on Decision and Control
Austin, Texas • December 1988

DYNAMICAL SYSTEMS THAT SORT LISTS, DIAGONALIZE MATRICES AND SOLVE LINEAR PROGRAMMING PROBLEMS

R. W. Brockett

Division of Applied Sciences
Harvard University
Cambridge, MA 02138

Abstract

We establish a number of properties associated with the dynamical system $\dot{H} = [H,[H,N]]$, where H and N are symmetric n by n matrices and $[A,B] = AB - BA$. The most important of these come from the fact that this equation is equivalent to a certain gradient flow on the space of orthogonal matrices. We are especially interested in the role of this equation as an analog computer. For example, we show how to map the data associated with a linear programming problem into $H(0)$ and N in such a way as to have $\dot{H} = [H[H,N]]$ evolve to a solution of the linear programming problem. This result can be applied to find systems which solve a variety of generic combinatorial optimization problems and it even provides an algorithm for diagonalizing symmetric matrices.

Introduction

The astounding resurgence in interest in analog computing which has come about recently under the banner of neural networks seems to be based on the hope that scientists and engineers will have more success using *analog* parallel computing than they have had using *digital* parallel computing. In engineering analog computation is usually used as a simulation tool; however in this paper we approach it as a means to solve optimization problems. By this we mean that one will associate with each optimization problem in a given class a description of f and an initial condition $x(0)$ such that $\dot{x} = f(x)$ evolves to a state $x(\infty)$ which characterizes the solution of the given problem. Of course the computational complexity associated with the identification of f and $x(0)$, together with the complexity of passing from $x(\infty)$ to the variables of interest, must be modest. Toward that end, in this paper we establish some results on the solution of linear programming problems with special emphasis on the question of sorting lists of real numbers by analog computation.

Our main result is that given appropriate choices for $H(0)$ and N the equation $\dot{H} = [H,[H,N]]$ can be used to solve certain standard problems in applied mathematics. From a mathematical point of view, this equation is a way of studying the gradient (steepest descent or steepest ascent) equation associated with functions of the form $\text{tr}(Q\Theta N\Theta^T)$, viewed as functions of Θ with Θ belonging to the orthogonal group. This approach is an outgrowth of the work on matching done in [1]. We are not aware of any other study of such gradient flows, however, the results we obtain make contact with a variety of interesting work including, (i) Schur-Horn theory (e.g. how the eigenvalues of a symmetric matrix relate to its diagonal entries) [2,3], (ii) Lax pairs and isospectral flows (the Hamiltonian theory of equations for the form $\dot{L} = [B, A(L)]$) [4,6], and (iii) von Neumann's results on the relationship between eigenvalues and singular values [5].

This work was supported in part by the U.S. Department of the Air Force under grant AFOSR-96-0197, in part by the U.S. Army Research Office under grant DAAL03-86-K-0171, and by the National Science Foundation under grant CDR-85-00108.

A Gradient Flow on Orthogonal Matrices

The results to be described later depend on properties of two closely related differential equations. These are matrix equations which describe a gradient flow; that is, they are differential equations evolving in a Riemannian manifold and are the counterparts of the familiar

$$\dot{x} = \nabla\phi$$

evolving in Euclidean n-space. Although from our point of view it looks accidental, the equations turn out to have properties in common with equations which have appeared in the theory of completely integrable Hamiltonian systems. In particular, although they are not the same, the equations are similar to the matrix version of the (finite) Toda lattice problem when studied from the "Lax pair" point of view. (See, for example, the work of Deift *et al.* [4] or the recent survey [6].)

Let $SO(n)$ denote the set of n by n orthogonal matrices and let $so(n)$ denote the set of n by n skew symmetric matrices. If Q and N are fixed n by n symmetric matrices, and if tr M denotes the sum of the diagonal entries of a square matrix M, then $\text{tr}(Q\Theta N\Theta^T)$ defines a smooth function on $SO(n)$. In [1] we show that the gradient flow corresponding to this function (using the natural Riemannian metric on $SO(n)$) is

$$\dot{\Theta} = \Theta N\Theta^T Q\Theta - Q\Theta N\,. \qquad (1)$$

For the sake of completeness we make this calculation in appendix 1. Using the fact that $\Theta^T\Theta = I$, equation 1 can also be expressed as

$$\dot{\Theta} = \Theta(N\Theta^T Q\Theta - \Theta^T Q\Theta N)$$

Because $(N\Theta^T Q\Theta - \Theta^T Q\Theta N)$ is skew symmetric, Θ will remain orthogonal if $\Theta(0)$ is orthogonal. Moreover, the values of Θ which make the right-hand side vanish are easily characterized. If we limit our attention to the case where the eigenvalues of Q and N are distinct, the following theorem describes the equilibrium states.

Theorem 1: Suppose that $Q = \Psi^T D_Q \Psi$ and $N = \Xi^T D_N \Xi$ with Ψ and Ξ orthogonal, $D_Q = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ and $D_N = \text{diag}(\mu_1, \mu_2, \dots, \mu_n)$. Assume that $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ and $\mu_1 > \mu_2 > \cdots > \mu_n$. Then the values of Θ such that $\Theta N\Theta^T Q\Theta - Q\Theta N$ vanish are the values of Θ such that

$$\tilde{\Theta} = \Psi\Theta\Xi^T$$

is the product ΠD of a permutation matrix Π and a diagonal square root of the identity D. That is, $D = \text{diag}(\pm1, \pm1, \dots, \pm1)$.

Proof (Compare with [1].): Without loss of generality we can suppose that Q and N are diagonal. Because the ij^{th} element of $NP - PN$ is $(n_{ii} - n_{jj})p_{ij}$ we see that $NP - PN$

vanishes only when P is diagonal. Thus $N\Theta^T Q\Theta - \Theta^T Q\Theta N$ vanishes only if $\Theta^T Q\Theta$ is diagonal. If Q is diagonal and if there is a diagonal matrix D such that $\Theta^T Q\Theta = D$, then D has the same eigenvalues as Q and $Q\Theta = \Theta D$. For this to be true we must have

$$\lambda_i\theta_{ij} = \lambda_{\pi(j)}\theta_{ij} \quad ; \quad 1 \le i,j \le n$$

for some permutation π of the indices $(1,2,\ldots,n)$ whereby $i \mapsto \pi(i)$. This means that either $\lambda_i = \lambda_{\pi(j)}$, so that $i = \pi(j)$, or that $\theta_{ij} = 0$. Fixing i, we see that $\pi(j)$ equals j for exactly one value of j and so only one θ_{ij} can be nonzero. Because the rows of θ are unit vectors, the nonzero value must be ± 1. From this all else follows.

Equation 1 evolves in the space of orthogonal matrices and is cubic in Θ. We now introduce a very useful change of variables which recasts this equation as one which evolves in the space symmetric matrices and which is quadratic. Let $H = \Theta^T Q\Theta$. Then

$$\begin{aligned}\dot{H} &= \Theta^T Q(\Theta N\Theta^T Q\Theta - Q\Theta N) + (\Theta^T Q\Theta N\Theta^T - N\Theta^T Q)Q\Theta \\ &= HNH - H^2N + HNH - NH^2 \\ &= [H,[N,H]] \qquad (2)\end{aligned}$$

where $[A,B] = AB - BA$ is the Lie bracket.

Because H satisfies $H = \Theta^T Q\Theta$ with Q constant, this last equation defines an isospectral flow; that is, as H evolves its entries change in such a way as to leave the eigenvalues of H invariant.

Theorem 2: Suppose that N is a real diagonal matrix with unrepeated eigenvalues. If $H(0)$ is symmetric and $\dot{H} = [H,[H,N]]$, then

$$\lim_{t\to\infty} H(t) = H(\infty)$$

exists and is a diagonal matrix.

Proof: First of all, because $[H,N]$ is skew-symmetric $[H[H,N]]$ is symmetric and this equation evolves in the space of symmetric matrices. A direct calculation shows that

$$\begin{aligned}\frac{d}{dt}\mathrm{tr}(HN) &= \mathrm{tr}N[H,[H,N]] \\ &= -\mathrm{tr}(HN - NH)^2\end{aligned}$$

Because $HN - NH$ is skew-symmetric, this is expressible as

$$\frac{d}{dt}\mathrm{tr}(HN) = \mathrm{tr}(HN - NH)(HN - NH)^T$$

Thus $\mathrm{tr}(HN)$ is monotone increasing and is bounded from above. Therefore, it has a limit and its derivative goes to zero. However, its time derivative vanishes only if H and N commute, and since N has distinct eigenvalues, $HN = NH$ if and only if H is diagonal. Thus H approaches a diagonal matrix.

It is important to note that in theorem 2, $H(0)$ (and hence $H(t)$ for all t) is permitted to have repeated eigenvalues. Of course $\dot{H} = [H,[H,N]]$ has many equilibrium points and theorem 2 claims only that its solution approaches one of them. The situation for the gradient flow on $SO(n)$ expressed in terms of Θ is a bit more complex and is described by the following theorem only in the case where both N and H have unrepeated eigenvalues.

Theorem 3: Suppose that Q and N have unrepeated eigenvalues. With the exception of certain initial points contained in a finite union of codimension-one submanifolds of $SO(n)$, all solutions of equation (1) approach a matrix of the form ΠD with D being a diagonal square root of I and Π being that permutation matrix which reorders the diagonal of Q in such a way as to maximize

$$\eta = \sum_{i=1}^{n} q_{\pi(i)\pi(i)} n_{ii}$$

The proof of this is given in [1].

In Hardy, Littlewood and Polya's classic book on inequalities [5] they study the problem of rearranging sequences so as to maximize the sum of the products such as those which arose above. The basic result is that the product is maximized when the sequences are "similarly ordered", e.g. when both are monotone decreasing. One point of view which makes this obvious, at least in the case where all the entries are positive, is to identify the sum of the products with the moment generated when one starts with a straight horizontal beam supported by a pivot and fastens weights of weight u_i to hooks positioned at distance y_i. It is intuitively clear that the way to get the largest moment is to attach the heaviest weight as far as possible from the pivot, the next heaviest at the next furthest, etc.

Theorem 4: If N is symmetric and n by n with distinct eigenvalues and if $H(0)$ is symmetric with distinct eigenvalues, then $\dot{H} = [H,[H,N]]$ has $n!$ equilibrium points. The eigenvalues of the linearization of $[H,[H,N]]$ at an equilibrium point take the form

$$(\lambda_{\pi(i)} - \lambda_{\pi(j)})(\mu_j - \mu_i) \quad ; \quad 1 \le i \le j \le n$$

for some permutation π. In particular if N is diagonal, all the eigenvalues of the linearization of $[H,[H,N]]$ at a point H_0 where H is diagonal have negative real parts if and only if the eigenvalues of H_0 and N are similarly ordered.

Proof: As in the proof of theorem 1 we can make a preliminary transformation to put N in diagonal form. From theorem 2 we see that H approaches a diagonal matrix which must be of the form

$$H = \mathrm{diag}(\lambda_{\pi(1)}, \lambda_{\pi(2)}, \ldots, \lambda_{\pi(n)}) .$$

Near the diagonal we can linearize $[H,[H,N]]$ as

$$\dot{h}_{ij} = (\lambda_{\pi(i)} - \lambda_{\pi(j)})(\mu_j - \mu_i)h_{ij}$$

for $i > j$.

We see from this theorem that exactly one of the $n!$ equilibrium points is asymptotically stable. The proof makes it clear that if either $H(0)$ or N has repeated eigenvalues, exponential stability (but not asymptotic stability) is lost.

In a 1937 paper von Neumann [5] studied the problem of maximizing $\mathrm{tr}(UAVB)$ for A and B fixed Hermetian matrices and U and V unitary matrices. He computed the maximum and investigated the Hessian at all the stationary points of $\mathrm{tr}(UAVB)$. He did not look at a gradient flow, but his critical point results are similar to those given here.

The Symmetric Eigenvalue Problem

The gradient flow, as expressed in terms of H, provides a means for diagonalizing a symmetric matrix. Techniques of this type have been investigated before, see, e.g., Deift *et al.* [4] and Chu [6]. The usual scheme involves a map Π which projects a symmetric matrix onto its strictly lower triangular part and/or a preliminary transformation to Hessenberg form. Our approach is, we believe, more elegant but it does involve the selection of a diagonal matrix N whose role is to guide the algorithm as to the order in which the eigenvalues are to appear on the diagonal. Figure 1 shows the evolution of the diagonalization process for a seven by seven symmetric matrix.

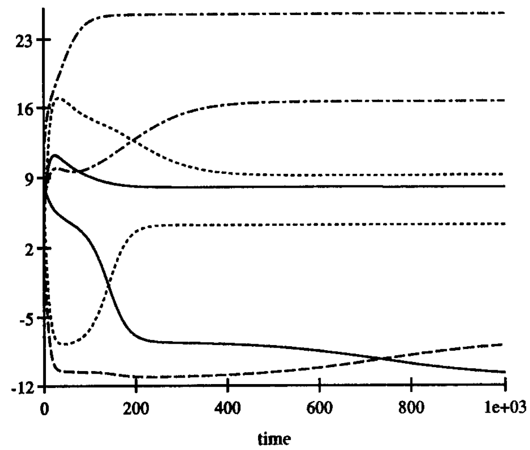

Figure 1: A plot of the diagonal elements of H as t evolves.

Linear Programming

A convex polytope is a bounded subset of $\mathbf{R}^n$ which has nonempty interior and which takes the form $X = \{x | Ax \leq b\}$. If we are given a vector c in $\mathbf{R}^n$ and asked to find $x \in X$ such that $\eta = \langle c, x\rangle$ is as large as possible, we have a linear programming problem. The simplex method is usually used to find a solution. Of course in recent years other methods have been proposed but a clear consensus as to the relative merits of these newer algorithms does not seem to have emerged. In this section we want to show how the descent equation introduced above can be used to solve linear programming problems when the constraint set is a polytope. This fails to be the general case by virtue of the assumption that X is bounded. Of course in any specific instance this could be circumvented.

Lemma 1: Given a convex polytope P in $\mathbf{R}^n$ having k vertices there exists a map $T : \mathbf{R}^k \to \mathbf{R}^n$ such that T maps the "standard simplex" in $\mathbf{R}^k$, i.e. the set

$$S = \{x | x \in \mathbf{R}^k \quad ; \quad x_i \geq 0 \quad ; \quad \Sigma x_i = 1\}$$

onto the given polytope.

Proof: Define T as

$$T = \begin{bmatrix} | & | & & | \\ a_1, & a_2, & \dots & a_k \\ | & | & & | \end{bmatrix}$$

with $a_1, a_2, \dots, a_k$ being the vertices of P in $\mathbf{R}^n$. Clearly T maps the extreme points of S in $\mathbf{R}^k$ onto extreme points in $\mathbf{R}^n$ and the remaining points of S serve to fill in the convex hull of P.

From this we can establish the following.

Theorem 5: Suppose that P is a convex polytope in $\mathbf{R}^n$ having k vertices. Then there exists a k by k symmetric matrix Q and a n by k matrix T such that P is the image under T of the convex hull of points generated by the permutations of the vector

$$v = \begin{bmatrix} \lambda_1(Q) \\ \lambda_2(Q) \\ \vdots \\ \lambda_k(Q) \end{bmatrix}$$

Proof: Let Q be a symmetric matrix with eigenvalues $(1, 0, 0, \dots, 0)$ and apply the preceding lemma.

The convex hull of the vector of eigenvalues plays a key role in Schur-Horn theory; see [2,3,9].

Using this lemma we can proceed to give a recipe to solve linear programming problems.

Theorem 6: Let X be a convex polytope in $\mathbf{R}^n$ with p extreme points. Suppose that we are to solve the linear programming problem consisting of maximizing $\langle c, x\rangle$ over $x \in X$. Let T be as in lemma 1. Then there exist diagonal matrices Q and N such that for almost all $\Theta \in SO(n)$,

$$\dot{H} = [H, [H, N]] \qquad ; \qquad H(0) = \Theta^T Q \Theta$$

converges to $H = \text{diag}(d_1, d_2, \dots, d_m)$ with the optimal x being given by Td.

Proof: Let $Q = \text{diag}(1, 0, 0, \dots, 0)$ and let $N = \text{diag}(\mu_1, \mu_2, \dots, \mu_n)$ with $\mu = cT$. The theorem then follows from theorems 2 and 4, the only further explanation being that we must choose Θ so that $H(0)$ avoids the thin sets associated with the stable manifolds of unstable critical points.

Since this choice of Q has repeated eigenvalues, the stable equilibrium of $\dot{H} = [H, [H, N]]$ is not exponentially stable. Thus although the way of generating the convex set X described in theorem 5 is mathematically convenient, it might not be the best for algorithmic purposes.

The Sorter

If we have a smooth dynamical system $\dot{x} = f(x, u)$; $y = h(x)$, with u being a vector whose components are fixed constants, and if y is to approach a sorted version of u, what can we say about the intrinsic complexity of the system which carries out this computation? We see, for example, that as u varies it must happen that the equilibrium value of x depends on u. The following is obvious but worth stating.

Lemma 2: No differentiable function of u sorts the components of u.

Proof: The function which sorts a list is continuous but not differentiable at those points in the domain for which two or more values of the components of u are equal.

Since the sorted rearrangement of u is not a smooth function of u and since we have set out to achieve the sorting by a smooth system, we see that we cannot achieve the desired result without inserting some dynamics. From theorem 2 it is clear that the differential equation $\dot{H} = [H, [H, N]]$ can be viewed as defining a sorting mechanism. Again the Hardy-Littlewood-Polya theory of rearrangement insures that if N is, for example, given by $\text{diag}(1, 2, \dots, n)$, then for almost all Θ

and for

$$H(0) = \Theta^T(\text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n))\Theta$$

the solution of $\dot{H} = [H, [H, N]]$ will approach

$$H(\infty) = \text{diag}(\lambda_{\pi(1)}, \lambda_{\pi(2)}, \dots, \lambda_{\pi(n)})$$

with the final list sorted by size.

In the equation $\dot{H} = [H, [H, N]]$, $H(0)$ and N play dual roles. Although we have viewed $H(0)$ as characterizing the list, we could reverse the process and let N characterize the list and let H generate the labels.

Rearranging Functions

In this section we briefly indicate an infinite dimensional version of $\dot{H} = [H, [H, N]]$ which rearranges functions. Although our solution seems to be new, there is some literature on this problem, see, e.g. [8].

Suppose that we are given two functions, $q(\cdot)$ and $p(\cdot)$, defined on [0,1]. Following Hardy, Littlewood and Polya [7] we say that they are *equi-measurable* if for all real numbers α the measure of the two sets

$$\begin{aligned} A(\alpha) &= \{x | q(x) \geq \alpha\} \\ B(\alpha) &= \{x | p(x) \geq \alpha\} \end{aligned}$$

is the same. In such a case one may, with some justification, refer to $q(\cdot)$ as being a *rearrangement* of $p(\cdot)$. It is intuitively clear that there is a rearrangement of any given function which is monotone decreasing. In view of the results of our earlier sections it is reasonable to seek an evolution equation which computes the monotone rearrangement.

Theorem 7: Suppose $p(\cdot) : [0,1] \to \mathbf{R}$ is a given, strictly monotone decreasing function. Suppose that

$$\phi(\cdot,\cdot,\cdot) : [0,\infty) \times [0,1] \times [0,1] \to \mathbf{R}\,.$$

satisfies the evolution equation

$$\frac{\partial \phi(t,x,y)}{\partial t} = \int_0^1 \phi(t,x,\xi)\phi(t,\xi,y)(p(x) + p(y) - 2p(\xi))d\xi$$

then in the limit as t goes to infinity, $(x-y)\phi(t,x,y)$ approaches zero and there exists a monotone decreasing function γ depending only on $\phi(0,\cdot,\cdot)$ such that for any smooth function $\psi(\cdot,\cdot)$

$$\lim_{t\to\infty} \int_0^1 \int_0^1 \phi(t,x,\xi)\psi(\xi,x)d\xi dx = \int_0^1 \gamma(x)\psi(x,x)dx$$

Proof: The proof follows from two calculations which are analogous to results developed above. In the first place we need to show that the flow is isospectral in the following sense.

For any smooth function $f : \mathbf{R}^1 \to \mathbf{R}^1$ we have

$$\frac{d}{dt}\int_0^1 f(\phi(t,x,x))dx = 0$$

This follows from the above formula for $\frac{\partial \phi}{\partial t}$ using the fact that, by reasons of symmetry,

$$\int_0^1 \int_0^1 f'\,\phi(t,x,\xi)\phi(t,\xi,x)(p(x) - p(\xi))dx d\xi = 0$$

Add to this the observation

$$\begin{aligned} &\tfrac{d}{dt}\textstyle\int_0^1 \phi(t,y,y)p(y)dy \\ &\quad = \textstyle\iint \phi(t,x,\xi)(p(\xi) - p(x))(\phi(t,\xi,x)p(\xi) - p(x))d\xi dx \\ &\quad = \textstyle\int_0^1\int_0^1 \phi(t,x,\xi)\cdot\phi(t,\xi,x)(p(\xi) - p(x))^2 d\xi dx \\ &\quad = \textstyle\int_0^1\int_0^1 \phi^2(t,x,\xi)(p(\xi) - p(x))^2 d\xi dx \end{aligned}$$

Because the integral is nonnegative, $\phi(t,x,\xi)(p(\xi) - p(x))$ must go to zero. Because p is strictly decreasing we see that $p(\xi) = p(x)$ if and only if $\xi = x$ and so $\phi(t,x,\xi)$ must approach zero for $\xi \neq x$.

Relationship with Realization Theory

Since system theoretic work done over the past 25 years on the realization problem is directed toward a deep and fundamental understanding of analog computation, it makes sense to approach the problem of what a neural net can compute from a system theoretic point of view. As studied in system theory, the realization problem is usually formulated as the problem of discovering a differential system which generates a given family of input/output pairs, with the inputs and outputs being defined on $[0,T]$ or $[0,\infty)$. However, in the case of digital electronics it is the equilibrium states which matter; how an equilibrium state is reached is usually unimportant. This suggests that it would be useful to formulate a modified realization problem in which only the steady state behavior is specified. In fact in an earlier paper [10] we studied such questions for the case in which the inputs are sine waves. Here we have discussed an analogous situation, that for which the inputs are constant values. The main interest here is in systems which compute discontinuous functions of the input, such as the maximum of the components of the input vector or a completely ordered version of the components of u.

Acknowledgments

The suggestions of Tony Bloch and K. S. Krishnaprasad are gratefully acknowledged.

References

[1] R. W. Brockett, "A Geometrical Matching Problem," (submitted for publication).

[2] A. Horn, "Doubly Stochastic Matrices and the Diagonal of a Rotation Matrix," *American J. of Math.*, Vol. 76, (1954), pp. 620-630.

[3] Bertram Kostant, "On Convexity, the Weyl Group, and the Iwasawa Decomposition," *Annales Scientifiques de L'ecole Normale Superieure*, 1973, pp. 413-455.

[4] P. Deift, T. Nanda, and C. Tomei, "Differential Equations for the Symmetric Eigenvalue Problem," *SIAM J. on Numerical Analysis*, Vol. 20, (1983), pp. 1-22.

[5] John von Neumann, "Some Matrix-Inequalities and Metrization of Matric-Spaces," *Tomsk Univ. Rev.*, Vol. 1, pp. 286-300, (1937). (See also *John von Neumann Collected Works*, Vol. IV, A. H. Taub, Ed., Pergamon Press, New York, pp. 205-218, 1962.)

[6] Moody T. Chu, "On the Continuous Realization of Interactive Processes," *SIAM Review*, Vol. 30, No. 3 (Sept. 1988), pp. 375-389.

[7] G. H. Hardy, J. E. Littlewood and G. Polya, *Inequalities*, 2nd edition, Cambridge University Press, Cambridge, England, 1952.

[8] Giorgio Talenti, "Assembling a Rearrangement," *Archive for Mechanics and Rational Analysis*, Vol. 98, pp. 285-293, (1987).

[9] Albert Marshall and Ingram Olkin, *Inequalities: Theory of Majorization and Its Applications*, Academic Press, New York, 1979.

[10] R. W. Brockett, "On Steady State Nonlinear Synthesis," *Proceedings of the 6th SICE Symposium on Control Theory*, May 1977, Hachioji, Japan.

Appendix

This appendix is devoted to the derivation of the gradient equation on the manifold of orthogonal matrices associated with the function $\mathrm{tr}(Q\Theta N\Theta^T)$. We parametrize a neighborhood of an orthogonal matrix Θ_0 by

$$\Theta(\Omega) = \Theta(I + \Omega + \Omega^2/2! \ldots)$$

with Ω skew symmetric. To first order in Ω we have

$$\begin{aligned}\mathrm{tr}(Q\Theta(I+\Omega)N(I-\Omega)\Theta^T \;&=\; \mathrm{tr}(Q\Theta N\Theta^T) + \mathrm{tr}(Q\Theta\Omega N\Theta^T)\\ &\quad - \mathrm{tr}(Q\Theta N\Omega\Theta^T)\end{aligned}$$

Using the fact that $\mathrm{tr}(ABC) = \mathrm{tr}(BCA) = \mathrm{tr}(CAB)$ we see that

$$\mathrm{tr}(Q\Theta(I+\Omega)N(I-\Omega)\Theta^T - \mathrm{tr}(Q\Theta N\Theta^T) = \mathrm{tr}((N\Theta^T Q\Theta - \Theta^T Q\Theta N)\Omega)$$

Define the matrix inner product as $\langle A, B\rangle = \mathrm{tr}(A^T B)$, (on $SO(n)$ this is proportional to the Killing form). We see that $\langle (N\Theta^T Q\Theta - \Theta^T QN\Theta)^T, \cdot\rangle$ represents the gradient at Θ. Using $\dot\Theta = \Theta\Omega$ we can express the gradient flow as

$$\Theta^T\dot\Theta = N\Theta^T Q\Theta - \Theta^T Q\Theta N$$

or

$$\dot\Theta = \Theta N\Theta^T Q\Theta - Q\Theta N\,.$$