can easily integrate over the ε_i. The resulting posterior distribution for μ is a Gaussian,

$$p(\mu|x,\sigma,\tau)d\mu = \frac{1}{\sqrt{2\pi v}} \exp\left[-\frac{(\mu-\bar{x})^2}{2v}\right] d\mu \quad , \quad -\infty < \mu < \infty \; , \tag{4}$$

so we recommend (under quadratic loss) for μ and its squared uncertainty

$$\langle\mu\rangle = \bar{x} \quad , \tag{5}$$

$$v = \frac{1}{n}\,\overline{(\sigma^2 + \tau^2)} \quad , \tag{6}$$

where the overbar denotes an average over i (i.e. over experiments) with weights $1/(\sigma_i^2+\tau_i^2)$. If we integrate (3) over μ we find the joint distribution of the ε_i,

$$p(\varepsilon|x,\sigma,\tau)d^n\varepsilon \propto \exp\left[-\frac{1}{2}(\varepsilon-x)^+A^{-1}(\varepsilon-x) - \frac{1}{2}\varepsilon^+B^{-1}\varepsilon\right] d^n\varepsilon \quad , \tag{7}$$

where A^{-1} and B^{-1} are positive definite, symmetric matrices defined by

$$(A^{-1})_{ij} \equiv \sigma_i^{-2}\,\delta_{ij} - \frac{\sigma_i^{-2}\,\sigma_j^{-2}}{\sum_k \sigma_k^{-2}} \quad , \tag{8}$$

$$(B^{-1})_{ij} \equiv \tau_i^{-2}\,\delta_{ij} \quad . \tag{9}$$

This product of two multivariate Gaussians is a multivariate Gaussian again, with mean vector $\langle\varepsilon\rangle = CA^{-1}x$ and covariance matrix C, where $C^{-1} = A^{-1}+B^{-1}$, so that $(A^{-1}+B^{-1})\langle\varepsilon\rangle = A^{-1}x$. Solving the last equation for $\langle\varepsilon_i\rangle$ one gets

$$\langle\varepsilon_i\rangle = \frac{\tau_i^2}{\sigma_i^2+\tau_i^2}(x_i - \bar{x}) \quad . \tag{10}$$

Thus the best estimate of ε_i is the deviation $x_i-\bar{x}$ of the i-th datum from the (weighted) mean, multiplied by a "shrinking factor" $\tau_i^2/(\sigma_i^2+\tau_i^2)$ which is close to zero if the expected unrecognised error is much smaller, and close to unity if it is much larger, than the known uncertainty σ_i. Of course, this is the trivial case: If we know the variances τ_i^2 of the unrecognised errors we know as much about them as about the other errors. We can thus simply add variances to get the total mean square errors $\sigma_i^2+\tau_i^2$ whose reciprocals appear then as weights in all i-averages as we have just found.

The simplest nontrivial case is obtained if we consider the τ_i as our best estimates of the (root-mean-square) unrecognised errors, based on the quality of the various measurements, on the accuracy of the techniques employed, perhaps even on the credibility of the experimentalists as judged from their past record. (Note that it is perfectly alright to put $\tau_i = 0$ for those experiments which can be considered as unaffected by unrecognised errors). The unknown true variances may then be taken as τ_i^2/c where c is an adjustable common scale parameter with prior p(c)dc, and the joint probability for μ and the vector ε as

$$p(\mu,\varepsilon|x,\sigma,\tau)d\mu d^n\varepsilon \propto d\mu d^n\varepsilon \int_0^\infty dc\,p(c) \prod_{i=1}^n c^{1/2} \exp\left[-\frac{(x_i-\mu-\varepsilon_i)^2}{2\sigma_i^2} - \frac{c\varepsilon_i^2}{2\tau_i^2}\right]. \tag{11}$$

Integrating over all ε_i one gets the posterior distribution of μ,

$$p(\mu|x,\sigma,\tau)d\mu \propto d\mu \int_0^\infty dc\,p(c) \prod_{i=1}^n (\sigma_i^2 + \tau_i^2/c)^{-1/2} \exp\left[-\frac{1}{2}\frac{(\mu-x_i)^2}{\sigma_i^2+\tau_i^2/c}\right]. \tag{12}$$

If we have no numerical information at all about the scale parameter Jeffreys' prior dc/c appears appropriate [3,4]. The integration over

c is easy if the known uncertainties are unimportant. With $\sigma_i = 0$ for all i the integrand becomes essentially a gamma distribution in c, integration over which yields Student's t-distribution

$$p(\mu|x,0,\tau)d\mu \propto \frac{du}{(1+u^2)^{n/2}} \quad , \quad -\infty < u \equiv \frac{\mu-\bar{x}}{s'} < \infty \quad , \tag{13}$$

with

$$\langle\mu\rangle = \bar{x} \, , \tag{14}$$

$$\text{var } \mu = \frac{s'^2}{n-3} \quad , \tag{15}$$

where $s'^2 \equiv \overline{x^2}-\bar{x}^2$, $\bar{x}$ and $\overline{x^2}$ are averages weighted by $1/\tau_i^2$. This is the same result as if the data from the various experiments were a sample from a Gaussian, affected by uncertainties τ_i.

In the general case, $\sigma_i > 0$, $\tau_i > 0$, the integral (11) with the noninformative prior, $p(c)dc \propto dc/c$, diverges logarithmically because the integrand becomes proportional to 1/c for $c \to \infty$. The Bayesian formalism signals in this way that the prior information is insufficient for definite predictions. Is there anything we know in addition to the fact that c is a scale parameter? Actually, if the τ_i are our best estimates of the uncertainties caused by unrecognised errors, we expect c to be close to unity. The maximum-entropy prior constrained by $\langle c \rangle = 1$ is [4]

$$p(c)dc = e^{-c} \, dc \quad , \qquad 0 < c < \infty \quad . \tag{16}$$

This is almost as noncommittal as Jeffreys' prior, decreasing also monotonically as c increases, but normalizable and giving less weight to the extrema. With this prior both the c-integral and the normalization constant of the posterior μ-distribution (12) are finite and can be calculated numerically without difficulty. Fig. 1 shows a real-life example, Gaussian distributions representing the results of six measurements of the ^{239}Pu fission cross section for 14.7 MeV neutrons listed in Table I, together with the estimated posterior distribution. Prior uncertainties of $\tau_i = 0.1$ b were assigned to all experiments indiscriminately, based on the state of the art. The posterior mean and and the root-mean-square uncertainty computed numerically are also given in Table I.

3. ESTIMATION OF UNKNOWN SYSTEMATIC ERRORS

What can we learn about the unrecognised systematic errors ε_i from the set of inconsistent data, $x_i \pm \sigma_i$, i = 1,...n? With the prior (16) it is easy to integrate the posterior probability distribution (11) first over the gamma distribution of c, then over the Gaussian distribution of μ. The result can be written in the form

$$p(\varepsilon|x,\sigma,\tau)d^n\varepsilon \propto \exp\left[-\frac{1}{2}(\varepsilon-x)^+A^{-1}(\varepsilon-x)\right]\left[1+\frac{1}{2}\varepsilon^+B^{-1}\varepsilon\right]^{-n/2-1}d^n\varepsilon \tag{17}$$

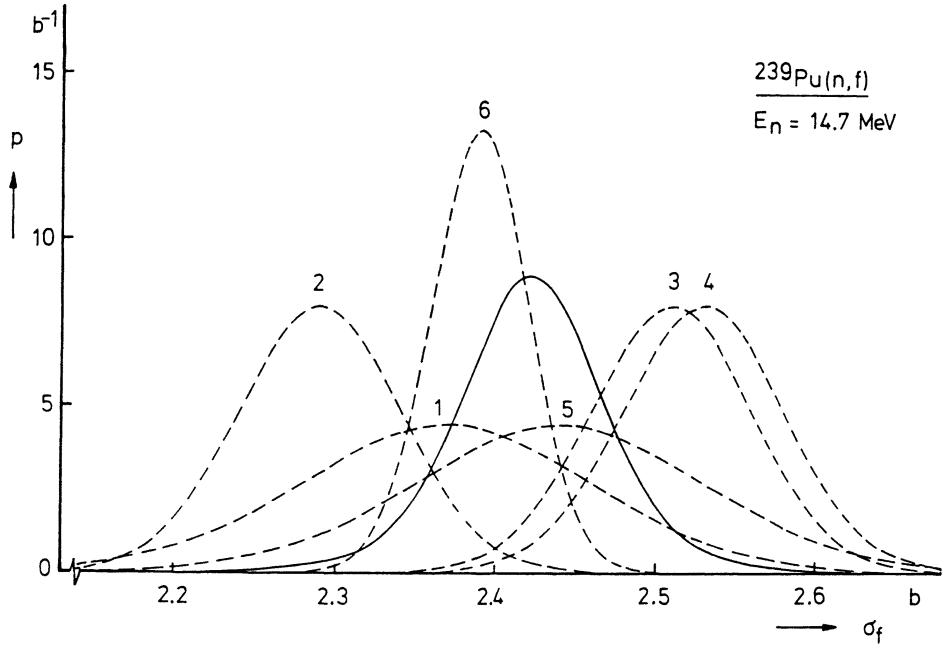


Figure 1. Probability densities representing the experimental results in Table I (dashed Gaussians), and posterior density of the true value (solid curve) estimated with the two-stage Bayesian model, Eq. 11, with hyperprior Eq. 16. Inconsistencies are evident between the experiments 2, 3 and 4 (curve labels correspond to the first column of the table).

TABLE I Experimental input and estimation results for the ^{239}Pu(n,f) reaction at 14.7 MeV incident neutron energy

i	Authors	Year	Ref.	measured fission cross section (barn)	estimated unrecognized error (barn)
1	Kari	1978	[6]	2.37 ± .09	-.019 ± .056
2	Cancé et al.	1978	[7]	2.29 ± .05	-.086 ± .050
3	Adamov et al.	1979	[8]	2.51 ± .05	.056 ± .048
4	Li et al.	1982	[9]	2.53 ± .05	.069 ± .049
5	Mahdawi et al.	1982	[10]	2.44 ± .09	.006 ± .056
6	Arlt et al.	1983	[11]	2.39 ± .03	-.027 ± .041
		best estimate:		2.42 ± .05	
		assumed:		τ_i = 0.1 b (for all i)	

where the matrices A and B are defined as before (in Eqs. 8 and 9). In order to get the mean vector $\langle\varepsilon\rangle$ and the covariance matrix $C \equiv \langle\delta\varepsilon\delta\varepsilon^+\rangle$ in analytic form we employ the saddle point approximation, i. e. we replace the ε-distribution by a multivariate Gaussian with the same maximum and the same curvature at the maximum,

$$p(\varepsilon|x,\sigma,\tau) \equiv e^{-F(\varepsilon)} \simeq \exp\left[-F(\hat{\varepsilon}) - \frac{1}{2}(\varepsilon-\hat{\varepsilon})^+(\nabla\nabla^+F)_{\varepsilon=\hat{\varepsilon}}(\varepsilon-\hat{\varepsilon})\right] \tag{18}$$

where the vector operator ∇ has coordinates $\nabla_i \equiv \partial/\partial\varepsilon_i$. This shows that

$$\langle\varepsilon\rangle \simeq \hat{\varepsilon}\ , \tag{19}$$

$$\left\langle\delta\varepsilon\delta\varepsilon^+\right\rangle = \left[(\nabla\nabla^+F)_{\varepsilon=\hat{\varepsilon}}\right]^{-1}\ . \tag{20}$$

The most probable vector $\hat{\varepsilon}$ must be found as the solution of

$$\nabla F = A^{-1}(\varepsilon-x) + \frac{n+2}{2}\frac{B^{-1}\varepsilon}{1+\varepsilon^+B^{-1}\varepsilon/2} = 0 \tag{21}$$

and the (approximate) covariance matrix $\langle\delta\varepsilon\delta\varepsilon^+\rangle$ as the inverse of

$$\nabla\nabla^+F = A^{-1} + \frac{n+2}{2}\frac{B^{-1}(1+\varepsilon^+B^{-1}\varepsilon/2) - B^{-1}\varepsilon\varepsilon^+B^{-1}}{(1+\varepsilon^+B^{-1}\varepsilon/2)^2} \tag{22}$$

evaluated at $\varepsilon = \hat{\varepsilon}$. With the definitions of A and B, Eqs. 8 and 9, we get from (21)

$$\hat{\varepsilon}_i = x_i - \bar{x} + \bar{\hat{\varepsilon}} - \frac{n+2}{2}\frac{\tau_i^{-2}\sigma_i^2\hat{\varepsilon}_i}{1+\sum_j \tau_j^{-2}\hat{\varepsilon}_j^2/2}\ , \tag{23}$$

where $\bar{x}$ and $\bar{\hat{\varepsilon}}$ are $1/\sigma_i^2$-weighted averages. This is suitable for iteration. With $\hat{\varepsilon}_i \simeq x_i - \bar{x}$ for all i as first approximation one finds the second approximation

$$\hat{\varepsilon}_i \simeq \left[1 - \frac{n+2}{2}\frac{\tau_i^{-2}\sigma_i^2}{1+\sum_j \tau_j^{-2}(x_j-\bar{x})^2}\right](x_i-\bar{x})\ , \tag{24}$$

and then the higher ones. This treatment of systematic errors is an example of the "hierarchical" (here: two-stage) Bayesian method which involves repeated application of Bayes' theorem: The sampling distribution (2) depends on parameters ε_i with the prior (1) which, in turn, depends on the "hyperparameter" c with the "hyperprior" (16).

The final estimates and their uncertainties (square roots of the diagonal elements of the matrix C) for our ^{239}Pu problem, obtained in this way, are given in the last column of Table I. No significant unrecognised errors are found for the measurements 1, 5 and 6, whereas 2, 4 and perhaps 3 seem affected by unrecognized errors which are of the same order of magnitude as the uncertainties stated by the authors. Of course these conclusions could have been drawn already from the experimental data (and especially from the pictorial representation by Gaussians in Fig. 1), but the formalism provides quantitative support for our common sense also under less obvious circumstances.

4. COMPARISON WITH JAMES-STEIN ESTIMATORS

Our second approximation, Eq. 24, resembles the James-Stein estimators [12] which, since their introduction, have caused a great deal of excitement, confusion and a flood of papers. C. Stein showed, using the frequentist definition of risk (square error averaged over all possible samples, for a given set of parameters) that estimators similar to (24) have lower risk than the estimates resulting from Bayesian estimation under quadratic loss (which minimise the square error averaged over all possible parameters, for the sample at hand). Many improvements have been suggested to Stein's original estimators, based on distribution theory and educated guesswork. For instance, the "plus rule" estimator

$$\varepsilon_i^* \simeq \left[1 - \frac{n-2}{n}\frac{\sigma^2}{s'^2}\right]_+ (x_i - \bar{x}) \quad , \tag{25}$$

has been proposed for a situation which corresponds to data with equal uncertainties ($\sigma_i = \sigma$ for all i). The subscript + means that only positive values of the shrinking factor are accepted, for negative values one puts $\varepsilon_i^* = 0$. Moreover, the estimator (25) is restricted to $n \geq 3$. Wild discussions arose about the "paradox" that the estimator for ε_i depends on all the other, independently sampled x_j, $j \neq i$. The "speed of light" question was asked whether inclusion of other unrelated data would not improve the estimate. ("Do you mean that if I want to estimate tea consumption in Taiwan I will do better to estimate simultaneously the speed of light and the weight of hogs in Montana?" [13]) So-called parametric empirical Bayes recipes seemed to offer some insight, e. g. replacement of τ^2 in Eq. 10 (in the case $\tau_i = \tau$ for all i) by the sample variance s'^2. However plausible such recipes may be (see e.g. [14]), without rigorous justification they remain adhockeries.

Under the same circumstances ($\sigma_i = \sigma$, $\tau_i = \tau$) Eq. 24 yields

$$\langle \varepsilon_i \rangle \simeq \left[1 - \frac{(n+2)\sigma^2}{ns'^2 + 2\tau^2}\right] (x_i - \bar{x}) \quad , \tag{26}$$

valid for all $n \geq 2$, without discontinuities or interpretational problems. A paradox exists only for frequentists who deny themselves the use of priors. For a Bayesian the fact that he considers data from a set of similar experiments, formally related by the hyperprior p(c)dc, induces correlations and shrinking factors in a perfectly natural way. From the Bayesian viewpoint, on the other hand, it looks odd to base the risk criterion not on the one available sample and on prior knowledge, but on the available sample and all other unobserved samples that can be imagined. Frequentist claims about the superiority of estimators "in the long run" (averaged over many samples) are not very relevant in data evaluation work, where one must infer best values (for quadratic or any other loss) from one available sample. It may be true that an estimator with low risk is closer to the true value for a larger fraction of all possible samples than an estimator which ensures minimal quadratic loss, but for the remaining fraction of samples the errors tend to be so much worse that the apparent advantage turns into a net disadvantage [15]. In

any case, the Bayesian two-stage method yields, in second saddle point approximation, estimators which are similar to, but especially for small samples better than, James-Stein estimators. Moreover, by iteration one finds all possible improvements and also the uncertainties in a systematic and unambiguous way, without the bizarre discontinuities of many improved James-Stein estimators. The Bayesian approach leaves no room for guesswork once statistical model, priors and loss function are fixed.

ACKNOWLEDGEMENT

It is a pleasure to thank E.T. Jaynes and F.G. Perey for valuable information and discussions, and G. Kessler, H. Küsters and the Fast Breeder Project (PSB) at Karlsruhe for support of this work.

REFERENCES

[1] J.W.M. DuMond and E.R. Cohen, Rev. Mod. Phys. 25 (1953) 691; E.R. Cohen, K.M. Crowe, J.W.M. DuMond, "Fundamental Constants of Physics", Interscience, New York - London (1957); E.R. Cohen and B.N. Taylor, Physics Today, August 1987, BG 11

[2] F.H. Fröhner, Karlsruhe Report KfK 4099 (1986)

[3] H. Jeffreys, "Theory of Probability", Oxford (1939)

[4] E.T. Jaynes, "Prior Probabilities", IEEE Trans. Syst. Cybern. SSC-4 (1968) 227; reprinted in [5] p. 114

[5] E.T. Jaynes, "Papers on Probability, Statistics and Statistical Physics", R.D. Rosenkrantz, ed., Reidel, Dordrecht (1983)

[6] K. Kari, Karlsruhe report KfK 2673 (1978)

[7] M. Cance and G. Grenier, Nucl. Sci. Eng. 68 (1978) 197

[8] V.M. Adamov et al., Proc Conf. on Nucl. Cross Sections and Technol., Knoxville, NBS SP 594, p. 995 (1979)

[9] Li Jingwen et al., Proc. Conf. on Nucl. Data for Sci. and Technol., Antwerp 1982, K.H. Böckhoff, ed., Reidel, Dordrecht (1982), p. 55

[10] M. Mahdawi and G.F. Knoll, Proc. Conf. on Nucl. Data for Sci. and Technol., Antwerp 1982, K.H. Böckhoff, ed., Reidel, Dordrecht (1982), p. 58

[11] R. Arlt et al., 6-th All Union Conf. on Neutron Physics, Kiev, vol. 2, p. 129 (1983)

[12] C. Stein, Proc. Third Berkeley Symp. on Math. Statistics and Probab., U. of California Press, vol. I, p. 197 (1956); W. James and C. Stein, Proc. Fourth Berkeley Symp. on Math. Statistics and Probab., U. of California Press, vol. I, p. 361 (1961)

[13] B. Efron and C. Morris, J. Royal Statist. Soc., Series B, 35 (1973) 379

[14] J.O. Berger, "Statistical Decision Theory and Bayesian Analysis", Springer, New York etc. (1985)

[15] E.T. Jaynes, in "Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science", W.L. Harper and C.A. Hooker, eds., Reidel, Dordrecht (1976); reprint in [5] p. 149

Superresolution limit for signal recovery

E.L.Kosarev

Institute for Physical Problems
USSR Academy of Sciences
Moscow 117334

Abstract

It is shown there is an absolute limit for resolution enchancement in comparison with the Rayleigh's classic diffraction limit. The maximum value of superresolution which can be obtained in principle is determined by noise and may be computed via Shannon's theorem concerning the maximum information transmission speed through the connecting channel having noise. A restoration algorithm based on maximum likelihood method which has the Shannon's supremum superresolution is documented. The numerical tests of this algorithm is presented and the result of its application to nuclear magnetic resonance experiment is shown. Superresolution depends logarithmically on the signal/noise ratio.

1 Introduction

The problem of signal recovery (or inverse problem) from noisy data has a long history and there are a lot of papers and books covering this subject [1-8] from which should be noted the splendid review of B.R.Frieden [4]. In the simplest case the problem is reduced to the first kind integral equation

$$\int_a^b K(x,y)f_0(y)dy = F_0(x), \qquad c \le x \le d, \tag{1}$$

which should be solved for the right hand side $F_0(x)$ is known only together with the noise

$$F(x) = F_0(x) + N(x). \tag{2}$$

The statistical characteristics of the noise $N(x)$ (distribution function) is considered to be known. The goal of the inverse problem is to recover with maximum accuracy the unknown function $f_0(y)$ from the measurements of experimental data $F(x)$. The function $K(x,y)$ is the kernel of the integral equation (1). It depends on the physical problem from which we have equation (1) and in this paper we shall call the function $K(x,y)$ the apparatus function or point spread function (PSF).

For the case when the PSF is space invariant $K(x,y) = K(x-y)$ and the function $K(x-y)$ has a bounded Fourier spectrum support the general limitation on possibilities to

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 475–480.

distinguish the closely spaced signals $f_0(y)$ from the noisy data $F(x)$ follows from the well-known Shannon's theorem [9] concerning the maximum speed of information transmission through a noisy channel.

In fact the assumption that the function $K(x-y)$ has a finite spectrum support is not necessary for Shannon's theorem to be correct. According to A.N.Kolmogorov [10] it is sufficient for the dimension of signal space $F(x)$ to be finite. This is the case for almost all experimental data $F(x)$, because the function $F(x)$ is always measured in finite number of points.

We introduce here such a definition for the superresolution coefficient

$$\mathbf{SR} = \Delta/\delta, \tag{3}$$

where Δ is the range of apparatus function $K(x-y)$

$$\Delta = \int_{-\infty}^{\infty} K^2(x)dx, \tag{4}$$

providing $K(0) = 1$, and δ is the minimal distance between signals $f_0(y)$ which we can recognize to be different. This definition is very natural and it was introduced at first by Rayleigh [11] as a measure of resolving power for optical devices and in these cases Δ was defined as the diffraction limit. We intend to show in this paper that the modern methods for signal restoration has much more resolving power in comparison with Rayleigh's classic diffraction limit and that the ultimate limit is determined by noise rather than diffraction.

2 Shannon's limit for superresolution enchancement

The standard form of Shannon's formula for apparatus functions $K(x-y)$ having a bounded Fourier spectrum support can be written in the form

$$B/X = W log_2(1 + P_s/P_n), \tag{5}$$

where $X = d - c$, B/X is the maximum number of information bits per unit measure of x which can be obtained in principle, W is the Fourier spectrum width of apparatus function $K(x-y)$,

$$P_s = \int_{-\infty}^{\infty} F_0^2(x)dx$$

is the energy of signal $F_0(x)$,

$$P_n = n\sigma^2$$

is the energy of noise, and σ^2 equals to the standard deviation of noise in each point x_i

$$\overline{N(x_i)N(x_j)} = \sigma^2\delta_{ij}, \quad i,j = 1,2,...,n$$

of experimental data. Of course formula (5) should be changed for apparatus functions having unbounded spectrum support because for these functions the quantity W has no sense.

From considerations based on dimensional analysis instead of the factor $B/(XW)$ we can use the expression Δ/δ which has the same dimensionality and instead of standard formula (5) we obtain a new one

$$\Delta/\delta = \mathbf{Const} \cdot \log_2(1 + \mathbf{P_s}/\mathbf{P_n}). \tag{6}$$

The numerical value of **Const** can not be recognized from dimensional analysis alone. We can determine it in two ways.

In the first way we can compute the range (4) for a PSF having a bounded spectrum support and in the second way we find the value of **Const** from numerical tests for different PSF's including the unbounded spectrum support functions such a Gaussian and a Lorentzian PSF's. Both ways give the same result: the maximum superresolution practical does not depend on the shape of the apparatus function and the most important conclusion is that the superresolution has a logarithmic dependence on the signal/noise ratio (SNR). The numerical value of $\mathbf{Const} = \mathbf{1/3}$. The full description of all details of this conclusion will be published by the author elsewhere. In next section we describe the algorithm for signal recovery, which has the Shanon's supremum superresolution, results of its numerical tests and application to nuclear magnetic resonance (NMR) experiment.

3 Restoration algorithm and its application

For restoration we use the maximum likelihood (ML) method [12] which is the generalization of M.Z.Tarasko's iteration algorithm [13]. For the case where the right hand side $F(x_i)$ has binomial (or Poisson) distribution in each separate point x_i and jointly polynomial distribution for the whole set of $\{F(x_i)\}$ values for $i = 1, 2, ..., n$. the iteration formula can be written in the form

$$g_k^{(s+1)} = g_k^{(s)} + h g_k^{(s)} \sum_{i=1}^{n} p_{ik} \left(\frac{f_i}{\sum_{j=1}^{m} p_{ij} g_j^{(s)}} - 1 \right) \quad . \tag{7}$$

In this formula $s = 1, 2, ...$ is the iteration number, the vector g_k, $k = 1, 2, ..., m$ equal to the values of the unknown function $f_0(y_i)$ in points $y_1, y_2, ..., y_m$, the vector f_i, $i = 1, 2, ..., n$ equal to right hand side function $F(x_i)$ in points $x_1, x_2, ..., x_n$ and matrix p_{ij} is equal to the values of PSF: $p_{ij} = K(x_i, y_j)$, h is the length step in the space of unknown vector $\{g_k\}$ in the direction which is close to the gradient direction. At $h = 1$ the iteration procedure (7) coincides with M.Z.Tarasko's one. For more detailed specification of the restoration algorithm we refer readers to Ref.[12]. The actual program implementation will be described in a forthcoming publication.

All computation results which are presented in this paper were obtained on HP-1000 minicomputer at the Institute for Physical Problems with the accuracy of 39 bits per mantissa ($\sim 1.8 \times 10^{-12}$). For number of data points $n = 512$ every 50 iterations took about 5 min CPU time. The total number of iterations depends first of all on how close the value of δ is to the resolution limit and sometimes it takes up to 5–10 thousands.

In numerical tests we studied the dependence of the resolution limit for two and three line signals as a function of signal/noise ratio for different PSF's. It was chosen three different PSF's:

$$K(x, y) = k_i(s), \quad s = (x - y)/D, \quad i = 1, 2, 3$$

having a Gaussian shape $k_1(s) = exp(-s^2)$, a Lorentzian one $k_2(s) = 1/(1+s^2)$ and a form $k_3(s) = (\frac{sin\ s}{s})^2$. Every of these functions we convolved with 2 or 3 Gaussian shape narrow lines

$$f_0(y) = exp(-u^2) + exp(-v^2), \qquad \delta = y_2 - y_1,$$

where $u = (y - y_1)/D_1$, $v = (y - y_2)/D_1$, $D_1 = D/40$ and convolution integral (1) is computed by fast Fourier transform. The parameter D is chosen to be equal to such a value so that the PSF decreases to zero inside the interval (a, b). The noise $N(x)$ is added to the computed values of $F_0(x)$ and this values are concidered as initial data $F(x)$ for numerical tests.

The example of two line restoration is shown in Fig.1 for Lorentzian PSF with $D = 60$, SNR=20 dB and distances between the two lines $\delta = 35$ and $\delta = 28$. In the first case this distance is larger than the Shannon's resolution limit and in the second one it is less. We can see on this figure that for $\delta = 35$ the two lines are well resolved and for $\delta = 28$ they are not.

Summary of such tests is presented in Fig.2 for all three PSF types mentioned above and SNR between 10 and 50 dB. In this figure is also shown Shannon's resolution limit

$$\mathbf{SR} = \frac{1}{3}\mathbf{log_2(1 + P_s/P_n)}, \tag{8}$$

the regression line for numerical experiment data and the confidence interval for this line. There is a good agreement between them.

This result is quite a new one and it differs from the factorial dependence of superresolution as a function of SNR which was stated in earlier papers [14-16]. This is because we do not use any parametrization for restoration. If we could have some information concerning the unknown signal $f_0(y)$ to be restored the superresolution limit (8) can be exceeded.

The application of the ML algorithm (7) to NMR experiment [17] is presented in Fig.3. All restored lines which are quite distinguished in this figure are fully interpreted in [17]. This example demonstrates not only the efficiency of restoration method (7) and also reveals information in the raw experimental data which would be fully unrecognized and lost without the use the modern restoration methods.

4 Acknowledgements

Author would like to thank Professors L.A.Vainstein and K.Sh.Zigangirov for interest in this work and critical comments; his colleaques E.R.Podolyak and V.I.Gelfgat for helping in program implementation of restoration algorithm; the officials of SERC Daresbury Laboratory who facilitated participation in the MEM Workshop; P.A.Ridley and E.Pantos for interesting discussions and help in the preparation of the manuscript.

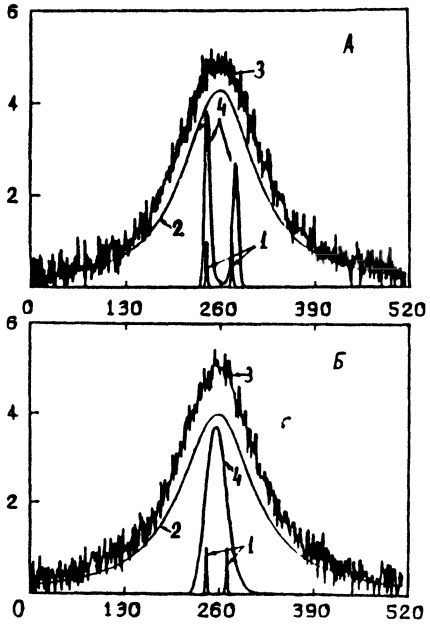


Fig.1 Restoration of 2 lines convolved with Lorentzian shape point spread function and signal to noise ratio 20 dB. A - space between each of lines $\delta = 35$ is greater than resolution limit (8); B - space $\delta = 28$ is less than resolution limit; 1 - the original lines; 2 - point spread function; 3 - initial data for restoration algorithm; 4 - the restoration results.

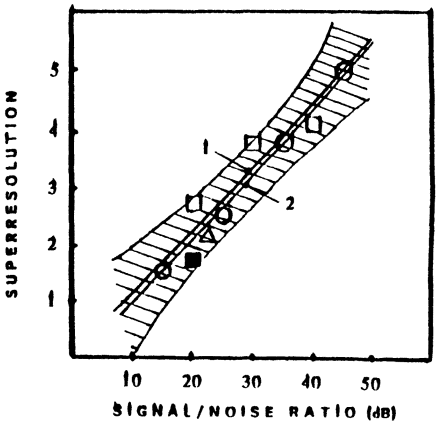


Fig.2 Summary of numerical tests for 2 and 3 lines restoration.
Circles - results for Gaussian PSF, 2 lines; white squares - the Lorentzian PSF, 2 lines; black square - the Lorentzian PSF, 3 lines; triangles - PSF in the form of $(\frac{\sin x}{x})^2$, 2 lines. 1 - the Shannon's superresolution limit (8); 2 - the regression line for numerical tests, which coincides with Shannon's limit within the corridor of errors. The dashed strip is the 95% confidence interval between domains above and under the superresolution limit.

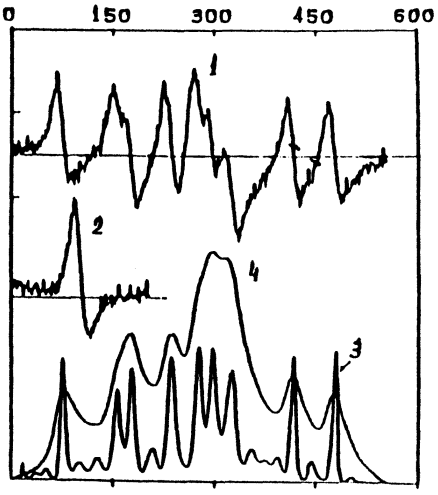


Fig.3 Application of ML algorithm (7) to NMR data processing [17].
1 - original experimental data for NMR absorbtion; 2 - PSF, which was determined by gluing together the left and the right tails of curve 1; 3 - restoration result for NMR absorbtion; 4 - numerical integral of the curve 1.

5 References

1. S.G.Rautian. *Uspechi fisicheskih nauk.* **66**, No.3, 475-517, 1958 (in Russian)
2. V.F.Turchin et al. *Uspechi fisicheskih nauk.* **102**, No.3, 345-386, 1970 (in Russian)
3. L.A.Vainstein. *Doklady AN SSSR.* **204**, No.5, 1067-1070, No.6, 1331-1334, 1972 (in Russian)
4. B.R.Frieden. Image Enchancement and Restoration. In *Picture processing and digital filtering*, pp.178-248. Ed. T.S.Huang, Springer, Berlin, 1979
5. E.L.Kosarev. *Comput.Phys.Comm.* **20**, 69-75, 1980
6. A.N.Tichonov and Ya.Arsenin. *Solution of ill-posed problems.* Halsted, N.Y.,1977
7. J.M.Mandel. *Optimal Seismic Deconvolution.* Academic Press, N.Y.,1983
8. *Deconvolution with application in spectroscopy.* Ed. D.A.Jansson. AP, N.Y., 1984
9. C.Shannon. Communications in the presence of noise. *Proc.IRE.* **37**, No.1, 10-21, 1949
10. A.N.Kolmogorov. Theory of information transmission. In *Information and algorithm theory.* Collected papers, Nauka, Moscow, 1987 (in Russian)
11. Lord Rayleigh. The wave theory of light. In *Encyclopedia Britanica*, **24**, 1884
Also in *The Scientific Papers of Lord Rayleigh.* **3**, 47-189, Cambridge Univ.Press, 1887-1892
12. E.L.Kosarev, V.D.Peskov, E.R.Podolyak. *Nucl. Instr. and Meth.*, **208**, 637-645, 1983
13. M.Z.Tarasko. Preprint *FEI-156*, Obninsk, 1969 (in Russian)
14. N.J.Bershad. *J.Opt.Soc.Am.***59**, 157-163, 1969
15. W.F.Gabriel.*Proc. IEEE*, **68**, No.6, 654-666, 1980
16. V.V.Karavaev, V.S.Molodtsov. *Radiotechnika i Electronika*, **32**, No,1, 22-26, 1987 (in Russian)
17. N.E.Alekseevskii, E.G.Nikolaev. *Journ.Exper.and Theor.Phys.*. **91**, No.5(11), 1820-1831, 1986 (in Russian)

A MONOTONIC PROPERTY OF DISTRIBUTIONS BASED ON ENTROPY WITH FRACTILE CONSTRAINTS

V. SOLANA and N.C. LIND
Consejo Superior de Investigaciones Científicas, CECIME
Serrano 123, E-28006 Madrid
Spain.

ABSTRACT. A consistency principle for distribution functions used in reliability and risk analysis has been proposed recently as follows: The distribution function $Q_X(x, x_1, x_2, \ldots, x_r)$ of a real random variable X assigned on the basis of a random sample $[x_1, ..., x_r]$ of X, should everywhere be monotonically non-increasing in all elements of the sample $(\partial Q_X/\partial x_i \leq 0, \ \forall\ i \ \subset 1, \ldots, r)$. This principle is sometimes, but generally not, satisfied by conventional methods of estimation that employ sample moments.

The paper shows that this principle of consistency is satisfied by the posterior distribution obtained by minimizing the cross-entropy with respect to any reference ("prior") distribution under fractile constraints. This "data monotonicity" is also shown to extend to the case of multivariate distributions. The marginal distribution function obtained by a probability-preserving monotonic transformation of a finite set of random variables is everywhere monotonically non-increasing with respect to any jointly observed realizations of the random variables.

The consistency property is further shown to hold for system reliabilities in the multivariate case. This consistency of entropy-based reliability analysis is methodologically of great importance and provides a strong reason to use the entropy method together with fractile constraints in the assignment of distributions.

1. INTRODUCTION

The subject of this paper is the problem of assigning a probability distribution $q(x)$ to a random variable X, given only a random sample. Each solution to this inference problem rests on a specific set of assumptions, the solution "rationale", that must not be in conflict with the reality that the distribution is meant to represent. Increasingly there are severe requirements placed on such rationales, particularly in reliability and risk analysis when the distribution is part of a model used to assess great risks, risks to the public, or risk to people who receive little concomitant benefit. Arbitrary assumptions of distribution type, as employed in many classical and Bayesian methods to solve this problem, are often unacceptable.

Even when the distribution type assumption may be acceptable, there may be shortcomings in the methods of parameter estimation. One such defect occurs in reliability analysis when using classical estimation methods. In reliability analysis the set of failed states is a region of the space of basic random variables. Failure is the crossing of the limit states surface separating the failed states from the safe states. The probability of failure is calculated as the integral of the probability density of states over this region. The simplest example is a system having two independent basic random variables, namely a load S and the capacity R to resist this load. The reliability measure is calculated by a subtracted convolution of R and S. Suppose that the load variable S is modelled by a normal distribution, with parameters estimated from the sample mean m_s and sample variance s_s^2 of a set of observed loads. Now imagine that one sample value x_i below the mean is reduced;

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 481–490.

this lowers the mean load but increases the standard deviation. The combined effect may be to lower the reliability, which is absurd, as pointed out by Öfverbeck (1984) and others.

To avoid this absurdity, a consistency principle for distribution functions in reliability and risk analysis has been proposed as follows (Lind and Chen 1987): The distribution function $Q_X(x, x_1, x_2, \ldots, x_r)$ of a random variable X, defined on a domain D^{r+1}, where D is a continuous subset of the set of real numbers, and assigned on the basis of a random sample $[x_1, \ldots, x_r]$ of X, should be monotonically non-increasing in all elements of the sample $(\partial Q_X/\partial x_i \leq 0,\ \forall i \subset 1, \ldots, r)$. This principle is not normally satisfied by the conventional methods of estimation that employ sample moments. The main result of this paper is that this principle is satisfied by distributions that minimize the cross-entropy subject to fractile constraints. A function of the elements of a random sample is said in this paper to be *data monotonic* if it is monotonically non-increasing in every sample element everywhere in its domain. Otherwise, the monotonically non-decreasing condition in every sample element will be specified. The consistency principle states that distribution functions in reliability and risk analysis should be data monotonic.

Unjustifiable assumptions of distribution type may be avoided if Jaynes' principle of maximum entropy is invoked to identify the unique solution to the problem that can be called the least biased distribution. This is accomplished by minimizing the Kullback-Leibler cross-entropy (Shore and Johnson 1980). It must be admitted that this involves a reference distribution, usually called the "prior" distribution, which must be assumed.

The constraints in the entropy method are normally in the form of prescribed expectations of functions of x. When the information about X is obtained directly in the form of observed moments, it is natural to use these measured moments as constraints. But the case considered in this paper, when the data are a set of individual sample values of X, is perhaps more common. The straightforward way to produce the constraints from such data is to calculate some set of sample moments $[m_k]$ and put this equal to the corresponding set of moments of $q(x)$. Unfortunately, several objections can be raised against this "method of moments". It is sufficient to note that a selection of a particular subset of moments $[m_k]$ to represent the data involves an arbitrary choice and induces extraneous information. This common adaptation of the method of moments is incorrect.

2. DISTRIBUTIONS WITH FRACTILE CONSTRAINTS

However, there is a practical alternative to moment constraints, namely fractile constraints. The basis is the following well known exact property of random sampling (Feller 1968; Madsen et al. 1986). A new observation of a random variable X has equal probability of falling in the $r+1$ intervals into which the elements of a random sample of size r divides the domain of X. This may be expressed in the form of the following *Sample Rule*: The elements of a random sample of size r of a random variable are the $i/(r+1)$-fractiles $(i = 1, 2, \ldots, r)$ of the distribution of the random variable.

Consider a real-valued random variable X that has the finite or infinite domain $I = [x_0, x_{r+1}]$, partitioned into $r+1$ subintervals $I_0 = [x_0, x_1), I_1 = [x_1, x_2), \ldots, I_r = [x_r, x_{r+1}]$. Given is a reference distribution P(x) having density function $p(x)$ that is positive everywhere in I. The value of $P(x)$ at $x = x_i$ is denoted by P_i. The fractile pairs

$$(x, Q(x))_i = (x_i, Q_i),\ \ i = 1, 2, \ldots, r \tag{1}$$

are prescribed, for example according the sample rule; we seek a *posterior distribution* $q(x)$, with distribution function $Q(x)$, that minimizes the cross-entropy functional

$$D(q,p) = \int_I q(x)[log\ q(x) - log\ p(x)]dx, \tag{2}$$

and satisfies the fractile constraints (1). The general solution, determined using discontinuity functions and Lagrange's multiplier method (Lind and Solana 1988), may be expressed in terms of the interval multipliers

$$\mu_i = (Q_{i+1} - Q_i)/(P_{i+1} - P_i). \tag{3}$$

The solution is

$$q(x) = \mu_i\ p(x),\ x \in I_i, i = 1,\ldots,r, \tag{4}$$

$$Q(x) = Q_i + \mu_i\ (P(x) - P_i), x \in I_i, i = 1,\ldots,r. \tag{5}$$

The cross-entropy functional (2) takes the minimum value

$$D_{min} = \sum_0^r (Q_{i+1} - Q_i)\ log[(Q_{i+1} - Q_i)/(P_{i+1} - P_i)]. \tag{6}$$

The posterior density function $q(x)$ in (4) has piecewise the form of the reference density function $p(x)$ scaled over each interval I_i by the constant factor μ_i. The minimum value of the cross-entropy functional is functionally independent of the reference density $p(x)$, and of the reference distribution function $P(x)$ except for the values P_i assumed at points x_i , $i = 1,\ldots,r$. The expression for the minimum value of the cross-entropy (6) is analogous to the minimum cross-entropy for discrete distributions (Kapur and Kesavan 1987), corresponding to a set of $r+1$ possible events.

3. DATA MONOTONICITY OF UNIVARIATE DISTRIBUTION FUNCTIONS

Consider now the effect on $Q(x) \equiv Q_X(x, x_1, x_2, \ldots, x_r)$ of a change in one of the fractile constraints (1). Assume that there is a constraint on such changes in the form of a functional dependence of the probability value on the fractile. This dependence may be written as $Q = F(x)$, i.e. formally as a distribution function containing the fractiles $Q_i = F(x_i), i = 1,\ldots,r$. By differentiation of (3),

$$\frac{\partial \mu_i}{\partial x_i} = \frac{1}{P_{i+1} - P_i}\left[\mu_i p(x_i) - \frac{dQ_i}{dx_i}\right],\ i = 1,\ldots,r \tag{7}$$

$$\frac{\partial \mu_i}{\partial x_{i+1}} = -\frac{1}{P_{i+1} - P_i}\left[\mu_i p(x_{i+1}) - \frac{dQ_{i+1}}{dx_{i+1}}\right],\ i = 1,\ldots,r \tag{8}$$

Substituting (7) into the partial derivatives of (5) gives

$$\frac{\partial Q(x)}{\partial x_i} = -\left[1 - \frac{P(x) - P_i}{P_{i+1} - P_i}\right]\left[\mu_i p(x_i) - \frac{dQ_i}{dx_i}\right], x_i \in I_i, i = 1,\ldots,r \tag{9}$$

$$\frac{\partial Q(x)}{\partial x_{i+1}} = -\left[\frac{P(x) - P_i}{P_{i+1} - P_i}\right]\left[\mu_i p(x_{i+1}) - \frac{dQ_{i+1}}{dx_{i+1}}\right],\ x_i \in I_i, i = 1,\ldots,r \tag{10}$$

All other partial derivatives of $Q(x)$ with respect to the sample elements vanish. Since the terms in the first brackets on the right hand side of (9) and (10) are positive, a condition that $Q(x)$ is data monotonic is simply that the second brackets are positive. These conditions depend, however, on the density values $q(x_i), i = 1, \ldots, r$ of the posterior distribution, which also depend on the prior density (4). $Q(x)$ is data monotonic if the slope dF/dx of the function constraining the changes in the fractiles is less than the slopes $q(x)$ of the posterior distribution function in a neighborhood of all constraints (x_i, Q_i).

In particular, when the fractile constraints arise from the sample rule given in section 2, the probability values Q_i are independent of the observations x_i, i.e. $dQ_i/dx_i = 0$. Then, the second brackets in (9) and (10) are always positive. Thus, if the probabilities of the fractile constrain pairs (x_i, Q_i), are as stated by the sample rule, then the value of the assigned posterior distribution $Q(x)$, for any argument, is data monotonic.

This data monotonicity of univariate distribution functions is the same as the "consistency principle" for distributions in reliability and risk analysis stated by Lind and Chen (1987). However, data monotonicity is found herein not as a principle, but as a property of distributions that satisfy the sample rule and minimize the cross-entropy with respect to a fixed reference distribution.

The data monotonicity property extends immediately to the distributions $Q_Z(z)$ of any random variable Z defined by a monotonic transformation of $X, z = g(x)$, preserving the probability measure. This follows as a consequence of the invariance of cross-entropy methods (Shore and Johnson 1980) and the invariance of the sample rule under monotonic transformations of the random variable. Hence, data monotonicity of a distribution $Q_X(x)$ is an invariant property under the group of monotonic transformations of variables.

4. DATA MONOTONICITY IN THE MULTIVARIATE CASE

The general solution of cross-entropy estimation of multivariate distributions subjected to fractile constraints has been obtained by Lind and Solana (1988a) in the cases of stochastic dependence and independence of random variables. In this solution the sample rule given in section 2 was employed to define the fractile constraints.

In the following we consider only the case of stochastic independence. The main result in this case is that data monotonicity is a property of the marginal distributions of the minimum cross-entropy multivariate distributions. This is a consequence of the system independence property of entropy methods (Shore and Johnson. 1980). The solution of distributions minimizing cross-entropy and the data monotonicity property of their marginal distributions are only illustrated for the case of two random variables.

Let $(x_k, y_k), k \in N = (1, 2, \ldots, r)$, be a set of observed pairs of values of a continuous vector-valued random variable (X, Y) for some system that has a continuous set E of possible states, $E \subset R^2$. The components of the observed data vectors are reordered as increasing values. Thus, two sets of r-scalar pairs (x_i, Q_{Xi}) and $(y_j, Q_{Yj}), i, j \in N$, of marginal fractile data are obtained, in which Q_{Xi} and Q_{Yj} are the marginal cumulative probabilities at data points x_i and y_j, $Q_{Xi} = Q_X(x_i)$ and $Q_{Yj} = Q_Y(y_j)$. These probabilities are obtained, for instance, by the sample rule.

Next, the fractile constraints arise from the set of $2r$ marginal fractile pairs (x_i, Q_{Xi}) and (y_j, Q_{Yj}) and the condition that X and Y are stochastically independent. These constraints can be represented as a whole, in a unique way, by the set of r^2-vectorial fractile

constraints pairs $((x_i, y_j), Q_{XYij}), i, j \in N$, where Q_{XYij} are the cumulative probabilities at the vector data point, such that $Q_{XYij} = Q_{Xi}(x_i)\, Q_{Yj}(y_j)$.

Given is a reference distribution function $P_{XY}(x,y) = P_X(x)\, P_Y(y)$ having reference density $p_{XY}(x,y) = p_X(x)p_Y(y)$. Consider the cells E_{ij} in which the state domain E is partitioned, such that $E_{ij} = I_{Xi} \cap I_{Yj}$. I_{Xi} and I_{Yj} are the subintervals $I_{Xi} = [x_i, x_{i+1})$ and $I_{Yi} = (y_j, y_{j+1}]$, $i, j \in N \cup \{0\}$.

The general solution may be found by Lagrange's multiplier method. It may be expressed in terms of the cell multipliers

$$\mu_{ij} = \frac{(Q_{Xi+1} - Q_{Xi})(Q_{Yj+1} - Q_{Yj})}{(P_{Xi+1} - P_{Xi})(P_{Yj+1} - P_{Yj})} \tag{11}$$

The minimum cross-entropy density functions and distributions are

$$q_{XY}(x,y) = \mu_{ij}\, p_X(x)\, p_Y(y), \quad (x,y) \in E_{ij}; i, j \in N \cup \{0\} \tag{12}$$

$$[Q_X(x) - Q_{Xi}][Q_Y(y) - Q_{Yj}] = \mu_{ij}\, [P_X(x) - P_{Xi}][P_Y(y) - P_{Yj}], (x,y) \in E_{ij}; i, j \in N \cup \{0\}. \tag{13}$$

The posterior density $q_{XY}(x,y)$ in (12) has piecewise the form of the reference density function $p_{XY}(x,y)$ scaled over each cell E_{ij} by the constant factor μ_{ij}.

Now the marginal distributions of the minimum cross-entropy bivariate distribution $Q_{XY}(x,y)$ are derived from (11) and (13) as follows.

The specialization of (13) for $x = x_{i+1}$, and $y = y_{j+1}$, and the substitutions of the cell multiplier μ_{ij} given by (11), determine the following expressions for marginal distributions:

$$Q_X(x) - Q_{Xi} = \frac{Q_{Xi+1} - Q_{Xi}}{P_{Xi+1} - P_{Xi}} [P_X(x) - P_{Xi}], x \in I_{Xi}, i \in N \cup \{0\}. \tag{14}$$

$$Q_Y(y) - Q_{Yj} = \frac{Q_{Yj+1} - Q_{Yj}}{P_{Yj+1} - P_{Yj}} [P_Y(y) - P_{Yj}], y \in I_{Yj},\ j \in N \cup \{0\}. \tag{15}$$

The entropy method for estimation of univariate distributions with fractile constraints, given in section 2, may also be applied to the marginal reference distributions $P_X(x)$ and $P_Y(y)$. The results are equivalent to (14) and (15), and the interval multipliers are

$$\mu_{Xi} = (Q_{Xi+1} - Q_{Xi})/(P_{Xi+1} - P_{Xi}) \tag{16}$$

$$\mu_{Yj} = (Q_{Yj+1} - Q_{Yj})/(P_{Yj+1} - P_{Yj}). \tag{17}$$

Therefore, the data monotonicity of the minimum cross-entropy univariate distribution, established in section 3, is also a property of the marginal distributions (14) and (15). Hence, if the marginal fractile constraints (x_i, Q_{Xi}) and (y_j, Q_{Yj}), $i, j \in N$, are assigned from the sample rule, then the marginal distributions $Q_X(x)$ and $Q_Y(y)$ of the multivariate posterior distribution, for any values of their argument x or y, are monotonically non-increasing as functions of the observations x_i and $y_j; i, j \in N$.

5. INVARIANCE OF MARGINAL DISTRIBUTION DATA MONOTONICITY

In section 3 it was shown that data monotonicity of univariate distributions is an invariant property of the set of monotonic transformations of a random variable. Now, data monotonicity of the marginal distributions of multivariate distributions, established in section 4, is examined in the case of a change of random variables.

The main result in this section is that data monotonicity of marginal distributions is an invariant property of a set of random variables transformations. In the same way as in section 4, only the case of two random variables is presented. Extension of results to the case of multiple variables is easily derived.

Let T be a monotonic probability-preserving transformation from random variables X and Y to random variables U and V, given by the mapping $u = g(x, y)$ and $v = h(x, y)$. Monotonicity of T implies that the first order partial derivatives of the functions g and h do not vanish in any point of E. Let $Q_U(u)$ and $Q_V(v)$ be the marginal distributions of the multivariate posterior distribution $Q_{UV}(u, v)$. These marginal distributions correspond to the following integrals:

$$Q_U(u) = \int_{E_U} q_{XY}(x, y)\, dx\, dy \tag{18}$$

$$Q_V(v) = \int_{E_V} q_{XY}(x, y)\, dx\, dy \tag{19}$$

where the integral domains are the subsets E_U and E_V, such that $g(x, y) \leq u$, $\forall (x, y) \in E_U$, and $h(x, y) \leq v$, $\forall (x, y) \in E_V$.

Integrals (18) and (19) extend over all the cells E_{ij} which are total or partially contained in the subsets E_U and E_V. Thus, in view of (12),(18) and (19), the marginal distributions may be rewritten

$$Q_U(u) = \sum_{o}^{r} \sum_{o}^{r} Q_{U,ij}(u) \tag{20}$$

$$Q_V(v) = \sum_{o}^{r} \sum_{o}^{r} Q_{V,ij}(v) \tag{21}$$

$Q_{U,ij}(u)$ and $Q_{V,ij}(v)$ are auxiliary step functions for each cell E_{ij}, given by

$$Q_{U,ij}(u) = \mu_{ij} \int_{E_{U,ij}} p_X(x) p_Y(y)\, dx\, dy,\ (x, y) \in E_{ij} \tag{22}$$

$$Q_{V,ij}(u) = \mu_{ij} \int_{E_{V,ij}} p_X(x)\, p_Y(y)\, dx\, dy,\ (x, y) \in E_{ij} \tag{23}$$

where the integration domains are $E_{U,ij} = E_U \cap E_{ij}$and $E_{V,ij} = E_V \cap E_{ij}$.

The sums in (20) and (21) extend to all non-empty subsets $E_{U,ij}$ and $E_{V,ij}$, respectively.

The partial derivatives of marginal distributions $Q_U(u)$ and $Q_V(v)$ with respect to the components of the observed data vectors, x_i and y_i, $i, j \in N$, of vector random variable (X, Y) are calculated next. By (20) and (21), they reduce to the sums of partial derivatives of $Q_{U,ij}(u)$ and $Q_{V,ij}(v)$.

When one cell E_{ij} is totally contained in E_U and E_V, the auxiliary step functions obtained from (13) are reduced to the following constant values

$$\left.\begin{matrix} Q_{U,ij}(u) \\ \text{and} \\ Q_{V,ij}(v) \end{matrix}\right\} = (Q_{Xi+1} - Q_{Xi})(Q_{Yj+1} - Q_{Yj}), \text{ for } E_{U,ij} = E_{ij} \text{ and } E_{V,ij} = E_{ij} \quad (24)$$

Since the derivatives of (24) vanish, the non-zero contributions from $Q_{U,ij}(u)$ and $Q_{V,ij}(v)$ to the partial derivatives of $Q_U(u)$ and $Q_V(v)$ only correspond to the cells E_{ij} partially contained in E_U and E_V. These cells are expressed by the conditions $E_{U,ij} \neq E_{ij}$ and $E_{V,ij} \neq E_{ij}$. Next, such contributions are calculated for a typical cell.

Only the derivatives of one of the marginal distributions, typically $Q_U(u)$, will be necessary. Then, two types of functions $u = g(x, y)$ are possible, which correspond to the monotonic conditions $dy/dx < 0$, case a) and $dy/dx > 0$, case b), for only constant value of u . The analysis is made, for instance, in the first case $(dy/dx < 0)$. Yet, there are two possibilities as the way in which the values of u increase, such that $g(x, y) < u$. Suppose that u increase on the right hand part in which E is divided by a function $u = g(x, y)$ and $dy/dx < 0$ for only constant value of u.

Now, the partial derivatives of the integrals in (22) are equivalent to the following expressions:

$$\frac{\partial[Q_{U,ij}(u)/\mu_{ij}]}{\partial x_i} = -p_X(x_i)\,[P_Y(t_i) - P_{Yj}]$$

$$\frac{\partial[Q_{U,ij}(u)/\mu_{ij}]}{\partial y_j} = -p_Y(y_j)\,[P_X(s_j) - P_{Xi}] \quad (25)$$

$$\frac{\partial[Q_{U,ij}(u)/\mu_{ij}]}{\partial x_{i+1}} = p_X(x_{i+1})\,[P_Y(t_{i+1}) - P_{Yj}]$$

$$\frac{\partial[Q_{U,ij}(u)/\mu_{ij}]}{\partial y_{j+1}} = p_Y(y_{j+1})\,[P_X(s_{j+1}) - P_{Xi}] \quad (26)$$

In (25) and (26) the variables s and t define the crossing points of curves $g(x, y) = u$ and the cell boundaries of E_{ij}. Such points are two of the next four:

$$(x = x_i, y = t_i); (x = x_{i+1}, y = t_{i+1}); (x = s_{j+1}, y = y_{j+1}) \text{and } (x = s_j, y = y_j);$$

$$\text{for } y_j \leq t_{i+1} < t_i \leq y_{j+1} \text{and } x_i \leq s_{j+1} < s_j \leq x_{i+1}. \quad (27)$$

Substituting (25), (26) and derivatives of (11) into the partial derivatives of (22) give

$$\frac{\partial Q_{U,ij}(u)}{\partial x_i} = \frac{\mu_{ij}\, p_X(x_i)}{P_{Xi+1} - P_{Xi}} \left[\frac{Q_{U,ij}(u)}{\mu_{ij}} - [P_Y(t_i) - P_{Yj}](P_{Xi+1} - P_{Xi})\right]$$

$$\frac{\partial Q_{U,ij}(u)}{\partial x_{i+1}} = \frac{\mu_{ij}\, p_X(x_{i+1})}{P_{Xi+1} - P_{Xi}} \left[\frac{-Q_{U,ij}(u)}{\mu_{ij}} + [P_Y(t_{i+1}) - P_{Yj}](P_{Xi+1} - P_{Xi})\right]$$

$$\frac{\partial Q_{U,ij}(u)}{\partial y_j} = \frac{\mu_{ij}\, p_Y(y_j)}{P_{Yj+1} - P_{Yj}} \left[\frac{Q_{U,ij}(u)}{\mu_{ij}} - [P_X(s_j) - P_{Xi}](P_{Yj+1} - P_{Yj}) \right]$$

$$\frac{\partial Q_{U,ij}(u)}{\partial y_{j+1}} = \frac{\mu_{ij}\, p_Y(y_{j+1})}{P_{Yj+1} - P_{Yj}} \left[\frac{-Q_{U,ij}(u)}{\mu_{ij}} + [P_X(s_{j+1}) - P_{Xi}](P_{Yj+1} - P_{Yj}) \right] \tag{28}$$

Since the terms in the main brackets on (28) are not positive, the derivatives of $Q_{U,ij}(u)$ with respect to the sample values x_i, x_{i+1}, y_j and y_{j+1} are data monotonic.

Therefore, in the case a), i.e. when the values of u increase on the right hand side from the curve $g(x,y) = u$, the values of derivatives of $Q_U(u)$, for each cell such that $E_{U,ij} \neq E_{ij}$, are data monotonic. Otherwise, in the case b), i.e. when the values of u increase on the left hand side from $g(x,y) = u$, the values of derivatives of $Q_U(u)$, for each cell such that $E_{U,ij} \neq E_{ij}$, are data monotonic (non-decreasing).

As conclusion, in the case of a function $u = g(x,y)$ such that $dy/dx < 0$ when u is a constant, the derivatives of marginal distributions $Q_U(u)$ obtained from (20), as the sum of derivatives of $Q_{U,ij}(u)$ for individual cells, are data monotonic when:

(a) the values of u increase on the right hand side from the curve $u = g(x,y)$,

$$\frac{\partial Q_U(u)}{\partial x_i} \leq 0 \text{ and } \frac{\partial Q_U(u)}{\partial y_j} \leq 0, \text{ for } i = 1,\ldots,r, \quad j = 1,\ldots,r. \tag{29}$$

(b) the values of u increase on the left hand side from the curve $u = g(x,y)$,

$$\frac{\partial Q_U(u)}{\partial x_i} \geq 0 \text{ and } \frac{\partial Q_U(u)}{\partial y_j} \geq 0, \text{ for } i = 1,\ldots,r, \quad j = 1,\ldots,r. \tag{30}$$

The marginal distributions $Q_U(u, x_1,\ldots,x_r, y_1,\ldots,y_r)$ may be also given as functions of r^2 ordered values of u, defined by $u_k = g(x_i, y_j), k = 1,\ldots,r^2$, for $i = 1,\ldots,r$ and $j = 1,\ldots,r$. This is a consequence of transformation T being monotonic and probability-preserving, in which case the inverse transformation T^{-1} exist providing the inverse functions for x and y values.

Hence, in the case a), taken into account (29) and the positiveness of $\partial x_i/\partial u_k$ and $\partial y_j/\partial u_k$, give

$$\frac{\partial Q_U(u)}{\partial u_k} \leq 0 \tag{31}$$

In case b), taken into account (30) and the negativeness of $\partial x_i/\partial u_k$ and $\partial y_j/\partial u_k$, give the same inequality (31) as in the case a).

The same results are obtained in case of a function $u = g(x,y)$ such that $dy/dx > 0$ when u is a constant. The proof is immediately derived in this case by changing the random variable X for $(-X)$, reducing it to the first case $(dy/dx < 0)$.

Finally the following conclusion is drawn: Data monotonicity of the marginals $Q_X(x)$ and $Q_Y(y)$ of the multivariate distributions of $Q_{XY}(x,y)$ when X and Y are stochastic independent variables, extends to the marginal distributions $Q_U(u)$ and $Q_V(v)$ of multivariate distributions of random variables U and V given by a monotonic probability preserving transformation of X and Y, $u = g(x,y)$ and $v = h(x,y)$.

Therefore, the data monotonicity of the marginal distributions of multivariate distributions that minimize the cross-entropy subjected to fractile constraints with probabilities stated by the sample rule, is an invariant property of the group of monotonic probability-preserving transformations , given by

$$\frac{\partial Q_U(u)}{\partial u_k} \leq 0 \text{ , and } \frac{\partial Q_V(v)}{\partial v_\ell} \leq 0$$

$$\text{for } u_k = g(x_i, y_j), v_\ell = h(x_i, y_j); \text{ such that } k \text{ and } \ell = 1, \dots, r^2, \text{ for } i,j \in N. \tag{32}$$

6. APPLICATION TO RELIABILITY OF SYSTEMS

The data monotonicity property of marginal distributions of multivariate distributions based on entropy with fractile constrains is immediately applied to the analysis of system reliability, as follows.

A reliability measure is given by the probability of failure that corresponds to the probability density of states of a system over a failure region. Assume that the failure region is given by the limit state functions $u = g(x,y)$ and $v = h(x,y)$, which correspond to the group of monotonic probability-preserving transformations of variables. Then, the reliability distributions for these limit states can be interpreted as the marginal distributions $\phi_U(u)$ and $\phi_V(v)$ of the multivariate distributions of transformed variables U and V, respectively. Then, the data monotonicity of marginal distributions gives that the reliability distribution function for each limit state function, is monotonically non-increasing for any argument, with regard to the values u_k, $k = 1, \dots, r^2$, of the state function $u = g(x,y)$, for the set of r^2-vectorial fractile data points with components x_i and y_j; $i,j = 1, \dots, r$.

This amounts to a generalization of the "consistence principle" of distributions of random variables for reliability and risk analysis, stated by Lind and Chen (1987). In this generalization the "consistency principle" refers not only to the assigned univariate distributions of random variables, such as the entropy-based distributions with fractile constrains, but also to the reliability distributions for any limit state function.

7. CONCLUSIONS

(1) The conventional use of moments to provide constraints from a random sample for the entropy-based assignment of a random variable is incorrect, mainly because it introduces arbitrary information implied in the selection of moments.

(2) The method of minimum cross-entropy is preferable for the assignment of distributions to continuous random variables because it yields results that are invariant under monotonic variable transformations. This extends to the invariance of monotonicity with respect to observations in a random sample, i.e. data monotonicity.

(3) Distributions assigned by the minimum cross-entropy method using fractile constraints that arise from the sample rule are data consistent, i.e. monotonically non-increasing over their entire domain in all elements of the sample.

(4) Data consistency in the sense of this paper for a distribution is the same as the consistency principle for reliability and risk analysis stated by Lind and Chen (1987).

(5) Data monotonicity of univariate distributions is an invariant property under the group of monotonic transformations of variables.

(6) The marginal distributions of any multivariate distribution assigned by cross-entropy minimization, using independent reference distributions, subject to sample rule fractile constraints, are data monotonic. Data monotonicity of these marginal distributions is an invariant property of the group of monotonic probability-preserving variable transformations.

ACKNOWLEDGMENT. Sincere thanks are due to A. Arteaga who gave freely of his time and provided many useful discussions.

REFERENCES

[1] W. Feller, An Introduction to Probability Theory and its Applications, John Wiley and Sons, New York, NY, 1968.

[2] E. T. Jaynes, Papers on Probability, Statistics, and Statistical Physics, D. Reidel, Dordrecht, Netherlands, 1983.

[3] J.N. Kapur and H.K. Kesavan, The Generalized Maximum Entropy Principle (with Applications), Sandford Educational Press, Waterloo, ON, Canada, 1987.

[4] N. C. Lind and X. Chen, "Consistent Distribution Parameter Estimation for Reliability Analysis", Structural Safety, vol. 4, pp. 141-149, 1987.

[5] N. C. Lind and V. Solana, "Cross-Entropy Estimation Random Variables with Fractile Constraints", Report IRR-11, Institute for Risk Research, University of Waterloo, ON, Canada, March 1988.

[6] H.O. Madsen, S. Krenk and N.C. Lind, Methods of Structural Safety, Prentice–Hall Book Co., Inc. Englewood Cliffs, NJ, 1986.

[7] P. Öfverbeck, "Small Sample Control and Structural Safety", Report TVBK-3009, Dept. Struct. Eng., Lund University, Lund, Sweden, 1980.

[8] J. Shore and R. W. Johnson, "Axiomatic Derivation of the Principle of Maximum Entropy and the Principle of Minimum Cross-Entropy", IEEE Transactions on Information Theory, Vol. IT-26, No. 1, pp. 26-37, Jan, 1980.

[9] J. Shore and R. W. Johnson, "Properties of Cross-Entropy Minimization", IEEE Transactions on Information Theory, Vol. IT-27, No. 4, pp. 472-482, July, 1981.

KINETIC THEORY AND ENSEMBLES OF MAXIMUM ENTROPY

John Karkheck
Department of Science and Mathematics
GMI Engineering & Management Institute
Flint, Michigan 48504-4898
U.S.A.

ABSTRACT. A general formalism based upon maximization of entropy is utilized to derive a kinetic theory for simple liquids. The first two equations of the BBGKY hierarchy are rendered into a set of closed kinetic equations by expressing the two-particle function as a functional of the one-particle function through maximizing entropy subject to constraints. The theory exhibits a strong H-theorem and yields the canonical ensemble as its equilibrium solution. For non-equilibrium states the theory admits two temperature scales, the distinction between which has significant effect on the value of the bulk viscosity and on the dynamics of short-time density fluctuations. The kinetic theory subsumes or eclipses many of the theories commonly applied to liquids. These techniques represent a much simpler and more transparent method to obtaining the latter and they are more amenable for making generalizations.

1. INTRODUCTION

To reconcile observed macroscopic irreversibility with the inherent reversibility of the microscopic equations of motion, e.g. Newton's second law or the Liouville equation when discussing classical systems, remains an outstanding problem in nonequilibrium statistical mechanics[1].

The former is typified, for fluids, by the equations of hydrodynamics which emanate from macroscopic conservation of mass, momentum, and energy[2] when supplemented with phenomenological relations - Fourier's law of heat conduction and Newton's stress tensor - and an equation of state which relates energy to density and temperature.

Two remarkable features of these equations are pertinent here: their very small number - five coupled irreversible equations - and their formal "closedness" which in essence makes the hydrodynamic variables a complete set[3] for describing the state of the fluid. Though the applicability of these conventional hydrodynamic equations is limited to states for which the variables are well-defined and their spatial and temporal variation are slow[4], nonetheless, the viability of this contracted description demonstrates that the full many-body micro-

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 491–496.

scopic description contains considerably more information than is elicited in experiments. Since the Liouville equation is reversible, it is through the contraction procedure that irreversibility enters, though precisely how remains an open question.

From the statistical-mechanical viewpoint, one seeks to project out irrelevant information from the Liouville equation to produce simpler kinetic theories which not only support the hydrodynamic equations but also provide equations of state and microscopic expressions for the transport coefficients. How to project and what to project, and what are appropriate ensembles for nonequilibrium states, are examined here by using maximization of entropy to effect closure of the BBGKY hierarchy.

2. DESCRIPTION OF THE THEORY

For definiteness, we consider an interatomic potential, V(r), of the form

$$V(r) = \begin{cases} \infty & r < \sigma \\ \phi(r)\theta(R - r) & r > \sigma \end{cases} \tag{1}$$

where $\phi(r)$ is a smooth function. The discontinuity in the potential, $\varepsilon = -\phi(R) > 0$, introduced by the Heaviside function, θ, is used to effect instantaneous interchange of kinetic and potential energy between a pair of particles passing through a separation $r = R$.

The kinetic theory is required to yield hydrodynamic equations. To produce conservation of mass and momentum equations, it is sufficient to have an equation for f_1, the one-particle distribution function. But for conservation of energy we must append an equation for the potential energy density,

$$e_p(\vec{r},t) = \frac{1}{2}\int dx_1 dx_2\ V(r_{12})\ f_2(x_1,x_2,t)\ \delta(\vec{r} - \vec{r}_1), \tag{2}$$

where the position-velocity sextuple $x = \vec{r},\vec{v}$, $r_{12} = |\vec{r}_1 - \vec{r}_2|$, and f_2 is the two-particle distribution function. Exact equations for these, obtained from the BBGKY hierarchy, are

$$\left(\frac{\partial}{\partial t} + \vec{v}_1\cdot\nabla_1\right)f_1(x_1,t) = \frac{1}{m}\int_{\sigma<r_{12}<R} dx_2 \nabla_1\phi\cdot\frac{\partial}{\partial\vec{v}_1} f_2(x_1,x_2,t) +$$

$$+ \int d\vec{v}_2 \int d\hat{\sigma}\theta(\hat{\sigma}\cdot\vec{g})\hat{\sigma}\cdot\vec{g}\Big\{\sigma^2[f_2(\vec{r}_1,\vec{v}_1',\vec{r}_1+\vec{\sigma},\vec{v}_2',t) - f_2(x_1,\vec{r}_1-\vec{\sigma},\vec{v}_2,t)] +$$

$$+ R^2[f_2(\vec{r}_1,\vec{v}_1^{+},\vec{r}_1+\vec{R}^{-},t) - f_2(x_1,\vec{r}_1-\vec{R}^{+},\vec{v}_2,t)$$

$$+ \theta\left(\hat{\sigma}\cdot\vec{g} - \sqrt{\frac{4\varepsilon}{m}}\right)\left(f_2(\vec{r}_1,\vec{v}_1^{-},\vec{r}_1-\vec{R}^{+},\vec{v}_2^{-},t)-f_2(x_1,\vec{r}_1+\vec{R}^{-},\vec{v}_2,t)\right) +$$

$$+ \theta\left(\sqrt{\frac{4\varepsilon}{m}} - \hat{\sigma}\cdot\vec{g}\right)\left(f_2(\vec{r}_1,\vec{v}_1',\vec{r}_1-\vec{R}^{-},\vec{v}_2',t)-f_2(x_1,\vec{r}_1+\vec{R}^{-},\vec{v}_2,t)\right)]\Big\} \quad (3)$$

$$\frac{\partial}{\partial t} e_p(\vec{r},t) + \nabla\cdot[e_p\vec{u}(\vec{r},t) + \vec{J}_\phi(\vec{r},t)] =$$

$$\frac{1}{2}\int_{\sigma<r_{12}<R} dx_1dx_2\delta(\vec{r}-\vec{r}_1)f_2\vec{g}\cdot\nabla_2\phi(r_{12}) +$$

$$+ \frac{1}{2}\varepsilon R^2\int d\vec{v}_1d\vec{v}_2\int d\hat{\sigma}\ \hat{\sigma}\cdot\vec{g}\Big\{\theta\left(\hat{\sigma}\cdot\vec{g} - \sqrt{\frac{4\varepsilon}{m}}\right)f_2(\vec{r},\vec{v}_1,\vec{r}+\vec{R}^{-},\vec{v}_2,t)$$

$$- \theta(\hat{\sigma}\cdot\vec{g})f_2(\vec{r},\vec{v}_1,\vec{r}-\vec{R}^{+},\vec{v}_2,t)\Big\} \quad (4)$$

The singular nature of the collisions requires careful analysis in constructing the collision integrals.[5] As constructed, these equations evolve forward in time,[6] though the collision operator itself is symmetric under time reversal.[7] See ref. [8] for discussion of notation.

This pair of equations contains the unknown f_2. By expressing f_2 as a functional of f_1 and e_p, a closed set of equations is obtained. However, there is not a unique choice for f_2. Here we obtain a specific form via the approximate ensemble, ρ_m, which follows from maximizing the grand ensemble entropy

$$S[\rho] = -k_B \sum_{N=0}^{\infty} \int dx^N\ \rho(x^N,t)\ \ell n[\rho(x^N,t)a^N\ N!] \quad (5)$$

subject to the constraints of f_1 and e_p. These are not viewed as limitations on knowledge of the system. Rather, the approach affords an objective evaluation of what information is essential[9] to permit accurate inferences about behavior and properties of the many-body system. The $a = (h/m)^3$ and h and k_B are the Planck and Boltzmann constants, respectively. The resulting ensemble is

$$\rho_m(x^N,t) = (Za^NN!)^{-1}\exp\left[-\left(\sum_{i=1}^{N}\lambda(x_i,t) + \sum_{i=1<j}^{N}\beta_{ij}V(r_{ij})\right)\right], \quad (6)$$

where Z is a normalization factor, and $\beta_{ij} = \frac{1}{2}[\beta(\vec{r}_i,t)+\beta(\vec{r}_j,t)]$ and λ are Lagrange multiplier fields conjugate to e_p and f_1, respectively. This ensemble bears the same algebraic form as its equilibrium counterpart, and so yields

$$f_2(x_1,x_2,t) = f_1(x_1,t)\ f_1(x_2,t)\ g_2(\vec{r}_1,\vec{r}_2,t). \quad (7)$$

The pair correlation function g_2 can be expressed formally by the

cluster expansion

$$g_2(\vec{r}_1,\vec{r}_2,t) = \{f_{12}^M+1\}\{1+ \int dx_3 f_1(x_3,t) f_{13}^M f_{23}^M + \cdots\} \tag{8}$$

where the Mayer function now has the form $f_{ij}^M = \exp\{-\beta_{ij}V(r_{ij})\} - 1$. In principle, eq. (2) permits solving for β in terms of e_p and f_1 so that f_2 is a functional of these; using (7) in (3) and (4) renders them closed. Finally, inserting (6) into (5) yields the extremum entropy

$$S_m = k_B\{\ell nZ + \int dx\, f_1\lambda + \int d\vec{r}\, \beta e_p\}. \tag{9}$$

3. DISCUSSION

The form of ρ_m, eq. (6), permits the existence of two temperatures in the nonequilibrium fluid. The kinetic-energy temperature vested in λ and the potential-energy temperature vested in β are not equal, in general. Distinction between them is manifested in the value of the bulk viscosity.[10]

The theory exhibits a strong H-theorem: $dS_m/dt \geq 0$. From (9) and normalization of (6) it follows that $dS_m/dt = k_B\{\int dx\lambda(\partial f_1/\partial t) + \int d\vec{r}\beta(\partial e_p/\partial t)\}$. Employing (3), (4), and (7) there follows[8] $dS_m/dt =$

$$\frac{1}{2} k_B \int dx_1 dx_2\, f_2(x_1,x_2,t)\{\vec{v}_{12}\cdot\hat{r}_{12}[\delta(r_{12}-\sigma^+) + \delta(r_{12}-R^+) -$$

$$\delta(r_{12}-R^-)] + |\vec{v}_{12}\cdot\hat{r}_{12}| \sum_{\ell=1}^{4} \delta(r_{12}-R_\ell^*)\Theta_\ell(\vec{v}_{12}\cdot\hat{r}_{12})\ x$$

$$\ell n \frac{f_1(x_1,t)f_1(x_2,t)\exp(\beta_{12}E_\ell)}{b^{(\ell)}(\hat{r}_{12})f_1(x_1,t)f_1(x_2,t)} \} \geq 0.$$ Equality holds when the independent conditions are satisfied: $\delta(r_{12}-R_\ell)\Theta_\ell(\vec{v}_{12}\cdot\hat{r}_{12})\ x$ $[\exp(\beta_{12}E_\ell) - b^{(\ell)}(\hat{r}_{12})]f_1^{(0)}(x_1,t)f_1^{(0)}(x_2,t) = 0$. That at the hard core, $R_\ell = \sigma$, yields $f_1^{(0)}(x,t) = n(\vec{r},t)\left\{\frac{m}{2\pi k_B T(t)}\right\}^{3/2}\ x$ $\exp\left\{\frac{-m[\vec{v} - \vec{V}(t)]^2}{2\pi k_B T(t)}\right\}$. Those at the square-well edge, $R_\ell = R$, yield

$\beta(t) = 1/k_B T(t)$. At complete equilibrium, time derivatives vanish and the customary grand canonical ensemble is obtained: $\rho_m = \exp\{-\beta H\}$.

The H-theorem reveals that the mean-field terms in (3) and (4) do not contribute to the irreversible behavior. Thus, as exhibited here, irreversibility is a consequence of statistics - the form (7) is crucial to the analysis - and of collisional dynamics in which energy and momentum transfer occur.

The kinetic theory described here subsumes or encompasses a wide array of other theories that have been obtained by different methods. Setting ϕ = constant yields a square-well theory[8] that improves upon[8,10] the theory of Davis, Rice, and Sengers.[11] Setting $\phi = 0$ yields the revised Enskog theory which was derived through diagrammatic methods.[12] In the $R \to \infty$ limit such that $\varepsilon \to 0$, the linearized version[6,13] of the new theory encompasses the short-time kinetic theory of Lebowitz, Percus, and Sykes which was derived through linear response theory.[14]

The veracity of the theory stands in relation to the questions asked of it, since in an absolute sense the theory is only approximate. The ensemble (6) contains sufficient correlation to produce an exact short-time linearized theory[15], and the theory yields the hydrodynamic equations and provides expressions for the transport coefficients. However, the ensemble lacks velocity correlations which build up in time and appear to have a significant effect on the values of the transport coefficients, especially the shear viscosity, as compared to computer simulation results.[16] Since the transport coefficients depend upon different projections of f_2 than does e_p, an improved theory is expected if the full f_2 is used to characterize the ensemble. Such a theory has been proposed for hard spheres.[17] It is explicit but complex and has not been analyzed for transport.

Though the set of kinetic equations is Markovian, contraction down to an equation for f_1 alone yields a nonMarkovian closed equation. Thus, generally, memory can be built into a theory by imposing constraints that bear memory explicitly, as proposed by Jaynes[18], or through contraction of excess degrees of freedom. The latter seems more transparent than the former, which requires knowledge of duration of memory.

Finally, the techniques described here also yield the Boltzmann equation.[17,19]

4. ACKNOWLEDGEMENTS

I wish to acknowledge many fruitful discussions with Henk van Beijeren, Ignatz de Schepper, Jan Sengers, and George Stell, and financial support of our collaboration by NATO Grant No. 419/82. I greatly appreciate the support and encouragement provided by GMI and the typing of this manuscript by Corinne Anthony.

5. REFERENCES

1. O. Penrose, Rep. Prog. Phys., 42, 1937 (1979).
2. J.H. Irving and J.G. Kirkwood, J. Chem. Phys., 18, 817 (1950).
3. Though mass, momentum, and energy density are sufficient to describe the state of the fluid, their time evolution is governed by the transport coefficients whose conceptual status is deeper than the hydrodynamic variables themselves.
4. Generalized hydrodynamic equations have been proposed, see for example, J.P. Boon and S. Yip, Molecular Hydrodynamics, (McGraw-Hill, New

York, 1980).
5. M.H. Ernst, J.R. Dorfman, W.R. Hoegy, and J.M.J. van Leeuwen, Physica, 45, 127 (1969).
6. J. Karkheck, Kinam, 7A, 191 (1986).
7. J.A. Leegwater and H. van Beijeren, 'Linear Kinetic Theory of the Square-Well Fluid', preprint.
8. J. Karkheck, H. van Beijeren, I.M. de Schepper, and G. Stell, Phys. Rev. A, 32, 2517 (1985).
9. Y. Tikochinsky, N.Z. Tishby, and R.D. Levine, Phys. Rev. A, 30, 2638 (1984).
10. H. van Beijeren, J. Karkheck, and J.V. Sengers, Phys. Rev. A, 37, 2247 (1988).
11. H.T. Davis, S.A. Rice, and J.V. Sengers, J. Chem. Phys., 35, 2210 (1961).
12. H. van Beijeren and M.H. Ernst, J. Stat. Phys., 21, 125 (1979).
13. R.C. Castillo, E. Martina, M. Lopez de Haro, J. Karkheck, and G. Stell, 'Linearized Kinetic Variational Theory and Short Time Kinetic Theory', submitted to Phys. Rev. A.
14. J.L. Lebowitz, J.K. Percus, and J. Sykes, Phys. Rev., 188, 487 (1969).
15. J. Karkheck, Bull. Am. Phys. Soc., 32, 939 (1987).
16. J.P.J. Michels and N.J. Trappeniers, Physica A, 116, 516 (1982) and references therein.
17. J. Karkheck and G. Stell, Phys. Rev. A, 25, 3302 (1982).
18. E.T. Jaynes in The Maximum Entropy Formalism, ed. by R.D. Levine and M. Tribus (MIT, Cambridge, Mass., 1978).
19. R.M. Lewis, J. Math. Phys., 8, 1448 (1967).

ON THE USE OF QUADRATIC REGULARISATION WITHIN MAXIMUM ENTROPY IMAGE RESTORATION

A.M. Thompson
Department of Statistics
University of Glasgow
Glasgow G12 8QW
Scotland U.K.

ABSTRACT. In some circumstances, the objective function for Maximum Entropy image restoration can be closely approximated by a quadratic regularisation criterion. The details are set out for the case of the global Shannon entropy function as well as for local first-order and second order entropy functions for one- and two-dimensional pixellated images. It is proposed to use the quadratic approximation to choose the regularisation constant, which we then use, suitably scaled, as the regularisation constant in the Maximum Entropy method in order to construct a restoration of the image. Numerical simulations will be reported that demonstrate, in simple examples, the apparent effectiveness of this hybrid approach in terms of achieving a restoration that is likely to be close to the true scene.

1. INTRODUCTION

Many of the restoration techniques used for stabilising the deconvolution of pixellated images fall into one of two classes: Quadratic Regularisation techniques and Maximum Entropy Methods.

Quadratic Regularisation methods have many theoretical and computational attractions, associated with the fact that an explicit expression can be found for the recovered or estimated source function $\underline{\hat{f}}$ (regarded as a vector) in terms of the point spread matrix H, the smoothing or regularisation matrix C, and the data vector $\underline{g}$ i.e.

$$\underline{\hat{f}} = (H^T H + \lambda C)^{-1} H^T \underline{g} \qquad (1)$$

where $\lambda > 0$ is the smoothing parameter or regularisation constant.

Quadratic Regularisation can however produce unphysical solutions containing negative pixel values.

Maximum Entropy methods, on the other hand, impose positivity on the source function by means of a logarithmic term in the smoothing

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 497–504.

function, at the expense of introducing non-linearity into the problem of estimating the source function.

It is well known that, under certain circumstances, the global Shannon Entropy Function used in Maximum Entropy image reconstruction can be closely approximated by the Zeroth Order Quadratic Regularisation function. In this paper we shall set out the argument that leads to this approximate correspondence. We will also show that the same argument can be applied to demonstrate similar asymptotic relationships between higher orders of Quadratic Regularisation functions and localised forms of Maximum Entropy objective functions for both 1D and 2D pixellated images. In addition we suggest that, by suitable scaling, the smoothing parameter determined using a Quadratic Regularisation method can be used to calculate a suitable smoothing parameter for use in the corresponding Maximum Entropy image restoration method.

We shall also present the results of numerical simulations which show that, for a variety of source functions (including ones for which we would not expect the asymptotic relationships described above to hold), this quadratic method provides an 'acceptable' choice of smoothing parameter for use with Maximum Entropy methods. Thus we can use the benefits of Quadratic Regularisation in terms of speed and more effective choice of smoothing parameter while still retaining the positivity associated with Maximum Entropy techniques.

2. THE APPROXIMATE EQUIVALENCE OF MAXIMUM ENTROPY AND QUADRATIC REGULARISATION TECHNIQUES

Consider a one-dimensional true image $\underline{f}$, with elements f_i, distorted by a point spread matrix H with additive Gaussian white noise of variance σ^2, giving a data vector g. Thus

$$\underline{g} = H\underline{f} + \underline{\epsilon} \quad \text{where } \underline{\epsilon} \sim N(0,\sigma^2 I) \tag{2}$$

In order to determine the Maximum Entropy reconstruction $\hat{\underline{f}}$ of $\underline{f}$, we require to minimise the function

$$(\underline{g} - H\hat{\underline{f}})^T(\underline{g} - H\hat{\underline{f}}) + \lambda_{ME}\Sigma p_i \log p_i \tag{3}$$

where $p_i = \hat{f}_i / \Sigma_i \hat{f}_i$ and where the smoothing parameter λ_{ME} is a Lagrange multiplier which we obtain from some additional constraint, such as

$$\|\underline{g} - H\hat{\underline{f}}\|^2 = N_p\sigma^2, \tag{4}$$

where N_p is the number of pixels (e.g. Zhuang et al, 1987).

If, in the entropy term in expression (3), $\{\bar{f}-\hat{f}_i)/\bar{f}\}$ is small, where $\bar{f} = \Sigma\hat{f}_i/N_p$, then

$$\Sigma p_i \log p_i = (N\bar{f})^{-1}\Sigma f_i \log[1-\{(\bar{f}-\hat{f}_i)/\bar{f}\}] - \log N \tag{5}$$

$$= \Sigma\{\hat{f}_i[\hat{f}i/\bar{f} - 1]\}/N\bar{f} - \log N + \\ + O(\hat{f}_i[1-\hat{f}_i/\bar{f}]^2) \quad (6)$$

$$\approx (\Sigma f_i^2)/(\bar{f}N) - 1 - \log N \quad (7)$$

to first order.

Thus if $\bar{f}$ is a slowly varying function of λ

$$\Sigma p_i \log p_i \sim \Sigma f_i^2 \quad (8)$$

which is simply the Zeroth Order Regularisation function (Titterington 1985).

Thus expression (3) is approximately

$$\min\{\|\underline{g}-H\underline{f}\|^2 + \lambda_Q \Sigma f_i^2\} \quad (9)$$

where $$\lambda_Q = \lambda_{ME} N \bar{f}^2 \quad (10)$$

In obtaining this result we have made two assumptions. First we have assumed that $\bar{f} \approx$ constant. In many image processing problems, particularly those involving photographic type images, as obtained for instance from the HXIS instrument on the SMM satellite (Mackinnon et al, 1985) we expect this to be true since we would expect $\Sigma\hat{f}_i = \Sigma g_i$: the total observed number of counts in the blurred image is the same as in the true image.

The second assumption we have made is that each $\hat{f}_i$ is close to $\bar{f}$. Clearly, in a great many problems this will not be valid. Indeed, the elements of f may vary over several orders of magnitude. We shall however show, by means of numerical examples, that even although the values of $\hat{f}$ obtained using Maximum Entropy may be substantially different from those obtained using Quadratic Regularisation, the quadratic method can still be used to find a 'satisfactory' estimate for λ_{ME}.

The arguments used to get from (3) to (9) and (10) can equally well be applied to localised forms of the Maximum Entropy function. For example,

$$2\,\Sigma\, p_i \log(p_i/[p_{i+1}p_{i-1}]^{\frac{1}{2}}) = \Sigma\,(p_i - p_{i+1})(\log p_i - \log p_{i+1}) \quad (11)$$

$$\approx (N\bar{f}^2)^{-1}\Sigma(\hat{f}_i - \hat{f}_{i+1})^2 \,, \quad (12)$$

so that Maximum Entropy with a data adaptive prior $m_i = [p_{i+1}p_{i-1}]^{\frac{1}{2}}$ is asymptotically equivalent to first order regularisation.

In general we can write

$$\lambda_{ME}\underline{p}^T C \log \underline{p} \approx \lambda_Q \underline{f}^T C \underline{f} \quad (13)$$

where log $\underline{p}$ is the vector whose i^{th} element is $\log(p_i)$, C is some order of regularisation matrix and

$$\lambda_Q = \lambda_{ME}\, N\bar{f}^2 \tag{14}$$

Similar results can be obtained for two-dimensional images where the equivalent of (13) would be

$$\lambda_{ME} \sum_i \sum_j \sum_\mu \sum_\nu p_{ij} C_{ij\mu\nu} \log p_{\mu\nu} \approx \lambda_Q \sum_i \sum_j \sum_\mu \sum_\nu \hat{f}_{ij} C_{ij\mu\nu} \hat{f}_{\mu\nu} \tag{15}$$

where $\lambda_{ME} = \lambda_Q / N\bar{f}^2$, $\bar{f} = \sum \sum \hat{f}_{ij}/N$ and $C_{ij\mu\nu}$ are the elements of a smoothing or regularisation tensor.

For example, in the zeroth order case of the Shannon Entropy function,

$$C_{ijij} = 1 \quad \text{for all } ij \quad \text{and} \quad C_{ij\mu\nu} = 0 \ (\mu.\nu) \neq (i,j)$$

and, in the case of first order smoothing

$$C_{ijij} = 4 \qquad C_{ij,i+1,j} = C_{iji,j-1} = C_{ij,i+1,j} = C_{iji,j+1} = -1$$

for all ij and $C_{ij\mu\nu} = 0$ otherwise.

3. NUMERICAL SIMULATIONS

Maximum Entropy and Quadratic Regularisation image reconstruction problems can be split up into 5 characteristic components:

(1) True image $\underline{f}$
(2) Point spread function H
(3) Type of noise $\underline{\epsilon}$
(4) Smoothing function
(5) Choice of smoothing parameter.

In the remainder of this paper we present the results of some numerical simulations based on the 4 one-dimensional source functions shown in figure 1.

The source functions were convolved with a circulant point spread matrix

$$\begin{bmatrix} .6 & .2 & 0 & \cdots & 0 & 0 & .2 \\ .2 & .6 & .2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & .2 & .6 & .2 \\ .2 & 0 & 0 & \cdots & 0 & .2 & .6 \end{bmatrix} \tag{16}$$

and white Guassian noise with variance σ^2, was added. 500 random realisations were produced and the estimated source function $\underline{\hat{f}}$ was calculated using the zeroth and first order Quadratic Regularisation and the corresponding Maximum Entropy method described above.

The smoothing parameters for the Regularisation methods were chosen using 2 methods. The first was the commonly used chisquared method, in which λ is chosen to satisfy

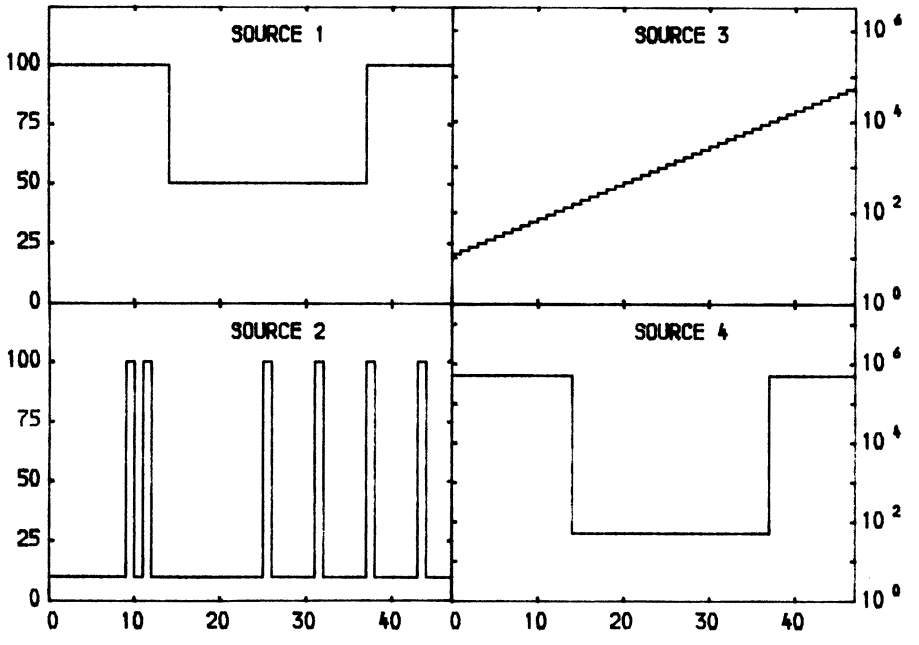


Figure 1: The four source functions used in the numerical simulations.

$$RSS(\lambda) = N_p\sigma^2 \tag{17}$$

where $RSS(\lambda) = \underline{g}^T(I-K(\lambda))^T(I-K(\lambda))\underline{g}$ and $K(\lambda) = H(H^TH+\lambda C)^{-1}H^T$. The second is called the Equivalent Degrees of Freedom method. This is essentially a method for correcting the over smoothing associated with the chisquared method (Wold & Wahba, 1975).

Instead of solving (17) we solve

$$RSS(\lambda) = \sigma^2 tr\{I-K(\lambda)\}, \tag{18}$$

where $tr\{K(\lambda)\}$ is interpreted as the number of degrees of freedom for estimating $\mathbf{E}(\underline{g}) = H\underline{f}$, and $N_p - tr\{K(\lambda)\}$ is the equivalent degrees of freedom for estimating error. (For a fuller discussion of these methods of choosing the smoothing parameter see Thompson et al (1988).)

In the case of the Maximum Entropy method we use two ways of estimating λ_{ME}. In the first we use the solution of $RSS(\lambda_{ME}) = N_p\sigma^2$, whereas, in the second, we take the values of λ_Q (defined in (9) and (10)) calculated using (17) and (18) and use the resulting value of λ_{ME} as the basis for a Maximum Entropy reconstruction.

The results of the numerical simulations described above are

presented in the form of three summary statistics. The first is

$$\hat{V} = \sum_{i=1}^{N_p} \sum_{j=1}^{N_s} (\bar{f}_i - \hat{f}_{ij})^2 / N_s N_p \tag{19}$$

where $\hat{f}_{ij}$ is the value of the intensity on ith pixel as estimated in the j^{th} simulation, $\bar{f}_i = \sum_{j=1}^{N_s} \hat{f}_{ij}/Ns$ is the estimate for ith pixel averaged over data sets, f_i is the true intensity, N_s is the number of simulations and N_p is the number of pixels. $\hat{V}$ is a measure of the instability or variability around the mean resulting from a given method of choosing λ.

The second statistic is

$$\hat{B} = \sum_{i=1}^{N_p} (\bar{f}_i - f_i)^2 / N_p, \tag{20}$$

which provides a measure of the bias imposed on the estimate of $\hat{f}$ as a result of the smoothing operation. Finally, the third statistic is the Total Mean Squared Error,

$$TMSE = \hat{V} + \hat{B} = \sum_{i=1}^{N_p} \sum_{j=1}^{N_s} (f_i - \hat{f}_{ij})^2 / N_p N_s \tag{21}$$

which provides a measure of the average proximity of $\underline{\hat{f}}$ is to the original image.

4. RESULTS AND CONCLUSIONS

Table 1 displays the summary statistics produced by the numerical simulations described in section 3. Source function 1, the small square well, satisfies the conditions outlined in section 2 for the asymptotic equivalence of Maximum Entropy and Quadratic Regularisation. We would therefore expect the results produced using these methods to be similar. This is confirmed by the almost identical results produced by each of these methods for both choices of smoothing parameter.

The results for source function 1 also illustrate the characteristic properties of the chisquared and EDF methods of choosing λ which is reproduced by all the data sets we have examined, the relative over-smoothing of the chisquared choice of λ and the under-smoothing of the EDF choice of λ: the chisquared method tends to produce a value for TMSE which is larger than the minimum value because it produces excessive bias $\hat{B}$, while the EDF method produces a larger than optimal TMSE because of an excessive value of $\hat{V}$.

Source function 2 is a one-dimensional image containing a series of narrow spikes. As can be seen by comparing the results produced using Maximum Entropy and Quadratic Regularisation, the asymptotic

approximation breaks down for this source function. It is however noticeable that the quadratic method of estimating λ_{ME} tends to perform better than the straightforward chisquared choice of smoothing parameter. Similar results can be seen for source functions 3 and 4. In particular the EDF method of choosing λ_{ME} does consistently better in terms of providing reconstructions which have a small Total Mean Squared Error thereby being in this sense closer to the true scene.

Another point of note is that if we compare the results using the Zeroth and First Order forms of both Maximum Entropy and Quadratic Regularisation it is clear that the localised First Order method frequently provides a reconstruction closer to the truth than the Zeroth Order method. In fact, of the 4 source functions examined here all are recovered better by the First Order method.

We should, however, add one cautionary note to these conclusions. The numerical simulations carried out here are very limited, particularly since we have examined only one kernel, and further simulations on other versions of the problem would be valuable. These preliminary results are, however, encouraging.

ACKNOWLEDGEMENT

This research has been carried out with the support of a research grant from the UK Science and Engineering Research Council.

REFERENCES

MacKinnon, A.L. Brown, J.C., Hayward, J. (1985) Solar Phys., **99**, 231.

Thompson, A.M., Brown, J.C., Kay, J.W., Titterington, D.M. (1988). Submitted for publication.

Titterington, D.M. (1985) Astron. Astrophys., **144**, 381.

Wold, S., Wahba, G. (1975) Communications in Statistics, **4**, No. 1, 1.

Zhuang, X, Ostevold, E., Haralick, R.M. (1987) IEEE Trans. ASSP, **ASSP-35,** No. 2, 208.

TABLE 1: Summary statistics for the numerical simulations described. The prefix Q- indicates that λ was chosen using the Quadratic Regularisation method then scaled and applied to maximum entropy.

SMOOTHING METHOD	SOURCE 1 $\sigma = 10.0$			SOURCE 2 $\sigma = 2.0$			SOURCE 3 $\sigma = 2.0$			SOURCE 4 $\sigma = 10.0$		
	$\hat{V}$	$\hat{B}$	TMSE	$\hat{V}$	$\hat{B}$	TMSE	$\hat{V}$	$\hat{B}$	TMSE	$\hat{V}$	$\hat{B}$	TMSE
SHANNON ENTROPY												
CHISQUARED	48.1	72.3	120.4	7.39	34.1	41.5	23.3	15.7	39.0	483	321	805
Q-CHISQUARED	46.7	74.6	121.3	10.11	15.1	25.2	23.4	19.4	42.7	413	4750	5160
Q-EDF	105.4	26.6	132.0	20.8	0.9	21.2	27.1	0.1	27.2	766	37	803
ZEROTH ORDER QUADRATIC REGULARISATION												
CHISQUARED	47.1	72.5	119.5	14.0	52.0	65.9	26.9	29.3	56.2	679	424	1098
EDF	108.1	25.2	133.3	24.8	1.2	26.0	27.0	0.1	27.1	675	1.3	676
FIRST ORDER MAXIMUM ENTROPY												
CHISQUARED	19.3	49.5	68.8	6.9	44.1	51.0	9.6	24.8	34.4	564	1095	1659
Q-CHISQUARED	19.0	50.4	69.4	8.8	22.6	31.4	15.6	0.7	16.3	330	389000	389000
Q-EDF	55.6	24.1	79.7	19.2	0.8	20.0	25.7	0.1	25.8	563	163	756
FIRST ORDER QUADRATIC REGULARISATION												
CHISQUARED	19.2	49.7	68.8	13.6	63.0	75.4	25.5	59.0	84.5	674	1690	2360
EDF	57.4	23.6	81.0	25.0	1.2	26.2	26.9	0.1	27.0	675	1.3	676

FROM CHIRP TO CHIP, A BEGINNING

GARY J. ERICKSON[1] AND PAUL O. NEUDORFER
DEPARTMENT OF ELECTRICAL ENGINEERING,
SEATTLE UNIVERSITY, SEATTLE, WASHINGTON U.S.A.
AND
C. RAY SMITH
RESEARCH, DEVELOPMENT & ENGINEERING CENTER,
U. S. ARMY MISSILE COMMAND,
REDSTONE ARSENAL, ALABAMA U.S.A.

ABSTRACT

Engineering workstations will be used in conjunction with a silicon compiler to design a Very Large Scale Integrated Circuit for a specific use, called an Application Specific Integrated Circuit, to analyze chirped signals of the form $f(t) = A_1 \cos(\omega t + \alpha t^2) + A_2 \sin(\omega t + \alpha t^2)$. This circuit should have the ability to analyze chirped signals from audio to 10MHz. The preliminary design of the chip is complete and fabrication will follow as soon as practical.

THE ALGORITHM

Very Large Scale Integrated Circuit (**VLSI**) design has entered in the undergraduate electrical engineering curriculum, and interesting, realistic design problems are always being sought. The practical analysis of chirped signals by means of the algorithm derived by Larry Bretthorst [1987, 1988] represents a worthy project for implementation. To test the problem for its suitability and degree of difficulty, we have taken it through the design and simulation phases on a commercial workstation.

To give the reader a better understanding of the algorithm to be implemented, we give a brief summary of the Bayesian analysis of chirped signals developed by Bretthorst, which proceeds from the earlier analysis of Jaynes [1987]

We have, in general, a signal of interest: $f(t) = f(t, \{\vartheta\})$, where $f(t)$ is assumed to be a known function containing unknown parameters $\{\vartheta\}$, some of which are of interest,and some of which are not of interest (called nuisance parameters). We refer to $f(t)$ as the *model.* Now the signal measured in some system

[1]and Puget Sound Power & Light Company, Bellevue, Washington U.S.A.

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 505–510.

will be: $y(t) = f(t) + e(t)$, where $e(t)$ is the noise, including errors in the system. Measurements are made at sample times:

$$t_i \quad , \quad i = 1, \ldots, N.$$

Thus we have data $D = (d_1, \ldots, d_N)$, where $e_i = e(t_i)$ and

$$d_i = y(t_i) = f(t_i) + e_i. \tag{1}$$

Noise (given that only the noise power is known) is given only through its probability density:

$$p(e_1, \ldots, e_N | \sigma, E) = \prod_{i=1}^{N} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{e_i^2}{2\sigma^2}\right). \tag{2}$$

Let H represent hypotheses regarding values of the parameters $\{\vartheta\}$. The H's are the objects of interest in this analysis. Our information concerning them is given, via Bayes' theorem, by

$$p(H|D, E) = \frac{p(H|E)\, p(D|H, E)}{p(D|E)}, \tag{3}$$

where E represents the prior information that we are able to include in the calculation. It is this expression which generates the algorithm that will be implimented in the integrated circuit. It is desirable to incorporate several prior probability distributions in the design of the chip.

We now substitute Eq. (1) into Eq. (2), $[e_i = d_i - f(t_i)]$: this represents a change of variables from the unknown $\{e_i\}$ to the unknown $\{\vartheta\}$. This will determine, therefore, $p(D|H, I)$. The result is

$$p(D|H, I) = p(D|\{\vartheta\}, \sigma, I) = \prod_{i=1}^{N} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{[d_i - f(t_i)]^2}{2\sigma^2}\right\} \tag{4}$$

$$= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^{N} [d_i - f(t_i)]^2\right\}. \tag{5}$$

The next step is to choose a model function. First, write

$$f(t) = \sum_{j=1}^{m} B_j G_j(t, \{\vartheta\}) \tag{6}$$

Then Eq. (5) becomes

$$p(D|\{B\}, \{\vartheta\}|\sigma, E) = \frac{\exp(-NQ/2\sigma^2)}{(2\pi\sigma^2)^{N/2}}, \tag{7}$$

where

$$Q = \overline{d^2} - \frac{2}{N}\sum_{j=1}^{m} B_j \sum_{i=1}^{N} d_i G_j(t_i) + \frac{1}{N}\sum_{j=1}^{m}\sum_{k=1}^{m} g_{ik} B_j B_k \tag{8}$$

$$\overline{d^2} = \frac{1}{N}\sum_{i=1}^{N} d_i^2 \tag{9}$$

$$g_{jk} = \sum_{i=1}^{N} G_j(t_i)\, G_k(t_i) \tag{10}$$

Now we diagonalize the matrix (g_{jk}) by introducing new functions $H_i(t)$ as follows:

$$f(t) = \sum_{i=1}^{m} A_i H_i(t)\ , \tag{11}$$

$$H_i(t) = \lambda_i^{-\frac{1}{2}} \sum_{j=1}^{m} e_{ij} G_j(t)\ , \tag{12}$$

$$\sum_{k=1}^{m} g_{jk} e_{lk} = \lambda_l e_{lk}\ ,\ \sum_{k=1}^{m} e_{lk} e_{lk} = 1\ , \tag{13}$$

$$\sum_{i=1}^{N} H_j(t_i)\, H_k(t_i) = \delta_{jk}\ . \tag{14}$$

Note that:

(1) (g_{jk}) is a function of $\{\vartheta\}$ and so is λ_j

(2) $B_k = \sum_{j=1}^{m} \frac{A_j e_{jk}}{\sqrt{\lambda_j}}\ ,\ A_k = \sqrt{\lambda_k} \sum_{j=1}^{m} B_j e_{kj}$ (15)

(3) $dB_1, \ldots, dB_m = (\lambda_1, \ldots, \lambda_m)^{-\frac{1}{2}}\, dA_1, \ldots, dA_m$ (16)

The Likelihood function in equation 7 becomes

$$p(D|\{A\},\{\vartheta\}|\sigma,E) = \left(2\pi\sigma^2\right)^{-\frac{N}{2}} \exp\left\{-\frac{N}{2\sigma^2}\left[\overline{d^2} - \frac{2}{N}\sum_{j=1}^{m} h_j A_j + \frac{1}{N}\sum_{j=1}^{m} A_j^2\right]\right\} \tag{17}$$

where

$$h_j = \sum_{i=1}^{N} d_i H_j(t_i)\,. \tag{18}$$

Finally, we introduce the chirped signal:

$$\begin{aligned} f(t) &= C\cos\left(\omega t + \alpha t^2 + \theta\right) \\ &= B_1 \cos\left(\omega t + \alpha t^2\right) + B_2 \sin\left(\omega t + \alpha t^2\right). \end{aligned} \tag{19}$$

So comparing with Eq. (6) we have

$$G_1(t) = \cos(\omega t + \alpha t^2) \tag{20a}$$

$$G_2(t) = \sin(\omega t + \alpha t^2) \tag{20b}$$

$$\{\vartheta\} = (\omega, \alpha). \tag{21}$$

The above material should give an indication of the nature of the calculations that the chip must be able to perform. Depending on the actual problem to be solved (which parameters are known, which ones are to be estimated and which are nuisance parameters), the specifics of the calculation will vary considerably and will go far beyond our summary — for a detailed analysis see Bretthorst [1987, 1988].

THE CHIP DESIGN

Preliminary design of an Application Specific Integrated Circuit (**ASIC**) to solve the above algorithm has been completed. The design was done on a Sun workstation with a silicon compiler, Concorde, from Seattle Silicon. This compiler allows one to easily enter the design as a schematic, and to simulate the design once it is done.

Our first design resembles Figure 1, which is a general design that is capable of using different algorithms. It was decided to design a chip with digital input, and to do the analog to digital conversion in another chip for two reasons: to simplify the first design, and to make it easy to work with computer simulated data. Simulations of this design indicate that it does do the mathematics that it was designed to do. A chip will be fabricated and placed on a circuit board, with the appropriate peripheral devices, which will be inserted in a small computer. This arrangement will allow easy simulation and display of the output of the processor.

The Random Access Memory (RAM), on the chip, stores the program, while the datapaths carry the data through a series of Arithmetic Logic Units (ALU's) that do the calculations. There are the equivalent of 238,000 transistors in this design.The chip should be capable of operating at 20 MHz, so signals up to 10MHz can be analyzed. Future versions of this chip, which will be faster, are under design. Parallel processors may be used to speed up the processing, or faster technology, such as Gallium Arsenide, may be used.

REFERENCES

Bretthorst, G.L. (1987), 'Bayesian Spectrum Analysis and Parameter Estimation,' Ph.D. Thesis, Washington University, St. Louis, Missouri. (Excerpts of this thesis are available in Erickson, G. J. and C. Ray Smith, eds. (1988), *Maximum-Entropy and Bayesian Methods in Science and Engineering.* **I.** *Foundations*, Kluwer Academic Publishers, Dordrecht.)

Bretthorst, G.L. (1988), 'Bayesian Spectrum Analysis and Parameter Estimation,'in *Lecture Notes in Statistics*, Springer–Verlag, Berlin.

Jaynes, E. T. (1987), 'Bayesian Spectrum and Chirp Analysis,' in Smith, C. Ray and G.J. Erickson, eds. *Maximum–Entropy and Bayesian Spectral Analysis and Estimation Problems*, D. Reidel, Dordrecht.

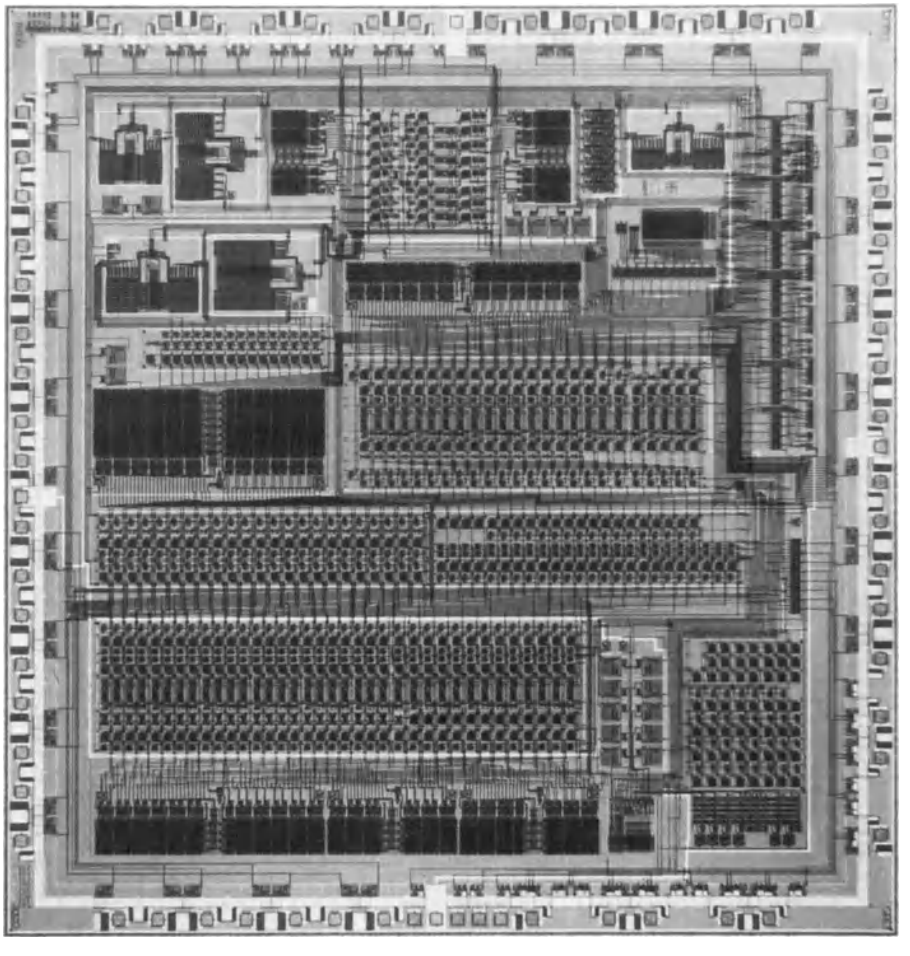


FIGURE 1.

Bayesian Data Analysis: Straight-line fitting

Stephen F. Gull
Cavendish Laboratory,
Madingley Road,
Cambridge CB3 0HE, U.K.

Abstract

A Bayesian solution is presented to the problem of straight-line fitting when both variables x and y are subject to error. The solution, which is fully symmetric with respect to x and y, contains a very surprising feature: it requires a informative prior for the distribution of sample positions. An uninformative prior leads to a bias in the estimated slope.

1. Introduction

An apparently simple data analysis problem that often arises is that of fitting a straight line to measurements of two quantities (x,y). Suppose that we have N such measurements $\{x_i,y_i\}$ and that they are each subject to independent Gaussian errors (σ_x,σ_y) (for the moment assumed known). Our task is to find a relationship of the form:

$$\hat{y} = a\hat{x} + b\ , \quad \text{where} \quad x_i = \hat{x}_i \pm \sigma_x\ ;\ y_i = \hat{y}_i \pm \sigma_y\ .$$

Note that we are considering a problem in which there is an underlying exact relation for the (unknown) quantities $(\hat{x},\hat{y})$ and that the measurements $\{x_i,y_i\}$ are subject to error. A related, but different problem is the case where there is very little experimental error, but the measurements refer to different objects with an intrinsic spread of (a,b) values. An example of this type would be the height and weight distributions of a set of students. Problems of this latter type are known as regression and, although they are clearly interesting, they are not the type of problem considered here.

2. The joint distribution

We now begin a careful Bayesian analysis of the straight-line-fitting problem, and will derive the joint probability distribution of the data and the parameters. For the case where both variables are subject to error we cannot avoid introducing the "hidden variables" $\{\hat{x}_i\}$ (and $\hat{y}_i=a\hat{x}_i+b$), which are a set of N nuisance parameters. We need these before we can even write down the likelihood function:

$$\mathrm{pr}(x,y\,|\,\hat{x},a,b,\sigma_x,\sigma_y) = (4\pi^2\sigma_x^2\sigma_y^2)^{-N/2}\ \exp -(\sum_i (x_i-\hat{x}_i)^2/\sigma_x^2 + \sum_i (y_i-\hat{y}_i)^2/\sigma_y^2)/2\ .$$

When we have completed the assignment of the joint p.d.f., we will integrate the nuisance parameters out of the posterior distribution.

To make further progress we need to refine our thinking about the nature of the problem. The variables x and y may not be of the same type, but it is

J. Skilling (ed.), Maximum Entropy and Bayesian Methods, 511–518.

usually as natural to plot x against y as y against x. We must therefore treat x and y in a symmetrical fashion. Recognising this, a sensible way to treat the problem is to suppose that there are separate scalings and offsets of the x and y variables that map them both into a given interval, for example (-1,+1). We define new scaled variables $\hat{X}$ and $\hat{Y}$, which are related as follows:

$$\hat{X} = (x-x_0) / R_x , \qquad \hat{Y} = (y-y_0) / R_y ,$$
$$a = R_y / R_x , \qquad b = y_0 - a\, x_0 .$$

This procedure closely follows what we do in practice when plotting points on graph-paper or on a display screen - we ascertain the range of both variables and plot accordingly. In this way our relationship takes the simple form:

$$\hat{X} = \pm\hat{Y} .$$

In what follows we will derive formulae appropriate for the positive sign. In order to cope with this ambiguity of the sign of the slope, we should, strictly, always compute both cases, and compare their posterior probabilities. In many cases it will be obvious which case is better. Two other extreme cases that might also be worth considering separately are the degenerate cases $\hat{X} = 0$ and $\hat{Y} = 0$.

At this point the reader may be forgiven for thinking that we have gone backwards; we started with two variables (a,b) and have replaced them by four (x_0,y_0,R_x,R_y). However, we will find that there are great advantages to be had from this more symmetrical formulation of the problem.

We start our development with the prior for $pr(x_0,y_0,R_x,R_y)$. Because the units of x and y are related to R_x and R_y, we can reasonably take R_x and R_y to be scale parameters, and the offsets x_0 and y_0 to be location parameters. We therefore take the prior as uniform in $\log R_x$, $\log R_y$, x_0 and y_0:

$$pr(x_0,y_0,R_x,R_y)\, dx_0\, dy_0\, dR_x\, dR_y \propto dx_0\, dy_0\, d(\log R_x)\, d(\log R_y) ,$$
$$\propto d(\log a)\, d(b a^{-1/2})\, d(\log R)\, d(x_0 a^{1/2}) ,$$

where $R = (R_x R_y)^{1/2}$ is a symmetric range parameter. We should also, for completeness, specify some sensible ranges for these parameters. In fact, the posterior distribution is normalisable over infinite ranges of x_0 and y_0 when there are more than two samples, and we shall return to the question of what (a_{min}, a_{max}) and (R_{min}, R_{max}) should be later.

The final expression for the prior in terms of our original variables a and b (and the range and offset of $\hat{X}$) is very instructive. In particular, the $(da\, db\, a^{-3/2})$ part of this prior can be compared to that obtained by Jaynes (1967) for an allegedly similar problem: he finds $(da\, db\, (a^2 + 1)^{-3/2})$, using a transformation group argument. Whilst I am always very wary of disagreeing with Ed, I note the following points.

1) the functional relationship derived by Jaynes:

$$a^3 f(a,b) = f(1/a, -b/a) \quad \text{(in the present notation)},$$

is satisfied by both candidate priors... and many others - this functional relationship is too weak to determine the prior uniquely.

2) The $(da\ db\ (a^2 + 1)^{-3/2})$ prior is the correct answer to a different problem. Suppose that (as in the Bertrand problem (Jaynes 1973)) a straw is thrown at random onto a piece of square graph paper. Imagine then that this straw defines a regression line. The rotational symmetry inherent in this second problem is now sufficient to determine the prior uniquely, but is not relevant to the straight-line fitting problem, even when the variables are of the same type. Indeed, for the particular example given by Jaynes, that of the daily temperature variations at New York and Boston (which is actually a regression problem, rather than line-fitting), it is rather difficult to understand why we should want to consider rotation of one axis onto the other.

For these reasons I believe that the hypothesis space defined here by (x_0, y_0, R_x, R_y) is more useful for the line-fitting problem than that implied by Jaynes' prior, but it was his prior (and the obvious non-uniqueness of the functional equation) that stimulated my interest in this problem.

We now arrive at a very interesting stage. The joint p.d.f. can be written as:

$$pr(x,y,\hat{x},x_0,y_0,R_x,R_y) = pr(x_0,y_0,R_x,R_y)\ pr(\hat{x}|x_0,y_0,R_x,R_y)\ pr(x,y|\hat{x},\sigma_x,\sigma_y),$$

where irrelevant conditionals have been dropped. Our remaining problem is the prior $pr(\hat{x}|x_0,y_0,R_x,R_y)$. At first sight it may seem peculiar that our answers are going to depend on our prior knowledge of the distribution of the "true" $\hat{x}$, and I imagine that strong objections will be voiced from some directions. However, my intuition about this matter has now been educated a little, and it is from this part of the prior that the most unexpected (and pleasing) feature of the Bayesian solution emerges. Let us take for this prior the independent Gaussian form:

$$pr(\hat{x}|x_0,y_0,R_x,R_y) = (2\pi R_x^2)^{-N/2} \exp -\sum_i((\hat{x}_i-x_0)^2/R_x^2)/2 .$$

This form can be derived by invoking the principle of Maximum Entropy, using constraints on $\langle\Sigma(\hat{x}-x_0)^2\rangle = N R_x^2$ and $\langle\Sigma\hat{x}\rangle = N x_0$. Note also that, because of the definition of the parameters, this prior is fully symmetric with respect to x and y. Perhaps the choice of a Gaussian prior for $\hat{x}$ and $\hat{y}$ does not really correspond to our best intuition for this problem; we might prefer to consider the points spread evenly over the graph paper. However, we shall continue to use a Gaussian prior, because it makes the algebra tractable, if not actually pleasant.

We now write down the full symmetric joint p.d.f.:

$$pr(x,y,\hat{x},x_0,y_0,\log R_x,\log R_y|\sigma_x,\sigma_y) =$$

$$(8\pi^3 R_x^2\sigma_x^2\sigma_y^2)^{-N/2} \exp -\sum_i ((\hat{x}_i-x_0)^2/R_x^2+(x_i-\hat{x}_i)^2/\sigma_x^2+(y_i-\hat{y}_i)^2/\sigma_y^2)/2 ,$$

which, using Bayes' theorem, is then proportional to the posterior distribution $pr(\hat{x},x_0,y_0,\log R_x,\log R_y|x,y,\sigma_x,\sigma_y)$.

3. Estimation of parameters

At this point we draw a polite veil over the algebra as we integrate the nuisance parameters $\hat{X}$ out of the problem. We note that the $\hat{X}$ have independent

Gaussian distributions which lead in turn to Gaussian distributions for x_0 and y_0:

$$\mathrm{pr}(x_0,y_0,\log R_x,\log R_y|x,y,\sigma_x,\sigma_y) = \int d^N\hat{x}\ \mathrm{pr}(\hat{x},x_0,y_0,\log R_x,\log R_y|x,y,\sigma_x,\sigma_y).$$

This yields estimators for x_0 and y_0, together with their covariance matrix:

$$\langle x_0\rangle = \sum_i x_i/N = \bar{x};\quad \langle y_0\rangle = \sum_i y_i/N = \bar{y};\quad \text{or}\quad \langle b\rangle = \bar{y} - a\,\bar{x}\ .$$

$$\langle \delta x_0^2\rangle = (\sigma_x^2 + aR^2)/N;\qquad \langle \delta x_0 \delta y_0\rangle = R^2/N;\qquad \langle \delta y_0^2\rangle = (\sigma_y^2 + a^{-1}R^2)/N,$$

and

$$\langle \delta b^2\rangle = (\sigma_y^2 + a^2\sigma_x^2)/N.$$

Note that the error estimates for x_0 and y_0 depend on the range parameter R, but that the error in the intercept value b depends only on the measurement errors σ_x and σ_y. We take this opportunity to integrate x_0 and y_0 out of the problem also, and to express the answer in terms of a and R. Finally, we find:

$$\log \mathrm{pr}(\log a,\log R|x,y) = \text{constant} - (N-1)/2\ \log(a\sigma_x^2R^2+\sigma_x^2\sigma_y^2+a^{-1}\sigma_y^2R^2)$$

$$- \frac{(V_{xx}(aR^2+\sigma_y^2)-2V_{xy}R^2+V_{yy}(a^{-1}R^2+\sigma_x^2))}{2\ (a\sigma_x^2R^2+\sigma_x^2\sigma_y^2+a^{-1}\sigma_y^2R^2)},$$

where the sample sum-squares are defined:

$$V_{xx} = \sum_i (x_i-\bar{x})^2;\qquad V_{xy} = \sum_i (x_i-\bar{x})(y_i-\bar{y});\qquad V_{yy} = \sum_i (y_i-\bar{y})^2.$$

There is little insight to be gained in developing this formula further analytically, but it is interesting to note its behaviour in certain limits. The estimated slope $\hat{a}$ is close to either V_{xy}/V_{xx} or V_{yy}/V_{xy}, depending on the relative sizes of σ_x and σ_y; its error is determined by the measurement errors, not the range parameter. The range parameter R is similarly determined by either $R_x^2 \sim V_{xx}/N$ or $R_y^2 \sim V_{yy}/N$, and its error $\delta\log R \sim N^{-1/2}$.

4. Discussion

We now illustrate this formula with a computer example. Figure 1 shows a dataset of 100 samples together with the best-fitted line. This looks to be a sensible fit, though we claim little credit for this in itself, because an equally good job can be done by eye. Figure 2 shows the posterior distribution of the interesting parameters R_x and R_y, confirming the presence of a single, well-defined maximum in the posterior p.d.f. We see also that there are certain problems of normalisation of the posterior distribution, because the p.d.f. tends to a constant value as $R_x \to 0$ and $R_y \to \infty$. As R_x or $R_y \to \infty$ the distribution falls off sufficiently fast to be integrable over an infinite range. This therefore raises again the question of a "sensible" cut-off for $R_{x\min}$ and $R_{y\min}$. We can answer the question of what a sensible cut-off means by investigating just what these cut-offs would have to be so that the contribution from the quadrant $(R_x,R_y) \to 0$ made a 50 per cent contribution to the posterior probability integral. For our dataset we find $R_{x\min}$ and $R_{y\min} < \exp(\exp(-1000))$. This is clearly a crazy number, and indicate that we are solving an essentially well-posed problem.

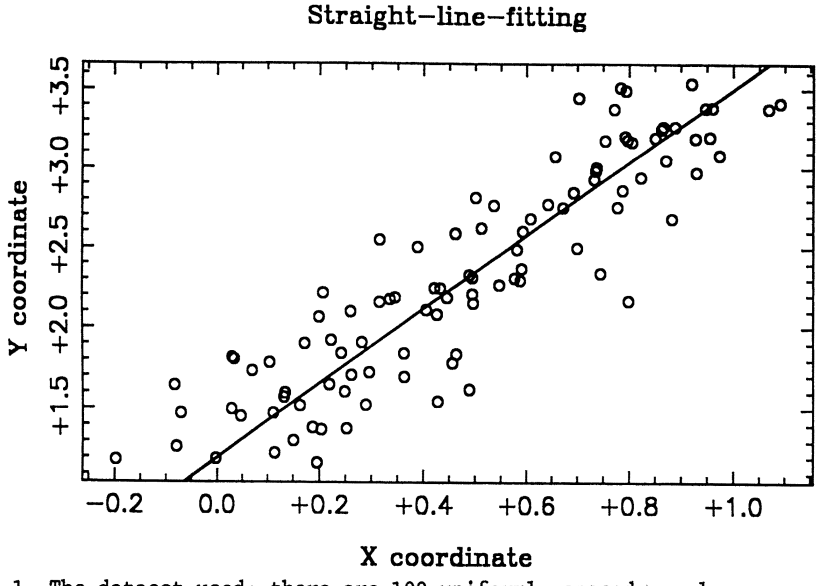


Figure 1. The dataset used: there are 100 uniformly-spaced samples.

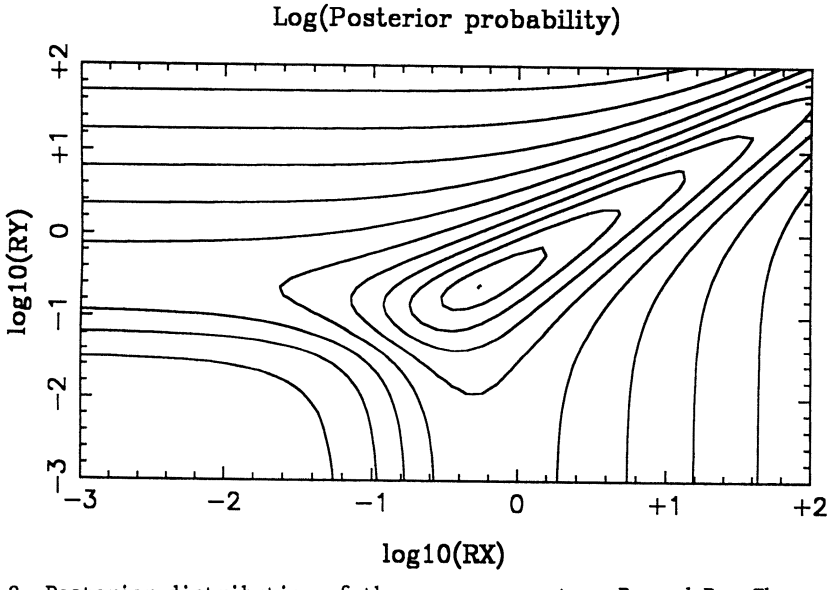


Figure 2. Posterior distribution of the range parameters R_x and R_y. The contour intervals are logarithmic, each level representing a probability difference of exp(100).

In other cases, though, we could well imagine that these numbers would not be so crazy, but instead give us insight into very real dilemmas. For example, suppose that the range of the data in one variable, say y, is very nearly covered by its error σ_y. This could easily happen, and implies $R_x \sim 0$. In this

case are we really so sure that there is any real variation of y present in the data? Only our prior probability of the range of R_{ymin} can help us here - R_{ymin} does matter. Of course, a change of prior for a, such as that suggested by Jaynes, can make this p.d.f. integrable, but at the cost of disguising what is an essential part of the problem. When the x variation is similar to σ_x, then R_{xmax} is also important.

Whilst the assignment of priors for a and b leads to interesting discussions, it does not actually affect the numerical estimates greatly. It is a bit like arguing about whether to use V/N or V/(N-1) when calculating standard deviations; the prior information becomes swamped as we gather more data. There are, however, much more important matters that are raised by our formula. The prior $pr(\hat{x}|x_0,y_0,R_x,R_y)$ is the most contentious part of the analysis, for the reason that there are so many nuisance parameters. We cannot swamp the $\hat{x}$ by gathering more data: we introduce a new $\hat{x}$ for each sample. We therefore have to be rather more careful about this prior.

Our first instinct, perhaps, would be to say "$\hat{x}$ is a location parameter" and assign to it a uniform prior over an infinite interval. There is no mathematical difficulty in this, indeed the analysis is far easier, and corresponds to our case R --> ∞. I freely admit that this was the first case that I tried, and I only abandoned it because it doesn't work. Indeed, if it had worked, then this analysis would have stayed in my research notebook as a trivial application of Bayesian methods. To see that the formula goes wrong, look at it in the limits $\sigma_x = 0$, $R = \infty$:

$$\log pr(a|x,y,\sigma_y) = \text{constant} - ((N-1)/2) \log a - ((a^2V_{xx} + aV_{xy})/\sigma_y^2)/2 \ .$$

The last term is fine, but the first term biasses the answer, increasing a by one-half of a standard deviation. But this term cannot just be dropped! We could get rid of it by re-formulating our hypothesis space in a different way, by dropping the symmetry with respect to x and y. But that in turn would exacerbate the problem for the complementary case $\sigma_y = 0$, where the present one-half standard deviation bias would be doubled.

All my Bayesian friends have objected at this point that "there's no such concept as bias in Bayesian analysis". It is true that there is no meaningful, exact definition of bias except in a frequentist sense. What I mean here is that the answer given by the R --> ∞ estimator is usually wrong, and in a given direction. The dictionary calls this "bias".

When a Bayesian calculation gives the wrong answer, it simply means that the hypothesis space contains wrong information. Here , we assembled the joint p.d.f. in a systematic way that I recommend be used in all Bayesian calculations (Gull 1988), so it is easy to see what went wrong. It was clear at the time that we needed the prior $pr(\hat{x}|x_0,y_0,R_x,R_y)$ for all the N samples simultaneously. We might swallow the "location parameter" argument for the first sample, but for N all at once it looks very strange. Suppose that the first (N-1) samples all lie in $-3 < x < 2.5$. Do we really believe that the next sample can be anywhere in $(-\infty, \infty)$? Our original Gaussian prior amounts to the reasonable suggestion that we learn about the mean and variance of the $\hat{x}$ distribution from the sample. We can of course do this, so that R is in general well-determined by the sample. Seen this way, we might it think it advisable to learn some other parameters of the shape of the $\hat{x}$ distribution as well. This will improve our results, but probably not very much, and at a terrible price: we would then be

unable to perform the required integrals analytically.

We can see now why the range parameter R corrects the bias of the estimated slope. Suppose, again, that $\sigma_x = 0$. The ((N-1)/2 log a) term from the determinant increases the slope by one-half of a standard deviation, but, as R is reduced, the $\hat{y}$ are gently squeezed in range, reducing the slope. When R reaches its most likely value the bias in the slope is exactly corrected. Seen the other way around: the range parameter biasses the slope against the weaker direction of the error bars; the determinant term corrects this. The formula given earlier seems to work for all combinations of σ_x and σ_y.

Can such a simple problem really require so complicated a solution? If all you want is the answer I can recommend an estimator for a:

$$\min \quad \frac{(a\ V_{xx} - 2\ V_{xy} + a^{-1}\ V_{yy})}{(a\ \sigma_x^2 + a^{-1}\ \sigma_y^2)} .$$

This is our answer with R --> ∞ and the determinant term dropped, so it will probably work. In can be derived by an ingenious argument (Brian Ripley, private communication, see also Ripley 1987 and Sprent 1969). The problem is scaled, not on the range of the data, but on the size of the errors σ_x and σ_y. The range itself is then allowed to go to infinity. If you scale on σ_x and σ_y, then there is no longer a 'weak' direction to be biassed, so no problems appear with the R --> ∞ solution. However, we note that a finite range $R^2 \sim V/N$ is still more likely than R = ∞, and because the problem is no longer scaled symmetrically on the range of the data, bias would return if R were reduced. In any case, scaling on the size of the errors looks a bit peculiar if either σ_x or σ_y is zero. Again, this modification of the hypothesis space seems to be an attempt to disguise what is a real problem. One is even led to speculate that scaling on σ_x and σ_y is a subconscious admission that some statisticians are more interested in the errors than they are in the data themselves!

Finally, we examine the problem of determining the level of the errors σ_x and σ_y if they are unknown. This does not involve any more analysis, because we have already been careful to retain all factors of σ_x and σ_y from the likelihood. We assign an uninformative prior for these variables, uniform in log σ_x and log σ_y, so that our previous formula will also be the posterior: $pr(\log\sigma_x, \log\sigma_y, \log a, \log R | x, y)$, which is illustrated in Figure 3. As we would expect, if only one of σ_x or σ_y is unknown, then the data determine the other extremely well, but it is too much to expect that <u>both</u> can be determined simultaneously. Rather, it is the combination $(a\sigma_x^2 + a^{-1}\sigma_y^2)$ that is accurately determined, but the error cannot be very reliably assigned to x or y individually. However, Figure 3 does show that there is just a little information about the ratio σ_y/σ_x contained in the dataset, presumably reflecting the fact that the $\hat{x}$ were uniformly sampled, rather than taken from a Gaussian distribution. Looking at the data by eye confirms this feeling; for a uniform distribution one can guess the relative contributions to the error. This indicates that there might be some real merit in using a more complicated hypothesis space, despite the difficulties of the computations involved.

5. Conclusions

The apparently simple Bayesian problem of straight-line fitting with both variables subject to error contains a subtle twist. The ranges of the variables

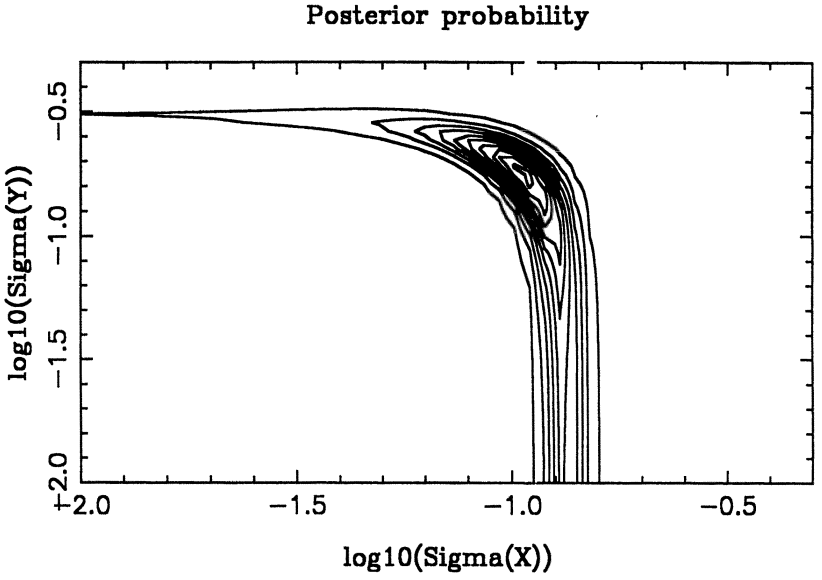


Figure 3. Posterior distribution of σ_x and σ_y. The contours are linear.

are usually well-determined by the dataset, equivalent to an "informative" prior for the sample positions. The use of a uniform, uninformative prior would lead to a bias in the estimated slope. The use of informative priors containing range parameters is a common feature of Bayesian analyses of this type: the "Classic" Maximum Entropy presented here by Skilling (1988) is another example.

Acknowledgments

I am grateful to John Skilling, Geoff Daniell and Brian Ripley for discussions about Bayesian regression.

References

Gull, S.F. (1988). Bayesian inductive inference and maximum entropy. In Maximum Entropy and Bayesian Methods in Science and Engineering, Vol. 1, ed. G.J. Erickson & C.R. Smith, pp 53-74. Kluwer, Dordrecht.

Jaynes, E.T. (1967). Reply to comments by Oscar Kempthorne, following Bayesian Intervals vs. Confidence Intervals. Reprinted in E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics, ed. R. Rosenkrantz (1983), pp190-209. Reidel, Dordrecht.

Jaynes, E.T. (1973). The well-posed problem. Reprinted in E.T. Jaynes: Papers on Probability, Statistics and Statistical Physics, ed. R. Rosenkrantz (1983), pp133-148. Reidel, Dordrecht.

Ripley, B.D. (1987). Regressian techniques for the detection of analytical bias, Analyst, 112, 377-383.

Skilling, J. (1988). Classic Maximum Entropy. In these Preceedings.

Sprent, P. (1969). Models in Regression and Related Topics. Methuen, London.

Index