CHAPTER 2

Double Bracket Isospectral Flows

2.1 Double Bracket Flows for Diagonalization

Brockett (1988) has introduced a new class of isospectral flows on the set of real symmetric matrices which have remarkable properties. He considers the ordinary differential equation

$$\boxed{\dot{H}(t) = [H(t), [H(t), N]], \qquad H(0) = H_0} \tag{1.1}$$

where $[A, B] = AB - BA$ denotes the Lie bracket for square matrices and N is an arbitrary real symmetric matrix. We term this the *double bracket* equation. Brockett proves that (1.1) defines an isospectral flow which, under suitable assumptions on N, diagonalizes any symmetric matrix $H(t)$ for $t \to \infty$. Also, he shows that the flow (1.1) can be used to solve various combinatorial optimization tasks such as linear programming problems and the sorting of lists of real numbers. Further applications to the travelling salesman problem and to digital quantization of continuous-time signals have been described, see Brockett (1989a; 1989b) and Brockett and Wong (1991). Note also the parallel efforts by Chu (1984b; 1984a), Driessel (1986), Chu and Driessel (1990) with applications to structured inverse eigenvalue problems and matrix least squares estimation.

The motivation for studying the double bracket equation (1.1) comes from the fact that it provides a solution to the following matrix least squares approximation problem.

Let $Q = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be a given real diagonal matrix and let

$$M(Q) = \{\Theta' Q \Theta \in \mathbb{R}^{n\times n} \mid \Theta\Theta' = I_n\} \tag{1.2}$$

denote the set of all real symmetric matrices $H = \Theta' Q \Theta$ orthogonally equivalent to Q. Thus $M(Q)$ is the set of all symmetric matrices with eigenvalues $\lambda_1, \dots, \lambda_n$. Given an arbitrary symmetric matrix $N \in \mathbb{R}^{n\times n}$, we consider the task of minimizing the matrix least squares index (distance)

$$\|N - H\|^2 = \|N\|^2 + \|H\|^2 - 2\operatorname{tr}(NH) \tag{1.3}$$

of N to any $H \in M(Q)$ for the Frobenius norm $\|A\|^2 = \operatorname{tr}(AA')$. This is a least squares estimation problem with spectral constraints. Since $\|H\|^2 = \sum_{i=1}^n \lambda_i^2$ is constant for $H \in M(Q)$, the minimization of (1.3) is equivalent to the maximization of the trace functional $\phi(H) = \operatorname{tr}(NH)$ defined on $M(Q)$.

Heuristically, if N is chosen to be diagonal we would expect that the matrices $H_* \in M(Q)$ which minimize (1.3) are of the same form, i.e. $H_* = \pi Q \pi'$ for a suitable $n \times n$ permutation matrix. Since $M(Q)$ is a smooth manifold (Proposition 1.1) it seems natural to apply steepest descent techniques in order to determine the minima H_* of the distance function (1.3) on $M(Q)$.

This program will be carried out in this section. The intuitive meaning of equation (1.1) will be explained by showing that it is actually the gradient flow of the distance function (1.3) on $M(Q)$. Since $M(Q)$ is a homogeneous space for the Lie group $O(n)$ of orthogonal matrices, we first digress to the topic of Lie groups and homogeneous spaces before starting our analysis of the flow (1.1).

Digression: Lie Groups and Homogeneous Spaces A *Lie group* G is a group, which is also a smooth manifold, such that

$$G \times G \to G, \qquad (x, y) \mapsto xy^{-1}$$

is a smooth map. Examples of Lie groups are

(a) The general linear group $GL(n, \mathbb{R}) = \{T \in \mathbb{R}^{n\times n} \mid \det(T) \neq 0\}$.

(b) The special linear group $SL(n, \mathbb{R}) = \{T \in \mathbb{R}^{n\times n} \mid \det(T) = 1\}$.

(c) The orthogonal group $O(n) = \{T \in \mathbb{R}^{n\times n} \mid TT' = I_n\}$.

(d) The unitary group $U(n) = \{T \in \mathbb{C}^{n\times n} \mid TT^* = I_n\}$.

The groups $O(n)$ and $U(n)$ are compact Lie groups, i.e. Lie groups which are compact spaces. Also, $GL(n, \mathbb{R})$ and $O(n)$ have two connected components while $SL(n, \mathbb{R})$ and $U(n)$ are connected.

The tangent space $\mathcal{G} := T_eG$ of a Lie group G at the identity element $e \in G$ carries in a natural way the structure of a *Lie algebra*. The Lie algebras of the above examples of Lie groups are

(a) $\mathrm{gl}(n,\mathbb{R}) = \{X \in \mathbb{R}^{n\times n}\}$

(b) $\mathrm{sl}(n,\mathbb{R}) = \{X \in \mathbb{R}^{n\times n} \mid \mathrm{tr}(X) = 0\}$

(c) $\mathrm{skew}(n,\mathbb{R}) = \{X \in \mathbb{R}^{n\times n} \mid X' = -X\}$

(d) $\mathrm{skew}(n,\mathbb{C}) = \{X \in \mathbb{C}^{n\times n} \mid X^* = -X\}$

In all of these cases the product structure on the Lie algebra is given by the *Lie bracket*

$$[X,Y] = XY - YX$$

of matrices X,Y.

A *Lie group action* of a Lie group G on a smooth manifold M is a smooth map

$$\sigma : G \times M \to M, \qquad (g,x) \mapsto g \cdot x$$

satisfying for all $g, h \in G$, and $x \in M$

$$g \cdot (h \cdot x) = (gh) \cdot x, \qquad e \cdot x = x.$$

The group action $\sigma : G \times M \to M$ is called *transitive* if there exists $x \in M$ such that every $y \in M$ satisfies $y = g \cdot x$ for some $g \in G$. A space M is called *homogeneous* if there exists a transitive G-action on M.

The *orbit* of $x \in M$ is defined by

$$\mathcal{O}(x) = \{g \cdot x \mid g \in G\}.$$

Thus the homogeneous spaces are the orbits of a group action. If G is a *compact* Lie group and $\sigma : G \times M \to M$ a Lie group action, then the orbits $\mathcal{O}(x)$ of σ are smooth, compact submanifolds of M.

Any Lie group action $\sigma : G \times M \to M$ induces an equivalence relation $\sim$ on M defined by

$$x \sim y \Longleftrightarrow \text{There exists} \quad g \in G \quad \text{with} \quad y = g \cdot x$$

for $x, y \in M$. Thus the equivalence classes of $\sim$ are the orbits of $\sigma : G \times M \to M$. The *orbit space* of $\sigma : G \times M \to M$, denoted by $M/G = M/\sim$, is defined as the set of all equivalence classes of M, i.e.

$$M/G := \{\mathcal{O}(x) \mid x \in M\}.$$

Here M/G carries a natural topology, the quotient topology, which is defined as the finest topology on M/G such that the quotient map

$$\pi : M \to M/G, \qquad \pi(x) := \mathcal{O}(x),$$

is continuous. Thus M/G is Hausdorff only if the orbits $\mathcal{O}(x)$, $x \in M$, are closed subsets of M.

Given a Lie group action $\sigma : G \times M \to M$ and a point $X \in M$, the *stabilizer subgroup* of x is defined as

$$\operatorname{Stab}(x) := \{g \in G \mid g \cdot x = x\}.$$

$\operatorname{Stab}(x)$ is a closed subgroup of G and is also a Lie group.

For any subgroup $H \subset G$ the orbit space of the H-action $\alpha : H \times G \to G$, $(h, g) \mapsto gh^{-1}$, is the set of coset classes

$$G/H = \{g \cdot H \mid g \in G\}$$

which is a homogeneous space. If H is a closed subgroup of G, then G/H is a smooth manifold. In particular, $G/\operatorname{Stab}(x)$, $x \in M$, is a smooth manifold for any Lie group action $\sigma : G \times M \to M$.

Consider the smooth map

$$\sigma_x : G \to M, \qquad g \mapsto g \cdot x,$$

for a given $x \in M$. Then the image $\sigma_x(G)$ of σ_x coincides with the G-orbit $\mathcal{O}(x)$ and σ_x induces a smooth bijection

$$\bar{\sigma}_x : G/\operatorname{Stab}(x) \to \mathcal{O}(x).$$

This map is a diffeomorphism if G is a compact Lie group. For arbitrary non-compact Lie groups G, the map $\bar{\sigma}_x$ need not be a homeomorphism. Note that the topologies of $G/\operatorname{Stab}(x)$ and $\mathcal{O}(x)$ are defined in a different way : $G/\operatorname{Stab}(x)$ is endowed with the quotient topology while $\mathcal{O}(x)$ carries the relative subspace topology induced from M. If G is compact, then these topologies are homeomorphic via $\bar{\sigma}_x$.

Returning to the study of the double bracket equation (1.1) as a gradient flow on the manifold $M(Q)$ of (1.2), let us first give a derivation of some elementary facts concerning the geometry of the set of symmetric matrices with fixed eigenvalues. Let

$$Q = \begin{bmatrix} \lambda_1 I_{n_1} & & 0 \\ & \ddots & \\ 0 & & \lambda_r I_{n_r} \end{bmatrix} \tag{1.4}$$

be a real diagonal $n \times n$ matrix with eigenvalues $\lambda_1 > \cdots > \lambda_r$ occurring with multiplicities $n_1, \ldots, n_r$, so that $n_1 + \cdots + n_r = n$.

Proposition 1.1 *$M(Q)$ is a smooth, connected, compact manifold of dimension*

$$\dim M(Q) = \frac{1}{2}\left(n^2 - \sum_{i=1}^{r} n_i^2\right).$$

Proof 1.2 The proof uses some elementary facts and the terminology from differential geometry which are summarized in the above digression on Lie groups and homogeneous spaces. Let $O(n)$ denote the compact Lie group of all orthogonal matrices $\Theta \in \mathbb{R}^{n\times n}$. We consider the smooth Lie group action $\sigma : O(n) \times \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ defined by $\sigma(\Theta, H) = \Theta H \Theta'$. Thus $M(Q)$ is an orbit of the group action σ and is therefore a compact manifold. The stabilizer subgroup $\operatorname{Stab}(Q) \subset O(n)$ of $Q \in \mathbb{R}^{n\times n}$ is defined as $\operatorname{Stab}(Q) = \{\Theta \in O(n) \mid \Theta Q \Theta' = Q\}$. Since $\Theta Q = Q\Theta$ if and only if $\Theta = \operatorname{diag}(\Theta_1, \dots, \Theta_r)$ with $\Theta_i \in O(n_i)$, we see that

$$\operatorname{Stab}(Q) = O(n_1) \times \cdots \times O(n_r) \subset O(n). \tag{1.5}$$

Therefore $M(Q) \cong O(n) / \operatorname{Stab}(Q)$ is diffeomorphic to the homogeneous space

$$M(Q) \cong O(n) / O(n_1) \times \cdots \times O(n_r), \tag{1.6}$$

and

$$\begin{aligned}
\dim M(Q) &= \dim O(n) - \sum_{i=1}^{r} \dim O(n_i) \\
&= \frac{n(n-1)}{2} - \frac{1}{2}\sum_{i=1}^{r} n_i(n_i - 1) \\
&= \frac{1}{2}\left(n^2 - \sum_{i=1}^{r} n_i^2\right).
\end{aligned}$$

To see the connectivity of $M(Q)$ note that the subgroup $SO(n) = \{\Theta \in O(n) \mid \det \Theta = 1\}$ of $O(n)$ is connected. Now $M(Q)$ is the image of $SO(n)$ under the continuous map $f : SO(n) \to M(Q)$, $f(\Theta) = \Theta Q \Theta'$, and therefore also connected. This completes the proof. ■

We need the following description of the tangent spaces of $M(Q)$.

Lemma 1.3 *The tangent space of $M(Q)$ at $H \in M(Q)$ is*

$$T_H M(Q) = \left\{[H, \Omega] \mid \Omega' = -\Omega \in \mathbb{R}^{n\times n}\right\}. \tag{1.7}$$

Proof 1.4 Consider the smooth map $\sigma_H : O(n) \to M(Q)$ defined by $\sigma_H(\Theta) = \Theta H \Theta'$. Note that σ_H is a submersion and therefore it induces a surjective map on tangent spaces. The tangent space of $O(n)$ at the $n \times n$ identity matrix I is $T_I O(n) = \{\Omega \in \mathbb{R}^{n\times n} \mid \Omega' = -\Omega\}$ (Appendix C) and the derivative of σ_H at I is the surjective linear map

$$\begin{aligned} D\sigma_H|_I : T_I O(n) \to & T_H M(Q), \\ \Omega \mapsto & \Omega H - H\Omega. \end{aligned} \tag{1.8}$$

This proves the result. ■

We now state and prove our main result on the double bracket equation (1.1). The result in this form is due to Bloch, Brockett and Ratiu (1990).

Theorem 1.5 *Let $N \in \mathbb{R}^{n\times n}$ be symmetric, and Q satisfy* (1.4).

(a) *The differential equation*(1.1),

$$\dot{H} = [H, [H, N]], \qquad H(0) = H'(0) = H_0,$$

defines an isospectral flow on the set of all symmetric matrices $H \in \mathbb{R}^{n\times n}$.

(b) *There exists a Riemannian metric on $M(Q)$ such that* (1.1) *is the gradient flow $\dot{H} = \operatorname{grad} f_N(H)$ of the function $f_N : M(Q) \to \mathbb{R}$, $f_N(H) = -\frac{1}{2}\|N - H\|^2$.*

(c) *The solutions of* (1.1) *exist for all $t \in \mathbb{R}$ and converge to a connected component of the set of equilibria points. The set of equilibria points H_∞ of* (1.1) *is characterized by $[H_\infty, N] = 0$, i.e. $NH_\infty = H_\infty N$.*

(d) *Let $N = \operatorname{diag}(\mu_1, \dots, \mu_n)$ with $\mu_1 > \dots > \mu_n$. Then every solution $H(t)$ of* (1.1) *converges for $t \to \pm\infty$ to a diagonal matrix $\pi \operatorname{diag}(\lambda_1, \dots, \lambda_n)\pi' = \operatorname{diag}(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)})$, where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of H_0 and π is a permutation matrix.*

(e) *Let $N = \operatorname{diag}(\mu_1, \dots, \mu_n)$ with $\mu_1 > \dots > \mu_n$ and define the function $f_N : M(Q) \to \mathbb{R}$, $f_N(H) = -\frac{1}{2}\|N - H\|^2$. The Hessian of $f_N : M(Q) \to \mathbb{R}$ at each critical point is nonsingular. For almost all initial conditions $H_0 \in M(Q)$ the solution $H(t)$ of* (1.1) *converges to Q as $t \to \infty$ with an exponential bound on the rate of convergence. For the case of distinct eigenvalues $\lambda_i \neq \lambda_j$, then the linearization at a critical point H_∞ is*

$$\dot{\xi}_{ij} = -\left(\lambda_{\pi(i)} - \lambda_{\pi(j)}\right)(\mu_i - \mu_j)\,\xi_{ij}, \qquad i > j, \tag{1.9}$$

where $H_\infty = \operatorname{diag}\left(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)}\right)$, for π a permutation matrix.

Proof 1.6 To prove (a) note that $[H, N] = HN - NH$ is skew-symmetric if H, N are symmetric. Thus the result follows immediately from Lemma 1.4.1. It follows that every solution $H(t)$, t belonging to an interval in $\mathbb{R}$, of (1.1) satisfies $H(t) \in M(H_0)$. By Proposition 1.1, $M(H_0)$ is compact and therefore $H(t)$ exists for all $t \in \mathbb{R}$. Consider the time derivative of the t function

$$f_N(H(t)) = -\tfrac{1}{2}\|N - H(t)\|^2 = -\tfrac{1}{2}\|N\|^2 - \tfrac{1}{2}\|H_0\|^2 + \operatorname{tr}(NH(t))$$

and note that

$$\frac{d}{dt} f_N(H(t)) = \operatorname{tr}\left(N\dot{H}(t)\right) = \operatorname{tr}(N[H(t), [H(t), N]]).$$

Since $\operatorname{tr}(A[B, C]) = \operatorname{tr}([A, B]C)$, and $[A, B] = -[B, A]$, then with A, B symmetric matrices, we have $[A, B]' = [B, A]$ and thus

$$\operatorname{tr}(N[H, [H, N]]) = \operatorname{tr}([N, H][H, N]) = \operatorname{tr}\left([H, N]'[H, N]\right).$$

Therefore

$$\frac{d}{dt} f_N(H(t)) = \|[H(t), N]\|^2. \tag{1.10}$$

Thus $f_N(H(t))$ increases monotonically. Since $f_N : M(H_0) \to \mathbb{R}$ is a continuous function on the compact space $M(H_0)$, then f_N is bounded from above and below. It follows that $f_N(H(t))$ converges to a finite value and its time derivative must go to zero. Therefore every solution of (1.1) converges to a connected component of the set of equilibria points as $t \to \infty$. By (1.10), the set of equilibria of (1.1) is characterized by $[N, H_\infty] = 0$. This proves (c). Now suppose that $N = \operatorname{diag}(\mu_1, \dots, \mu_n)$ with $\mu_1 > \dots > \mu_n$. By (c), the equilibria of (1.1) are the symmetric matrices $H_\infty = (h_{ij})$ which commute with N and thus

$$\mu_i h_{ij} = \mu_j h_{ij} \quad \text{for } i, j = 1, \dots, n. \tag{1.11}$$

Now (1.11) is equivalent to $(\mu_i - \mu_j) h_{ij} = 0$ for $i, j = 1, \dots, n$ and hence, by assumption on N, to $h_{ij} = 0$ for $i \neq j$. Thus H_∞ is a diagonal matrix with the same eigenvalues as H_0 and it follows that $H_\infty = \pi \operatorname{diag}(\lambda_1, \dots, \lambda_n) \pi'$. Therefore (1.1) has only a finite number of equilibrium points. We have already shown that every solution of (1.1) converges to a connected component of the set of equilibria and thus to a single equilibrium H_∞. This completes the proof of (d).

In order to prove (b) we need some preparation. For any $H \in M(Q)$ let $\sigma_H : O(n) \to M(Q)$ be the smooth map defined by $\sigma_H(\Theta) = \Theta' H \Theta$. The derivative at I_n is the surjective linear map

$$\begin{aligned} D\sigma_H|_I : T_I O(n) \to & T_H M(Q), \\ \Omega \mapsto & H\Omega - \Omega H. \end{aligned}$$

Let skew (n) denote the set of all real $n \times n$ skew-symmetric matrices. The kernel of $D\sigma_H|_I$ is then

$$K = \ker D\sigma_H|_I = \{\Omega \in \text{skew}(n) \mid H\Omega = \Omega H\}. \tag{1.12}$$

Let us endow the vector space skew (n) with the standard inner product defined by $(\Omega_1, \Omega_2) \mapsto \operatorname{tr}(\Omega_1'\Omega_2)$. The orthogonal complement of K in skew (n) then is

$$K^\perp = \{Z \in \text{skew}(n) \mid \operatorname{tr}(Z'\Omega) = 0 \quad \forall \Omega \in K\}. \tag{1.13}$$

Since

$$\operatorname{tr}\left([N,H]'\,\Omega\right) = -\operatorname{tr}(N[H,\Omega]) = 0$$

for all $\Omega \in K$, then $[N,H] \in K^\perp$ for all symmetric matrices N.

For any $\Omega \in$ skew (n) we have the unique decomposition

$$\Omega = \Omega_H + \Omega^H \tag{1.14}$$

with $\Omega_H \in K$ and $\Omega^H \in K^\perp$.

The gradient of a function on a smooth manifold M is only defined with respect to a fixed Riemannian metric on M (see Sections C.9 and C.10). To define the gradient of the distance function $f_N : M(Q) \to \mathbb{R}$ we therefore have to specify which Riemannian metric on $M(Q)$ we consider. Now to define a Riemannian metric on $M(Q)$ we have to define an inner product on each tangent space $T_H M(Q)$ of $M(Q)$. By Lemma 1.3, $D\sigma_H|_I$: skew $(n) \to T_H M(Q)$ is surjective with kernel K and hence induces an isomorphism of $K^\perp \subset$ skew (n) onto $T_H M(Q)$. Thus to define an inner product on $T_H M(Q)$ it is equivalent to define an inner product on $K^\perp$. We proceed as follows:

Define for tangent vectors $[H,\Omega_1], [H,\Omega_2] \in T_H M(Q)$

$$\langle\!\langle [H,\Omega_1], [H,\Omega_2] \rangle\!\rangle := \operatorname{tr}\left(\left(\Omega_1^H\right)' \Omega_2^H\right), \tag{1.15}$$

where Ω_i^H, $i = 1, 2$, are defined by (1.14). This defines an inner product on $T_H M(Q)$ and in fact a Riemannian metric on $M(Q)$. We refer to this as the normal Riemannian metric on $M(Q)$.

The gradient of $f_N : M(Q) \to \mathbb{R}$ with respect to this Riemannian metric is the unique vector field grad f_N on $M(Q)$ which satisfies the condition

(a) $$\operatorname{grad} f_N(H) \in T_H M(Q) \text{ for all } H \in M(Q)$$

(b) $$Df_N|_H([H,\Omega]) = \langle\!\langle \operatorname{grad} f_N(H), [H,\Omega] \rangle\!\rangle \quad \text{for all } [H,\Omega] \in T_H M(Q). \tag{1.16}$$

By Lemma 1.3, condition (1.16) implies that for all $H \in M(Q)$

$$\operatorname{grad} f_N(H) = [H, X] \tag{1.17}$$

for some skew-symmetric matrix X (which possibly depends on H). By computing the derivative of f_N at H we find

$$Df_N|_H([H,\Omega]) = \operatorname{tr}(N[H,\Omega]) = \operatorname{tr}\left([H,N]'\,\Omega\right).$$

Thus (1.16) implies

$$\begin{aligned} \operatorname{tr}\left([H,N]'\,\Omega\right) &= \langle\!\langle \operatorname{grad} f_N(H), [H,\Omega] \rangle\!\rangle \\ &= \langle\!\langle [H,X], [H,\Omega] \rangle\!\rangle \\ &= \operatorname{tr}\left(\left(X^H\right)'\,\Omega^H\right) \end{aligned} \tag{1.18}$$

for all $\Omega \in \operatorname{skew}(n)$.

Since $[H,N] \in K^\perp$ we have $[H,N] = [H,N]^H$ and therefore $\operatorname{tr}\left([H,N]'\,\Omega\right) = \operatorname{tr}\left([H,N]'\,\Omega^H\right)$. By (1.17) therefore

$$X^H = [H, N]. \tag{1.19}$$

This shows

$$\begin{aligned} \operatorname{grad} f_N(H) &= [H,X] = \left[H, X^H\right] \\ &= [H,[H,N]]. \end{aligned}$$

This completes the proof of (b).

For a proof that the Hessian of $f_N : M(Q) \to \mathbb{R}$ is nonsingular at each critical point, without assumption on the multiplicities of the eigenvalues of Q, we refer to Duistermaat et al. (1983). Here we only treat the generic case where $n_1 = \cdots = n_n = 1$, i.e. $Q = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$ with $\lambda_1 > \cdots > \lambda_n$.

The linearization of the flow (1.1) on $M(Q)$ around any equilibrium point $H_\infty \in M(Q)$ where $[H_\infty, N] = 0$ is given by

$$\dot{\xi} = H_\infty(\xi N - N\xi) - (\xi N - N\xi) H_\infty, \tag{1.20}$$

where $\xi \in T_{H_\infty} M(Q)$ is in the tangent space of $M(Q)$ at $H_\infty = \operatorname{diag}\left(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)}\right)$, for π a permutation matrix. By Lemma 1.3, $\xi =$

$[H_\infty, \Omega]$ holds for a skew-symmetric matrix Ω. Equivalently, in terms of the matrix entries, we have $\xi = (\xi_{ij})$, $\Omega = (\Omega_{ij})$ and $\xi_{ij} = \left(\lambda_{\pi(i)} - \lambda_{\pi(j)}\right) \Omega_{ij}$. Consequently, (1.20) is equivalent to the decoupled set of differential equations.

$$\dot{\xi}_{ij} = -\left(\lambda_{\pi(i)} - \lambda_{\pi(j)}\right) (\mu_i - \mu_j)\, \xi_{ij}$$

for $i, j = 1, \dots, n$. Since $\xi_{ij} = \xi_{ji}$ the tangent space $T_{H_\infty} M(Q)$ is parametrized by the coordinates ξ_{ij} for $i > j$. Therefore the eigenvalues of the linearization (1.20) are $-\left(\lambda_{\pi(i)} - \lambda_{\pi(j)}\right)(\mu_i - \mu_j)$ for $i > j$ which are all nonzero. Thus the Hessian is nonsingular. Similarly, $\pi = I_n$ is the unique permutation matrix for which $\left(\lambda_{\pi(i)} - \lambda_{\pi(j)}\right)(\mu_i - \mu_j) > 0$ for all $i > j$. Thus $H_\infty = Q$ is the unique critical point of $f_N : M(Q) \to \mathbb{R}$ for which the Hessian is negative definite.

The union of the unstable manifolds (see Section C.11) of the equilibria points $H_\infty = \operatorname{diag}\left(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)}\right)$ for $\pi \neq I_n$ forms a closed subset of $M(Q)$ of co-dimension at least one. (Actually, the co-dimension turns out to be exactly one). Thus the domain of attraction for the unique stable equilibrium $H_\infty = Q$ is an open and dense subset of $M(Q)$. This completes the proof of (e), and the theorem. ■

Remark 1.7 In the generic case, where the ordered eigenvalues λ_i, μ_i of H_0, N are distinct, then the property $\|N - H_0\|^2 \geq \|N - H_\infty\|^2$ leads to the *Wielandt-Hoffman inequality* $\|N - H_0\|^2 \geq \sum_{i=1}^{n} (\mu_i - \lambda_i)^2$. Since the eigenvalues of a matrix depend continuously on the coefficients, this last inequality holds in general, thus establishing the Wielandt-Hoffman inequality for all symmetric matrices N, H. □

Remark 1.8 In Part (c) of Theorem 1.5 it is stated that every solution of (1.1) converges to a connected component of the set of equilibria points. Actually Duistermaat et al. (1983) has shown that $f_N : M(Q) \to \mathbb{R}$ is always a Morse-Bott function. Thus, using Proposition 1.3.7, it follows that any solution of (1.1) is converging to a single equilibrium point rather than to a set of equilibrium points. □

Remark 1.9 If $Q = \left[\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right]$ is the standard rank k projection operator, then the distance function $f_N : M(Q) \to \mathbb{R}$ is a Morse-Bott function on the Grassmann manifold $\operatorname{Grass}_{\mathbb{R}}(k, n)$. This function then induces a Morse-Bott function on the Stiefel manifold $\operatorname{St}(k, n)$, which coincides with the Rayleigh quotient $R_N : \operatorname{St}(k, n) \to \mathbb{R}$. Therefore the Rayleigh quotient is seen as a Morse-Bott function on the Stiefel manifold. □

Remark 1.10 The rôle of the matrix N for the double bracket equation is that of a parameter which guides the equation to a desirable final state. In a diagonalization exercise one would choose N to be a diagonal matrix while the matrix to be diagonalized would enter as the initial condition of the double bracket flow. Then in the generic situation, the ordering of the diagonal entries of N will force the eigenvalues of H_0 to appear in the same order. □

Remark 1.11 It has been shown that the gradient flow of the least squares distance function $f_N : M(Q) \to \mathbb{R}$ with respect to the normal Riemannian metric is the double bracket flow. This is no longer true if other Riemannian metrices are considered. For example, the gradient of f_N with respect to the *induced* Riemannian metric is given by a very complicated formula in N and H which makes it hard to analyze. This is the main reason why the somewhat more complicated normal Riemannian metric is chosen. □

Flows on Orthogonal Matrices

The double bracket flow (1.1) provides a method to compute the eigenvalues of a symmetric matrix H_0. What about the eigenvectors of H_0? Following Brockett, we show that suitable gradient flows evolving on the group of orthogonal matrices converge to the various eigenvector basis of H_0.

Let $O(n)$ denote the Lie group of $n \times n$ real orthogonal matrices and let N and H_0 be $n \times n$ real symmetric matrices. We consider the smooth potential function

$$\phi : O(n) \to \mathbb{R}, \qquad \phi(\Theta) = \operatorname{tr}(N\Theta' H_0 \Theta) \tag{1.21}$$

defined on $O(n)$. Note that

$$\|N - \Theta' H_0 \Theta\|^2 = \|N\|^2 - 2\phi(\Theta) + \|H_0\|^2,$$

so that maximizing $\phi(\Theta)$ over $O(n)$ is equivalent to minimizing the least squares distances $\|N - \Theta' H_0 \Theta\|^2$ of N to $\Theta' H_0 \Theta \in M(H_0)$.

We need the following description of the tangent spaces of $O(n)$.

Lemma 1.12 *The tangent space of $O(n)$ at $\Theta \in O(n)$ is*

$$T_\Theta O(n) = \left\{ \Theta\Omega \mid \Omega' = -\Omega \in \mathbb{R}^{n\times n} \right\}. \tag{1.22}$$

Proof 1.13 Consider the smooth diffeomorphism $\lambda_\Theta : O(n) \to O(n)$ defined by left multiplication with Θ, i.e. $\lambda_\Theta(\psi) = \Theta\psi$ for $\psi \in O(n)$. The tangent space of $O(n)$ at the $n \times n$ identity matrix I is $T_I O(n) = \{\Omega \in \mathbb{R}^{n\times n} \mid \Omega' = -\Omega\}$; i.e. is equal to the Lie algebra of skew-symmetric matrices.

The derivative of λ_Θ at I is the linear isomorphism between tangent spaces

$$D(\lambda_\Theta)|_I : T_I O(n) \to T_\Theta O(n), \qquad \Omega \mapsto \Theta\Omega. \tag{1.23}$$

The result follows. ■

The standard Euclidean inner product $\langle A, B\rangle = \operatorname{tr}(A'B)$ on $\mathbb{R}^{n\times n}$ induces an inner product on each tangent space $T_\Theta O(n)$, defined by

$$\langle \Theta\Omega_1, \Theta\Omega_2\rangle = \operatorname{tr}\left((\Theta\Omega_1)'(\Theta\Omega_2)\right) = \operatorname{tr}(\Omega_1'\Omega_2). \tag{1.24}$$

This defines a Riemannian matrix on $O(n)$ to which we refer to as the *induced Riemannian metric* on $O(n)$. As an aside for readers with a background in Lie group theory, this metric coincides with the Killing form on the Lie algebra of $O(n)$, up to a constant scaling factor.

Theorem 1.14 *Let $N, H_0 \in \mathbb{R}^{n\times n}$ be symmetric. Then:*

(a) The differential equation

$$\boxed{\dot{\Theta}(t) = H_0\Theta(t)N - \Theta(t)N\Theta'(t)H_0\Theta(t), \qquad \Theta(0) \in O(n)} \tag{1.25}$$

induces a flow on the Lie group $O(n)$ of orthogonal matrices. Also (1.25) *is the gradient flow $\dot{\Theta} = \nabla\phi(\Theta)$ of the trace function* (1.21) *on $O(n)$ for the induced Riemannian metric on $O(n)$.*

(b) The solutions of (1.25) *exist for all $t \in \mathbb{R}$ and converge to a connected component of the set of equilibria points $\Theta_\infty \in O(n)$. The set of equilibria points Θ_∞ of* (1.25) *is characterized by $[N, \Theta_\infty' H_0 \Theta_\infty] = 0$.*

(c) Let $N = \operatorname{diag}(\mu_1, \dots, \mu_n)$ with $\mu_1 > \cdots > \mu_n$ and suppose H_0 has distinct eigenvalues $\lambda_1 > \cdots > \lambda_n$. Then every solution $\Theta(t)$ of (1.25) *converges for $t \to \infty$ to an orthogonal matrix Θ_∞ satisfying $H_0 = \Theta_\infty \operatorname{diag}\left(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)}\right)\Theta_\infty'$ for a permutation matrix π. The columns of Θ_∞ are eigenvectors of H_0.*

(d) Let $N = \operatorname{diag}(\mu_1, \dots, \mu_n)$ with $\mu_1 > \cdots > \mu_n$ and suppose H_0 has distinct eigenvalues $\lambda_1 > \cdots > \lambda_n$. The Hessian of $\phi : O(n) \to \mathbb{R}$ at each critical point is nonsingular. For almost all initial conditions $\Theta_0 \in O(n)$ the solution $\Theta(t)$ of (1.25) *converges exponentially fast to an eigenbasis Θ_∞ of H_0, such that $H_0 = \Theta_\infty \operatorname{diag}(\lambda_1, \dots, \lambda_n)\Theta_\infty'$ holds.*

Proof 1.15 The derivative of $\phi : O(n) \to \mathbb{R}$, $\phi(\Theta) = \operatorname{tr}(N\Theta' H_0 \Theta)$, at $\Theta \in O(n)$ is the linear map on the tangent space $T_\Theta O(n)$ defined by

$$\begin{aligned} D\phi|_\Theta(\Theta\Omega) &= \operatorname{tr}(N\Theta' H_0 \Theta\Omega - N\Omega\Theta' H_0 \Theta) \\ &= \operatorname{tr}([N, \Theta' H_0 \Theta]\Omega) \end{aligned} \tag{1.26}$$

for $\Omega' = -\Omega$. Let $\nabla\phi(\Theta)$ denote the gradient of ϕ at $\Theta \in O(n)$, defined with respect to the induced Riemannian metric on $O(n)$. Thus $\nabla\phi(\Theta)$ is characterized by

$$\text{(a)} \qquad \nabla\phi(\Theta) \in T_\Theta O(n),$$

$$\text{(b)} \qquad D\phi|_\phi(\Theta\Omega) = \langle\nabla\phi(\Theta), \Theta\Omega\rangle = \operatorname{tr}\left(\nabla\phi(\Theta)'\Theta\Omega\right), \tag{1.27}$$

for all skew-symmetric matrices $\Omega \in \mathbb{R}^{n\times n}$. By (1.26) and (1.27) then

$$\operatorname{tr}\left((\Theta'\nabla\phi(\Theta))'\Omega\right) = \operatorname{tr}([N, \Theta' H_0\Theta]\Omega) \tag{1.28}$$

for all $\Omega' = -\Omega$ and thus, by skew symmetry of $\Theta'\nabla\phi(\Theta)$ and $[N, \Theta' H_0\Theta]$, we have

$$\Theta'\nabla\phi(\Theta) = -[N, \Theta' H_0\Theta]. \tag{1.29}$$

Therefore (1.25) is the gradient flow of $\phi : O(n) \to \mathbb{R}$, proving (a).

By compactness of $O(n)$ and by the convergence properties of gradient flows, the solutions of (1.25) exist for all $t \in \mathbb{R}$ and converge to a connected component of the set of equilibria points $\Theta_\infty \in O(n)$, corresponding to a fixed value of the trace function $\phi : O(n) \to \mathbb{R}$. Also, $\Theta_\infty \in O(n)$ is an equilibrium print of (1.25) if and only if $\Theta_\infty[N, \Theta'_\infty H_0\Theta_\infty] = 0$, i.e. as Θ_∞ is invertible, if and only if $[N, \Theta'_\infty H_0\Theta_\infty] = 0$. This proves (b).

To prove (c) we need a lemma.

Lemma 1.16 *Let* $N = \operatorname{diag}(\mu_1, \ldots, \mu_n)$ *with* $\mu_1 > \cdots > \mu_n$. *Let* $\psi \in O(n)$ *with* $H_0 = \psi \operatorname{diag}(\lambda_1 I_{n_1}, \ldots, \lambda_r I_{n_r})\psi'$, $\lambda_1 > \cdots > \lambda_n$ *and* $\sum_{i=1}^r n_i = n$. *Then the set of equilibria points of* (1.25) *is characterized as*

$$\Theta_\infty = \psi D\pi \tag{1.30}$$

where $D = \operatorname{diag}(D_1, \ldots, D_r) \in O(n_1) \times \cdots \times O(n_r)$, $D_i \in O(n_i)$, $i = 1, \ldots, r$, *and* π *is an* $n \times n$ *permutation matrix.*

Proof 1.17 By (b), Θ_∞ is an equilibrium point if and only if $[N, \Theta'_\infty H_0\Theta_\infty] = 0$. As N has distinct diagonal entries this condition is equivalent to $\Theta'_\infty H_0\Theta_\infty$ being a diagonal matrix Λ. This is clearly true for matrices of the form (1.30). Now the diagonal elements of Λ are the eigenvalues of H_0 and therefore $\Lambda = \pi' \operatorname{diag}(\lambda_1 I_{n_1}, \ldots, \lambda_r I_{n_r})\pi$ for an $n \times n$ permutation matrix π. Thus

$$\begin{aligned}\operatorname{diag}(\lambda_1 I_{n_1}, \ldots, \lambda_r I_{n_r}) &= \pi\Theta'_\infty H_0\Theta_\infty\pi' \\ &= \pi\Theta'_\infty\psi \operatorname{diag}(\lambda_1 I_{n_1}, \ldots, \lambda_r I_{n_r})\psi'\Theta_\infty\pi'.\end{aligned} \tag{1.31}$$

Thus $\psi'\Theta_\infty\pi'$ commutes with $\operatorname{diag}(\lambda_1 I_{n_1}, \dots, \lambda_r I_{n_r})$. But any orthogonal matrix which commutes with $\operatorname{diag}(\lambda_1 I_{n_1}, \dots, \lambda_r I_{n_r})$, $\lambda_1 > \dots > \lambda_r$, is of the form $D = \operatorname{diag}(D_1, \dots, D_r)$ with $D_i \in O(n_i)$, $i = 1, \dots, r$, orthogonal. The result follows. ∎

It follows from this lemma and the genericity conditions on N and H_0 in (c) that the equilibria of (1.25) are of the form $\Theta_\infty = \psi D\pi$, where $D = \operatorname{diag}(\pm 1, \dots, \pm 1)$ is an arbitrary sign matrix and π is a permutation matrix. In particular there are exactly $2^n n!$ equilibrium points of (1.25). As this number is finite, part (b) implies that every solution $\Theta(t)$ converges to an element $\Theta_\infty \in \mathrm{O}(n)$ with $\Theta'_\infty H_0 \Theta_\infty = \pi' \operatorname{diag}(\lambda_1, \dots, \lambda_n)\pi = \operatorname{diag}(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)})$.

In particular, the column vectors of Θ_∞ are eigenvectors of H_0. This completes the proof of (c).

To prove (d) we linearize the flow (1.25) around each equilibrium point. The linearization of (1.25) at $\Theta_\infty = \psi D\pi$, where $D = \operatorname{diag}(d_1, \dots, d_n)$, $d_i \in \{\pm 1\}$, is

$$\dot{\xi} = -\Theta_\infty [N, \xi' H_0 \Theta_\infty + \Theta'_\infty H_0 \xi] \tag{1.32}$$

for $\xi = \Theta_\infty \Omega \in T_{\Theta_\infty} O(n)$. Thus (1.32) is equivalent to

$$\dot{\Omega} = -[N, [\Theta'_\infty H_0 \Theta_\infty, \Omega]] \tag{1.33}$$

on the linear space skew (n) of skew-symmetric $n \times n$ matrices Ω. As

$$\Theta'_\infty H_0 \Theta_\infty = \operatorname{diag}(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)}),$$

this is equivalent, in terms of the matrix entries of $\Omega = (\Omega_{ij})$, to the decoupled set of linear differential equations

$$\dot{\Omega}_{ij} = -\left(\lambda_{\pi(i)} - \lambda_{\pi(j)}\right)(\mu_i - \mu_j)\,\Omega_{ij}, \qquad i > j. \tag{1.34}$$

From this it follows immediately that the eigenvalues of the linearization at Θ_∞ are nonzero. Furthermore, they are all negative if and only if $(\lambda_{\pi(1)}, \dots, \lambda_{\pi(n)})$ and $(\mu_1, \dots, \mu_n)$ are similarly ordered, that is if and only if $\pi = I_n$. Arguing as in the proof of Theorem 1.5, the union of the unstable manifolds of the critical points $\Theta_\infty = \psi D\pi$ with $\pi \neq I_n$ is a closed subset of $\mathrm{O}(n)$ of co-dimension at least one. Its complement thus forms an open and dense subset of $\mathrm{O}(n)$. It is the union of the domains of attractions for the 2^n locally stable attractors $\Theta_\infty = \psi D$, with $D = \operatorname{diag}(\pm 1, \dots, \pm 1)$ arbitrary. This completes the proof of (d). ∎

Remark 1.18 In Part (b) of the above theorem it has been stated that every solution of (1.24) converges to a connected component of the set of equilibria points. Again, it can be shown that $\phi : O(n) \to \mathbb{R}$, $\phi(\Theta) = \operatorname{tr}(N\Theta' H_0 \Theta)$, is a Morse-Bott function. Thus, using Proposition 1.3.9, any solution of (1.24) is converging to an equilibrium point rather than a set of equilibria. □

It is easily verified that $\Theta(t)$ given from (1.25) implies that

$$H(t) = \Theta'(t) H_0 \Theta(t) \tag{1.35}$$

is a solution of the double bracket equation (1.1). In this sense, the double bracket equation is seen as a projected gradient flow from $O(n)$. A final observation is that in maximizing $\phi(\Theta) = \operatorname{tr}(N\Theta' H\Theta)$ over $O(n)$ in the case of generic matrices N, H_0 as in Theorem 1.14 Part (d), then of the $2^n n!$ possible equilibria $\Theta_\infty = \psi D\pi$, the 2^n maxima of $\phi : O(n) \to \mathbb{R}$ with $\pi = I_n$ maximize the sum of products of eigenvalues $\sum_{i=1}^n \lambda_{\pi(i)}\mu_i$. This ties in with a classical result that to maximize $\sum_{i=1}^n \lambda_{\pi(i)}\mu_i$ there must be a "similar" ordering; see Hardy, Littlewood and Polya (1952).

Problem 1.19 *For a nonsingular positive definite symmetric matrix A, show that the solutions of the equation $\dot{H} = [H, A[H,N]A]$ converges to the set of H_∞ satisfying $[H_\infty, N] = 0$. Explore the convergence properties in the case where A is diagonal.*

Problem 1.20 *Let $A, B \in \mathbb{R}^{n\times n}$ be symmetric. For any integer $i \in \mathbb{N}$ define $\operatorname{ad}_A^i(B)$ recursively by $\operatorname{ad}_A(B) := AB - BA$, $\operatorname{ad}_A^i(B) = \operatorname{ad}_A\left(\operatorname{ad}_A^{i-1}(B)\right)$, $i \geq 2$. Prove that $\operatorname{ad}_A^i(B) = 0$ for some $i \geq 1$ implies $\operatorname{ad}_A(B) = 0$. Deduce that for $i \geq 1$ arbitrary*

$$\dot{H} = \left[H, \operatorname{ad}_H^i(N)\right]$$

has the same equilibria as (1.1).

Problem 1.21 *Let $\operatorname{diag}(H) = \operatorname{diag}(h_{11}, \dots, h_{nn})$ denote the diagonal matrix with diagonal entries identical to those of H. Consider the problem of minimizing the distance function $g : M(Q) \to \mathbb{R}$ defined by $g(H) = \|H - \operatorname{diag}(H)\|^2$. Prove that the gradient flow of g with respect to the normal Riemannian metric is*

$$\dot{H} = [H, [H, \operatorname{diag}(H)]].$$

Show that the equilibrium points satisfy $[H_\infty, \operatorname{diag}(H_\infty)] = 0$.

Problem 1.22 *Let $N \in \mathbb{R}^{n\times n}$ be symmetric. Prove that the gradient flow of the function $H \mapsto \operatorname{tr}(NH^2)$ on $M(Q)$ with respect to the normal Riemannian metric is*

$$\dot{H} = \left[H, \left[H^2, N\right]\right]$$

Investigate the convergence properties of the flow. Generalize to $\operatorname{tr}(NH^m)$ *for* $m \in \mathbb{N}$ *arbitrary!*

Main Points of Section

An eigenvalue/eigenvector decomposition of a real symmetric matrix can be achieved by minimizing a matrix least squares distance function via a gradient flow on the Lie group of orthogonal matrices with an appropriate Riemannian metric. The distance function is a smooth Morse-Bott function.

The isospectral double bracket flow on homogeneous spaces of symmetric matrices converges to a diagonal matrix consisting of the eigenvalues of the initial condition. A specific choice of Riemannian metric allows a particularly simple form of the gradient. The convergence rate is exponential, with stability properties governed by a linearization of the equations around the critical points.

2.2 Toda Flows and the Riccati Equation

The Toda Flow

An important issue in numerical analysis is to exploit special structures of matrices to develop faster and more reliable algorithms. Thus, for example, in eigenvalue computations an initial matrix is often first transformed into Hessenberg form and the subsequent operations are performed on the set of Hessenberg matrices. In this section we study this issue for the double bracket equation. For appropriate choices of the parameter matrix N, the double bracket flow (1.1) restricts to a flow on certain subclasses of symmetric matrices. We treat only the case of symmetric Hessenberg matrices. These are termed *Jacobi matrices* and are banded, being tri-diagonal, symmetric matrices $H = (h_{ij})$ with $h_{ij} = 0$ for $|i - j| \geq 2$.

Lemma 2.1 *Let* $N = \operatorname{diag}(1, 2, \ldots, n)$ *and let* H *be a Jacobi matrix. Then* $[H, [H, N]]$ *is a Jacobi matrix, and* (1.1) *restricts to a flow on the set of isospectral Jacobi matrices.*

Proof 2.2 The (i, j)-th entry of $B = [H, [H, N]]$ is

$$b_{ij} = \sum_{k=1}^{n} (2k - i - j)\, h_{ik} h_{kj}.$$

Hence for $|i-j| \geq 3$ and $k = 1, \ldots, n$, then $h_{ik} = 0$ or $h_{kj} = 0$ and therefore $b_{ij} = 0$. Suppose $j = i + 2$. Then for $k = i + 1$

$$b_{ij} = (2(i+1) - i - j)\, h_{i,i+1} h_{i+1,i+2} = 0.$$

Similarly for $i = j + 2$. This completes the proof. ∎

Actually for $N = \operatorname{diag}(1, 2, \ldots, n)$ and H a Jacobi matrix

$$HN - NH = H_u - H_\ell, \tag{2.1}$$

where

$$H_u = \begin{bmatrix} h_{11} & h_{12} & & 0 \\ 0 & h_{22} & \ddots & \\ \vdots & \ddots & \ddots & h_{n-1,n} \\ 0 & \ldots & 0 & h_{nn}, \end{bmatrix}$$

and

$$H_\ell = \begin{bmatrix} h_{11} & 0 & \ldots & 0 \\ h_{21} & h_{22} & \ddots & \vdots \\ & \ddots & \ddots & 0 \\ 0 & & h_{n,n-1} & h_{nn} \end{bmatrix}$$

denote the upper triangular and lower triangular parts of H respectively. (This fails if H is not Jacobi; why?) Thus, for the above special choice of N, the double bracket flow induces the isospectral flow on Jacobi matrices

$$\boxed{\dot{H} = [H, H_u - H_\ell].} \tag{2.2}$$

The differential equation (2.2) is called the *Toda flow*. This connection between the double bracket flow (1.1) and the Toda flow has been first observed by Bloch (1990b). For a thorough study of the Toda lattice equation we refer to Kostant (1979). Flaschka (1974) and Moser (1975) have given the following Hamiltonian mechanics interpretation of (2.2). Consider the system of n idealized mass points $x_1 < \cdots < x_n$ on the real axis.

Let us suppose that the potential energy of this system of points $x_1, \ldots, x_n$ is given by

$$V(x_1, \ldots, x_n) = \sum_{k=1}^{n} e^{x_k - x_{k+1}}$$

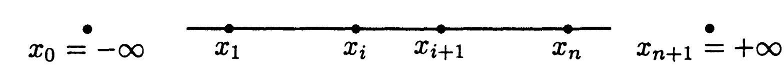

FIGURE 2.1. Mass points on the real axis

$x_{n+1} = +\infty$, while the kinetic energy is given as usual by $\frac{1}{2}\sum_{k=1}^{n} \dot{x}_k^2$. Thus the total energy of the system is given by the Hamiltonian

$$H(x,y) = \frac{1}{2}\sum_{k=1}^{n} y_k^2 + \sum_{k=1}^{n} e^{x_k - x_{k+1}}, \tag{2.3}$$

with $y_k = \dot{x}_k$ the momentum of the k-th mass point. Thus the associated Hamiltonian system is described by

$$\begin{aligned} \dot{x}_k &= \frac{\partial H}{\partial y_k} = y_k, \\ \dot{y}_k &= -\frac{\partial H}{\partial x_k} = e^{x_{k-1}-x_k} - e^{x_k - x_{k+1}}, \end{aligned} \tag{2.4}$$

for $k = 1, \dots, n$. In order to see the connection between (2.2) and (2.4) we introduce the new set of coordinates (this trick is due to Flaschka).

$$a_k = \tfrac{1}{2}e^{(x_k - x_{k+1})/2}, \qquad b_k = \tfrac{1}{2}y_k, \qquad k = 1, \dots, n. \tag{2.5}$$

Then with

$$H = \begin{bmatrix} b_1 & a_1 & & 0 \\ a_1 & b_2 & \ddots & \\ & \ddots & \ddots & a_{n-1} \\ 0 & & a_{n-1} & b_n \end{bmatrix},$$

the Jacobi matrix defined by a_k, b_k, it is easy to verify that (2.4) holds if and only if H satisfies the Toda lattice equation (2.2).

Let $\mathrm{Jac}(\lambda_1, \dots, \lambda_n)$ denote the set of $n \times n$ Jacobi matrices with eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$. The geometric structure of $\mathrm{Jac}(\lambda_1, \dots, \lambda_n)$ is rather complicated and not completely understood. However Tomei (1984) has shown that for distinct eigenvalues $\lambda_1 > \dots > \lambda_n$ the set $\mathrm{Jac}(\lambda_1, \dots, \lambda_n)$ is a smooth, $(n-1)$-dimensional compact manifold. Moreover, Tomei (1984) has determined the Euler characteristic of $\mathrm{Jac}(\lambda_1, \dots, \lambda_n)$. He shows that for $n = 3$ the isospectral set $\mathrm{Jac}(\lambda_1, \lambda_2, \lambda_3)$ for $\lambda_1 > \lambda_2 > \lambda_3$ is a compact Riemann surface of genus two. With these remarks in mind, the following result is an immediate consequence of Theorem 1.5.

Corollary 2.3

(a) *The Toda flow* (2.2) *is an isospectral flow on the set of real symmetric Jacobi matrices.*

(b) *The solution $H(t)$ of* (2.2) *exists for all $t \in \mathbb{R}$ and converges to a diagonal matrix as $t \to \pm\infty$.*

(c) *Let $N = \operatorname{diag}(1, 2, \dots, n)$ and $\lambda_1 > \cdots > \lambda_n$. The Toda flow on the isospectral manifold $\operatorname{Jac}(\lambda_1, \dots, \lambda_n)$ is the gradient flow for the least squares distance function $f_N : \operatorname{Jac}(\lambda_1, \dots, \lambda_n) \to \mathbb{R}$, $f_N(H) = -\frac{1}{2}\|N - H\|^2$.*

Remark 2.4 In Part (c) of the above Corollary, the underlying Riemannian metric on $\operatorname{Jac}(\lambda_1, \dots, \lambda_n)$ is just the restriction of the Riemannian metric $\langle\ ,\ \rangle$ appearing in Theorem 1.5 from $M(Q)$ to $\operatorname{Jac}(\lambda_1, \dots, \lambda_n)$. □

Connection with the Riccati equation

There is also an interesting connection of the double bracket equation (1.1) with the Riccati equation. For any real symmetric $n \times n$ matrix

$$Q = \operatorname{diag}(\lambda_1 I_{n_1}, \dots, \lambda_r I_{n_r}) \tag{2.6}$$

with distinct eigenvalues $\lambda_1 > \cdots > \lambda_r$ and multiplicities $n_1, \dots, n_r$, with $\sum_{i=1}^{r} n_i = n$, let

$$M(Q) = \{\Theta' Q \Theta \mid \Theta'\Theta = I_n\} \tag{2.7}$$

denote the isospectral manifold of all symmetric $n \times n$ matrices H which are orthogonally similar to Q. The geometry of $M(Q)$ is well understood, in fact, $M(Q)$ is known to be diffeomorphic to a flag manifold. By the isospectral property of the double bracket flow, it induces by restriction a flow on $M(Q)$ and therefore on a flag manifold. It seems difficult to analyze this induced flow in terms of the intrinsic geometry of the flag manifold; see however Duistermaat et al. (1983). We will therefore restrict ourselves to the simplest nontrivial case $r = 2$. But first let us digress on the topic of flag manifolds.

Digression: Flag Manifolds

A *flag* in $\mathbb{R}^n$ is an increasing sequence of subspaces $V_1 \subset \cdots \subset V_r \subset \mathbb{R}^n$ of $\mathbb{R}^n$, $1 \le r \le n$. Thus for $r = 1$ a flag is just a linear subspace of $\mathbb{R}^n$. Flags $V_1 \subset \cdots \subset V_n \subset \mathbb{R}^n$ with $\dim V_i = i$, $i = 1, \dots, n$, are called *complete*, and flags with $r < n$ are called *partial.*

Given any sequence $(n_1, \ldots, n_r)$ of nonnegative integers with $n_1 + \cdots + n_r \leq n$, the *flag manifold* $\operatorname{Flag}(n_1, \ldots, n_r)$ is defined as the set of all flags $(V_1, \ldots, V_r)$ of vector spaces with $V_1 \subset \cdots \subset V_r \subset \mathbb{R}^n$ and $\dim V_i = n_1 + \cdots + n_i$, $i = 1, \ldots, r$. $\operatorname{Flag}(n_1, \ldots, n_r)$ is a smooth, connected, compact manifold. For $r = 1$, $\operatorname{Flag}(n_1) = \operatorname{Grass}_{\mathbb{R}}(n_1, n)$ is just the Grassmann manifold of n_1-dimensional linear subspaces of $\mathbb{R}^n$. In particular, for $n = 2$ and $n_1 = 1$, $\operatorname{Flag}(1) = \mathbb{RP}^1$ is the real projective line and is thus homeomorphic to the circle $S^1 = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$.

Let $Q = \operatorname{diag}(\lambda_1 I_{n_1}, \ldots, \lambda_r I_{n_r})$ with $\lambda_1 > \cdots > \lambda_r$ and $n_1 + \cdots + n_r = n$ and let $M(Q)$ be defined by (1.2). Using (1.6) it can be shown that $M(Q)$ is diffeomorphic to the flag manifold $\operatorname{Flag}(n_1, \ldots, n_r)$. This *isospectral picture* of the flag manifold will be particularly useful to us. In particular for $r = 2$ we have a diffeomorphism of $M(Q)$ with the real Grassmann manifold $\operatorname{Grass}_{\mathbb{R}}(n_1, n)$ of n_1-dimensional linear subspaces of $\mathbb{R}^n$. This is in harmony with the fact, mentioned earlier in Chapter 1, that for $Q = \operatorname{diag}(I_k, 0)$ the real Grassmann manifold $\operatorname{Grass}_{\mathbb{R}}(k, n)$ is diffeomorphic to the set $M(Q)$ of rank k symmetric projection operators of $\mathbb{R}^n$.

Explicitly, to any orthogonal matrix $n \times n$ matrix

$$\Theta = \begin{bmatrix} \Theta_1 \\ \vdots \\ \Theta_r \end{bmatrix}$$

with $\Theta_i \in \mathbb{R}^{n_i \times n}$, $i = 1, \ldots, r$, and $n_1 + \cdots + n_r = n$, we associate the flag of vector spaces

$$V_\Theta := (V_1(\Theta), \ldots, V_r(\Theta)) \in \operatorname{Flag}(n_1, \ldots, n_r).$$

Here $V_i(\Theta)$, $i = 1, \ldots, r$, is defined as the $(n_1 + \cdots + n_i)$-dimensional vector space in $\mathbb{R}^n$ which is generated by the row vectors of the sub-matrix $[\Theta_1' \ldots \Theta_i']'$ This defines a map

$$\begin{aligned} f : M(Q) &\to \operatorname{Flag}(n_1, \ldots, n_r) \\ \Theta' Q \Theta &\mapsto V_\Theta. \end{aligned}$$

Note that, if $\Theta' Q \Theta = \hat{\Theta}' Q \hat{\Theta}$ then $\hat{\Theta} = \psi \Theta$ for an orthogonal matrix ψ which satisfies $\psi' Q \psi = Q$. But this implies that $\psi = \operatorname{diag}(\psi_1, \ldots, \psi_r)$ is block diagonal and therefore $V_{\hat{\Theta}} = V_\Theta$, so that f is well defined.

Conversely, $V_{\hat{\Theta}} = V_\Theta$ implies $\hat{\Theta} = \psi \Theta$ for an orthogonal matrix $\psi = \operatorname{diag}(\psi_1, \ldots, \psi_r)$. Thus the map f is well-defined and injective. It is easy to see that f is in fact a smooth bijection. The inverse of f is

$$\begin{aligned} f^{-1} : \operatorname{Flag}(n_1, \ldots, n_r) &\to M(Q) \\ V = (V_1, \ldots, V_r) &\mapsto \Theta_V' Q \Theta_V \end{aligned}$$

where

$$\Theta_V = \begin{bmatrix} \Theta_1 \\ \vdots \\ \Theta_r \end{bmatrix} \in O(n)$$

and Θ_i is any orthogonal basis of the orthogonal complement $V_{i-1}^{\perp} \cap V_i$ of V_{i-1} in V_i; $i = 1, \ldots, r$. It is then easy to check that f^{-1} is smooth and thus $f : M(Q) \to \mathrm{Flag}(n_1, \ldots, n_r)$ is a diffeomorphism.

With the above background material on flag manifolds, let us proceed with the connection of the double bracket equation to the Riccati equation. Let $\mathrm{Grass}_{\mathbb{R}}(k,n)$ denote the *Grassmann manifold* of k-dimensional linear subspaces of $\mathbb{R}^n$. Thus $\mathrm{Grass}_{\mathbb{R}}(k,n)$ is a compact manifold of dimension $k(n-k)$. For $k = 1$, $\mathrm{Grass}_{\mathbb{R}}(1,n) = \mathbb{RP}^{n-1}$ is the $(n-1)$-dimensional projective space of lines in $\mathbb{R}^n$ (see digression on projective spaces of Section 1.2).

For

$$Q = \begin{bmatrix} I_k & 0 \\ 0 & 0 \end{bmatrix},$$

$M(Q)$ coincides with the set of all rank k symmetric projection operators H of $\mathbb{R}^n$:

$$H' = H, \qquad H^2 = H, \qquad \operatorname{rank} H = k, \tag{2.8}$$

and we have already shown (see digression) that $M(Q)$ is diffeomorphic to the Grassmann manifold $\mathrm{Grass}_{\mathbb{R}}(k,n)$. The following result is a generalization of this observation.

Lemma 2.5 *For* $Q = \operatorname{diag}(\lambda_1 I_k, \lambda_2 I_{n-k})$, $\lambda_1 > \lambda_2$, *the isospectral manifold* $M(Q)$ *is diffeomorphic to the Grassmann manifold* $\mathrm{Grass}_{\mathbb{R}}(k,n)$.

Proof 2.6 To any orthogonal $n \times n$ matrix

$$\Theta = \begin{bmatrix} \Theta_1 \\ \Theta_2 \end{bmatrix}, \qquad \Theta_1 \in \mathbb{R}^{k \times n}, \qquad \Theta_2 \in \mathbb{R}^{(n-k) \times n},$$

we associate the k dimensional vector-space $V_{\Theta_1} \subset \mathbb{R}^n$, which is generated by the k orthogonal row vectors of Θ_1. This defines a map

$$\begin{aligned} f : M(Q) &\to \mathrm{Grass}_{\mathbb{R}}(k,n), \\ \Theta' Q \Theta &\mapsto V_{\Theta_1}. \end{aligned} \tag{2.9}$$

Note that, if $\Theta' Q \Theta = \widehat{\Theta}' Q \widehat{\Theta}$ then $\widehat{\Theta} = \psi \Theta$ for an orthogonal matrix ψ which satisfies $\psi' Q \psi = Q$. But this implies that $\psi = \operatorname{diag}(\psi_1, \psi_2)$ and therefore $V_{\widehat{\Theta}_1} = V_{\Theta_1}$ and f is well defined. Conversely, $V_{\widehat{\Theta}_1} = V_{\Theta_1}$ implies $\widehat{\Theta} = \psi \Theta$ for an orthogonal matrix $\psi = \operatorname{diag}(\psi_1, \psi_2)$. Thus (2.9) is injective. It is easy to see that f is a bijection and a diffeomorphism. In fact, as $M(Q)$ is the set of rank k symmetric projection operators, $f(H) \in \operatorname{Grass}_{\mathbb{R}}(k,n)$ is the image of $H \in M(Q)$. Conversely let $X \in \mathbb{R}^{n \times k}$ be such that the columns of X generate a k-dimensional linear subspace $V \subset \mathbb{R}^n$. Then

$$P_X := X (X'X)^{-1} X'$$

is the Hermitian projection operator onto V and

$$\begin{aligned} f^{-1} : \operatorname{Grass}_{\mathbb{R}}(k,n) &\to M(Q) \\ \text{column span}(X) &\mapsto X (X'X)^{-1} X' \end{aligned}$$

is the inverse of f. It is obviously smooth and a diffeomorphism. ■

Every $n \times n$ matrix $N \in \mathbb{R}^{n \times n}$ induces a flow on the Grassmann manifold

$$\Phi_N : \mathbb{R} \times \operatorname{Grass}_{\mathbb{R}}(k,n) \to \operatorname{Grass}_{\mathbb{R}}(k,n)$$

defined by

$$\Phi_N(t,V) = e^{tN} \cdot V, \tag{2.10}$$

where $e^{tN} \cdot V$ denotes the image of the k-dimensional subspace $V \subset \mathbb{R}^n$ under the invertible linear transformation $e^{tN} : \mathbb{R}^n \to \mathbb{R}^n$. We refer to (2.10) as the *flow* on $\operatorname{Grass}_{\mathbb{R}}(k,n)$ which is *linearly induced by* N.

We have already seen in Section 1.2, that linearly induced flows on the Grassmann manifold $\operatorname{Grass}_{\mathbb{R}}(k,n)$ correspond to the *matrix Riccati equation*

$$\dot{K} = A_{21} + A_{22}K - KA_{11} - KA_{12}K \tag{2.11}$$

on $\mathbb{R}^{(n-k)\times k}$, where A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Theorem 2.7 *Let $N \in \mathbb{R}^{n \times n}$ be symmetric and $Q = \operatorname{diag}(\lambda_1 I_k, \lambda_2 I_{n-k})$, $\lambda_1 > \lambda_2$, for $1 \leq k \leq n-1$. The double bracket flow* (1.1) *is equivalent via the map f, defined by* (2.9), *to the flow on the Grassmann manifold* $\operatorname{Grass}_{\mathbb{R}}(k,n)$ *linearly induced by $(\lambda_1 - \lambda_2) N$. It thus induces the Riccati equation* (2.11) *with generating matrix $A = (\lambda_1 - \lambda_2) N$.*

Proof 2.8 Let $H_0 = \Theta_0' Q \Theta_0 \in M(Q)$ and let $H(t) \in M(Q)$ be a solution of (1.1) with $H(0) = H_0$. We have to show that for all $t \in \mathbb{R}$

$$f(H(t)) = e^{t(\lambda_1 - \lambda_2)N} V_0, \tag{2.12}$$

where f is defined by (2.9) and $V_0 = V_{\Theta_0}$. By Theorem 1.14, $H(t) = \Theta(t)' Q \Theta(t)$ where $\Theta(t)$ satisfies the gradient flow on $O(n)$

$$\dot{\Theta} = \Theta(\Theta' Q \Theta N - N \Theta' Q \Theta). \tag{2.13}$$

Hence for $X = \Theta'$

$$\dot{X} = NXQ - XQX'NX. \tag{2.14}$$

Let $X(t)$, $t \in \mathbb{R}$, be any orthogonal matrix solution of (2.14). (Note that orthogonality of $X(t)$ holds automatically in case of orthogonal initial conditions.) Let $S(t) \in \mathbb{R}^{n \times n}$, $S(0) = I_n$, be the unique matrix solution of the linear time-varying system

$$\dot{S} = X(t)' NX(t)((\lambda_1 - \lambda_2) S - QS) + QX(t)' NX(t) S. \tag{2.15}$$

Let $S_{ij}(t)$ denote the (i,j)-block entry of $S(t)$. Suppose $S_{21}(t_0) = 0$ for some $t_0 \in \mathbb{R}$. A straightforward computation using (2.15) shows that then also $\dot{S}_{21}(t_0) = 0$. Therefore (2.15) restricts to a time-varying flow on the subset of block upper triangular matrices. In particular, the solution $S(t)$ with $S(0) = I_n$ is block upper triangular for all $t \in \mathbb{R}$

$$S(t) = \begin{bmatrix} S_{11}(t) & S_{12}(t) \\ 0 & S_{22}(t) \end{bmatrix}.$$

Lemma 2.9 *For any solution $X(t)$ of (2.14) let*

$$S = \begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix}$$

be the solution of (2.15) with $S(0) = I_n$. Then

$$X(t) \cdot S(t) = e^{t(\lambda_1 - \lambda_2)N} \cdot X(0). \tag{2.16}$$

Proof 2.10 Let $Y(t) = X(t) \cdot S(t)$. Then

$$\begin{aligned} \dot{Y} &= \dot{X}S + X\dot{S} \\ &= NXQS - XQX'NXS + NX((\lambda_1 - \lambda_2) S - QS) + XQX'NXS \\ &= (\lambda_1 - \lambda_2) NY. \end{aligned}$$

Let $Z(t) = e^{t(\lambda_1 - \lambda_2)N} \cdot X(0)$. Now Y and Z both satisfy the linear differential equation $\dot{\xi} = (\lambda_1 - \lambda_2) N\xi$ with identical initial condition $Y(0) = Z(0) = X(0)$. Thus $Y(t) = Z(t)$ for all $t \in \mathbb{R}$ and the lemma is proved. ■

We now have the proof of Theorem 2.7 in our hands. In fact, let $V(t) = f(H(t)) \in \mathrm{Grass}_{\mathbb{R}}(k,n)$ denote the vector space which is generated by the first k column vectors of $X(t)$. By the above lemma

$$V(t) = e^{t(\lambda_1-\lambda_2)N} V(0)$$

which completes the proof. ∎

One can use Theorem 2.7 to prove results on the dynamic Riccati equation arising in linear optimal control, see Anderson and Moore (1990). Let

$$N = \begin{bmatrix} A & -BB' \\ -C'C & -A' \end{bmatrix} \tag{2.17}$$

be the Hamiltonian matrix associated with a linear system

$$\dot{x} = Ax + Bu, \qquad y = Cx \tag{2.18}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$ and $C \in \mathbb{R}^{p\times n}$. If $m = p$ and $(A,B,C) = (A',C',B')$ then N is symmetric and we can apply Theorem 2.7.

Corollary 2.11 *Let $(A,B,C) = (A',C',B')$ be a symmetric controllable and observable realization. The Riccati equation*

$$\boxed{\dot{K} = -KA - A'K + KBB'K - C'C} \tag{2.19}$$

extends to a gradient flow on $\mathrm{Grass}_{\mathbb{R}}(n,2n)$ *given by the double bracket equation* (1.1) *under* (2.19). *Every solution in* $\mathrm{Grass}_{\mathbb{R}}(n,2n)$ *converges to an equilibrium point. Suppose*

$$N = \begin{bmatrix} A & -BB' \\ -C'C & -A' \end{bmatrix}$$

has distinct eigenvalues. Then (2.19) *has* $\binom{2n}{n}$ *equilibrium points in the Grassmannian* $\mathrm{Grass}_{\mathbb{R}}(n,2n)$, *exactly one of which is asymptotically stable.*

Proof 2.12 By Theorem 2.7 the double bracket flow on $M(Q)$ for $Q = \left[\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right]$ is equivalent to the flow on the Grassmannian $\mathrm{Grass}_{\mathbb{R}}(n,2n)$ which is linearly induced by N. If N has distinct eigenvalues, then the double bracket flow on $M(Q)$ has $\binom{2n}{n}$ equilibrium points with exactly one being asymptotically stable. Thus the result follows immediately from Theorem 1.5. ∎

Remark 2.13 A transfer function $G(s) = C(sI - A)^{-1} B$ has a symmetric realization $(A, B, C) = (A', C', B')$ if and only if $G(s) = G(s)'$ and the Cauchy index of $G(s)$ is equal to the McMillan degree; Youla and Tissi (1966), Brockett and Skoog (1971). Such transfer functions arise frequently in circuit theory. □

Remark 2.14 Of course the second part of the theorem is a well-known fact from linear optimal control theory. The proof here, based on the properties of the double bracket flow, however, is new and offers a rather different approach to the stability properties of the Riccati equation. □

Remark 2.15 A possible point of confusion arising here might be that in some cases the solutions of the Riccati equation (2.19) might not exist for all $t \in \mathbb{R}$ (finite escape time) while the solutions to the double bracket equation always exist for all $t \in \mathbb{R}$. One should keep in mind that Theorem 2.7 only says that the double bracket flow on $M(Q)$ is equivalent to a linearly induced flow on $\mathrm{Grass}_{\mathbb{R}}(n, 2n)$. Now the Riccati equation (2.19) is the vector field corresponding to the linear induced flow on an open coordinate chart of $\mathrm{Grass}_{\mathbb{R}}(n, 2n)$ and thus, up to the change of variables described by the diffeomorphism $f : M(Q) \to \mathrm{Grass}_{\mathbb{R}}(n, 2n)$, coincides on that open coordinate chart with the double bracket equation. Thus the double bracket equation on $M(Q)$ might be seen as an extension or completion of the Riccati vector field (2.19). □

Remark 2.16 For $Q = \left[\begin{smallmatrix} I_k & 0 \\ 0 & 0 \end{smallmatrix}\right]$ any $H \in M(Q)$ satisfies $H^2 = H$. Thus the double bracket equation on $M(Q)$ becomes the special Riccati equation on $\mathbb{R}^{n\times n}$

$$\boxed{\dot{H} = HN + NH - 2HNH}$$

which is shown in Theorem 2.7 to be equivalent to the *general* matrix Riccati equation (for N symmetric) on $\mathbb{R}^{(n-k)\times k}$. □

Finally we consider the case $k = 1$, i.e. the associated flow on the projective space $\mathbb{RP}^{n-1}$. Thus let $Q = \mathrm{diag}(1, 0, \ldots, 0)$. Any $H \in M(Q)$ has a representation $H = x \cdot x'$ where $x' = (x_1, \ldots, x_n)$, $\sum_{j=1}^{n} x_j^2 = 1$, and x is uniquely determined up to multiplication by ± 1. The double bracket flow on $M(Q)$ is equivalent to the gradient flow of the standard Morse function

$$\Phi(x) = \frac{1}{2}\operatorname{tr}(Nxx') = \frac{1}{2}\sum_{i,j=1}^{n} n_{ij} x_i x_j.$$

on $\mathbb{RP}^{n-1}$ see Milnor (1963). Moreover, the Lie bracket flow (1.25) on $\mathrm{O}(n)$ induces, for Q as above, the Rayleigh quotient flow on the $(n-1)$ sphere S^{n-1} of Section 1.3.

Problem 2.17 *Prove that every solution of*

$$\dot{H} = AH + HA - 2HAH, \qquad H(0) = H_0, \tag{2.20}$$

has the form

$$H(t) = e^{tA} H_0 \left(I_n - H_0 + e^{2tA} H_0\right)^{-1} e^{tA}, \qquad t \in \mathbb{R}.$$

Problem 2.18 *Show that the spectrum (i.e. the set of eigenvalues) $\sigma(H(t))$ of any solution is given by*

$$\sigma(H(t)) = \sigma\left(H_0 \left(e^{-2tA} (I_n - H_0) + H_0\right)\right).$$

Problem 2.19 *Show that for any solution $H(t)$ of* (2.20) *also $G(t) = I_n - H(-t)$ solves* (2.20).

Problem 2.20 *Derive similar formulas for time-varying matrices $A(t)$.*

Main Points of Section

The double bracket equation $\dot{H} = [H, [H, N]]$ with $H(0)$ a Jacobi matrix and $N = \operatorname{diag}(1, 2, \ldots, n)$ preserves the Jacobi property in $H(t)$ for all $t \geq 0$. In this case the double bracket equation is the *Toda flow* $\dot{H} = [H, H_u - H_\ell]$ which has interpretations in Hamiltonian mechanics.

For the case of symmetric matrices N and $Q = \operatorname{diag}(\lambda_1 I_k, \lambda_2 I_{n-k})$ with $\lambda_1 > \lambda_2$, the double bracket flow is equivalent to the flow on the Grassmann manifold $\operatorname{Grass}_{\mathbb{R}}(k, n)$ linearly induced by $(\lambda_1 - \lambda_2) N$. This in turn is equivalent to a Riccati equation (2.11) with generating matrix $A = (\lambda_1 - \lambda_2) N$. This result gives an interpretation of certain Riccati equations of linear optimal control as gradient flows.

For the special case $Q = \operatorname{diag}(I_k, 0, \ldots, 0)$, then $H^2 = H$ and the double bracket equation becomes the special Riccati equation $\dot{H} = HN + NH - 2HNH$. When $k = 1$, then $H = xx'$ with x a column vector and the double bracket equation is induced by $\dot{x} = (N - x'NxI_n)x$, which is the Rayleigh quotient gradient flow of $\operatorname{tr}(Nxx')$ on the sphere S^{n-1}.

2.3 Recursive Lie-Bracket Based Diagonalization

Now that the double bracket flow with its rich properties has been studied, it makes sense to ask whether or not there are corresponding recursive versions. Indeed there are. For some ideas and results in this direction we refer to Chu (1992b), Brockett (1993), and Moore, Mahony and Helmke

(1994). In this section we study the following recursion, termed the Lie-bracket recursion,

$$\boxed{H_{k+1} = e^{-\alpha[H_k,N]} H_k e^{\alpha[H_k,N]}, \qquad H_0 = H_0', \qquad k \in \mathbb{N}} \tag{3.1}$$

for arbitrary symmetric matrices $H_0 \in \mathbb{R}^{n\times n}$, and some suitably small scalar α, termed a step size scaling factor. A key property of the recursion (3.1) is that it is isospectral. This follows since $e^{\alpha[H_k,N]}$ is orthogonal, as indeed is any e^A where A is skew symmetric.

To motivate this recursion (3.1), observe that H_{k+1} is also the solution at time $t = \alpha$ to a linear matrix differential equation initialized by H_k as follows

$$\begin{aligned} \frac{d\bar{H}}{dt} &= [\bar{H}, [H_k, N]], \qquad \bar{H}(0) = H_k, \\ H_{k+1} &= \bar{H}(\alpha). \end{aligned}$$

For small t, and thus α small, $\bar{H}(t)$ appears to be a solution close to that of the corresponding double bracket flow $H(t)$ of $\dot{H} = [H, [H, N]]$, $H(0) = H_k$. This suggests, that for step-size scaling factors not too large, or not decaying to zero too rapidly, that the recursion (3.1) should inherit the exponential convergence rate to desired equilibria of the continuous-time double bracket equation. Indeed, in some applications this piece-wise constant, linear, and isospectral differential equation may be more attractive to implement than a nonlinear matrix differential equation.

Our approach in this section is to optimize in some sense the α selections in (3.1) according to the potential of the continuous-time gradient flow. The subsequent bounding arguments are similar to those developed by Brockett (1993), see also Moore et al. (1994). In particular, for the potential function

$$f_N(H_k) = -\tfrac{1}{2}\|N - H_k\|^2 = \operatorname{tr}(NH_k) - \tfrac{1}{2}\|N\|^2 - \tfrac{1}{2}\|H_0\|^2 \tag{3.2}$$

we seek to maximize at each iteration its increase

$$\Delta f_N(H_k, \alpha) := f_N(H_{k+1}) - f_N(H_k) = \operatorname{tr}(N(H_{k+1} - H_k)) \tag{3.3}$$

Lemma 3.1 *The constant step-size selection*

$$\alpha = \frac{1}{4\|H_0\| \cdot \|N\|} \tag{3.4}$$

satisfies $\Delta f_N(H_k, \alpha) > 0$ *if* $[H_k, N] \neq 0$.

Proof 3.2 Let $H_{k+1}(\tau) = e^{-\tau[H_k,N]} H_k e^{\tau[H_k,N]}$ be the $k+1$th iteration of (3.1) for an arbitrary step-size scaling factor $\tau \in \mathbb{R}$. It is easy to verify that

$$\begin{aligned}
\frac{d}{d\tau} H_{k+1}(\tau) &= [H_{k+1}(\tau), [H_k, N]] \\
\frac{d^2}{d\tau^2} H_{k+1}(\tau) &= [[H_{k+1}(\tau), [H_k, N]], [H_k, N]].
\end{aligned}$$

Applying Taylor's theorem, then since $H_{k+1}(0) = H_k$

$$H_{k+1}(\tau) = H_k + \tau [H_k, [H_k, N]] + \tau^2 \mathcal{R}_2(\tau),$$

where

$$\mathcal{R}_2(\tau) = \int_0^1 [[H_{k+1}(y\tau), [H_k, N]], [H_k, N]] (1-y)\, dy. \tag{3.5}$$

Substituting into (3.3) gives, using matrix norm inequalities outlined in Sections A.6 and A.9,

$$\begin{aligned}
\Delta f_N(H_k, \tau) &= \operatorname{tr}\left(N\left(\tau [H_k, [H_k, N]] + \tau^2 \mathcal{R}_2(\tau)\right)\right) \\
&= \tau \|[H_k, N]\|^2 + \tau^2 \operatorname{tr}(N \mathcal{R}_2(\tau)) \\
&\geq \tau \|[H_k, N]\|^2 - \tau^2 |\operatorname{tr}(N \mathcal{R}_2(\tau))| \\
&\geq \tau \|[H_k, N]\|^2 - \tau^2 \|N\| \cdot \|\mathcal{R}_2(\tau)\| \\
&\geq \tau \|[H_k, N]\|^2 - \tau^2 \|N\| \\
&\quad \cdot \int_0^1 \|[[H_{k+1}(y\tau), [H_k, N]], [H_k, N]]\| (1-y)\, dy \\
&\geq \tau \|[H_k, N]\|^2 - 2\tau^2 \|N\| \cdot \|H_0\| \cdot \|[H_k, N]\|^2 \\
&=: \Delta f_N^L(H_k, \tau).
\end{aligned} \tag{3.6}$$

Thus $\Delta f_N^L(H_k, \tau)$ is a lower bound for $\Delta f_N(H_k, \tau)$ and has the property that for sufficiently small $\tau > 0$, it is strictly positive, see Figure 3.1. Due to the explicit form of $\Delta f_N^L(H_k, \tau)$ in τ, it is immediately clear that if $[H_k, N] \neq 0$, then $\alpha = 1/(4\|H_0\|\,\|N\|)$ is the unique maximum of (3.6). Hence, $f_N(H_k, \alpha) \geq f_N^L(H_k, \alpha) > 0$ for $[H_k, N] \neq 0$. ■

This lemma leads to the main result of the section.

Theorem 3.3 *Let $H_0 = H_0'$ be a real symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and let $N = diag(\mu_1, \ldots, \mu_n)$, $\mu_1 > \cdots > \mu_n$. The Lie-bracket recursion* (3.1)*, restated as*

$$\boxed{H_{k+1} = e^{-\alpha[H_k,N]} H_k e^{\alpha[H_k,N]}, \qquad \alpha = 1/(4\|H_0\| \cdot \|N\|)} \tag{3.7}$$

with initial condition H_0, has the following properties:

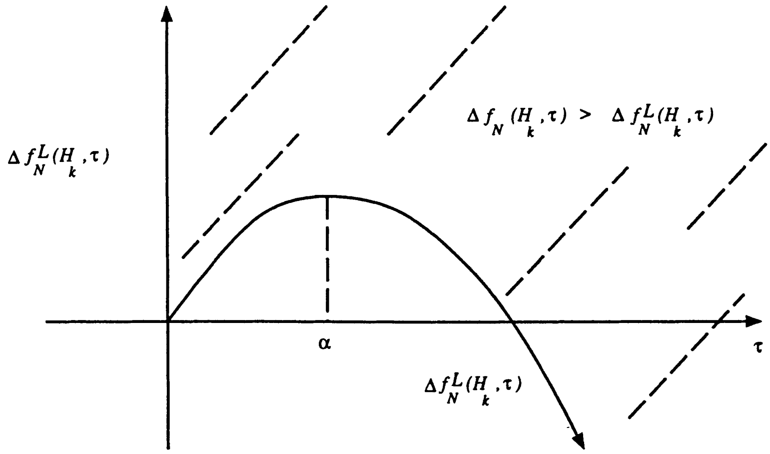

FIGURE 3.1. The lower bound on $\Delta f_N^L(H_k,\tau)$

(a) The recursion is isospectral.

(b) If (H_k) is a solution of the Lie-bracket algorithm, then $f_N(H_k)$ of (3.2) is a strictly monotonically increasing function of $k \in \mathbb{N}$, as long as $[H_k, N] \neq 0$.

(c) Fixed points of the recursive equation are characterised by matrices $H_\infty \in M(H_0)$ such that $[H_\infty, N] = 0$, being exactly the equilibrium points of the double bracket equation (1.1).

(d) Let (H_k), $k = 1, 2, \ldots$, be a solution to the recursive Lie-bracket algorithm, then H_k converges to a matrix $H_\infty \in M(H_0)$, $[H_\infty, N] = 0$, an equilibrium point of the recursion.

(e) All equilibrium points of the recursive Lie-bracket algorithm are unstable, except for $Q = diag(\lambda_1, \ldots, \lambda_n)$ which is locally exponentially stable.

Proof 3.4 To prove Part (a), note that the Lie-bracket $[H, N]' = -[H, N]$ is skew-symmetric. As the exponential of a skew-symmetric matrix is orthogonal, (3.1) is just conjugation by an orthogonal matrix, and hence is an isospectral transformation. Part (b) is a direct consequence of Lemma 3.1.

For Part (c) note that if $[H_k, N] = 0$, then by direct substitution into (3.1) we see $H_{k+l} = H_k$ for $l \geq 1$, and thus H_k is a fixed point. Conversely if $[H_k, N] \neq 0$, then from (3.6), $\Delta f_N(H_k, \alpha) \neq 0$, and thus $H_{k+1} \neq H_k$. In particular, fixed points of (3.1) are equilibrium points of (1.1), and furthermore, from Theorem 1.5, these are the only equilibrium points of (1.1).

To show Part (d), consider the sequence H_k generated by the recursive Lie-bracket algorithm for a fixed initial condition H_0. Observe that Part (b) implies that $f_N(H_k)$ is strictly monotonically increasing for all k where $[H_k, N] \neq 0$. Also, since f_N is a continuous function on the compact set $M(H_0)$, then f_N is bounded from above, and $f_N(H_k)$ will converge to some $f_\infty \leq 0$ as $k \to \infty$. As $f_N(H_k) \to f_\infty$ then $\Delta f_N(H_k, \alpha^c) \to 0$.

For some small positive number ε, define an open set $D_\epsilon \subset M(H_0)$ consisting of all points of $M(H_0)$, within an ϵ-neighborhood of some equilibrium point of (3.1). The set $M(H_0) - D_\varepsilon$ is a closed, compact subset of $M(H_0)$, on which the matrix function $H \mapsto \|[H, N]\|$ does not vanish. As a consequence, the difference function (3.3) is continuous and strictly positive on $M(H_0) - D_\varepsilon$, and thus, is under bounded by some strictly positive number $\delta_1 > 0$. Moreover, as $\Delta f_N(H_k, \alpha) \to 0$, there exists a $K = K(\delta_1)$ such that for all $k > K$ then $0 \leq \Delta f_N(H_k, \alpha) < \delta_1$. This ensures that $H_k \in D_\varepsilon$ for all $k > K$. In other words, (H_k) is converging to some subset of possible equilibrium points.

From Theorem 1.5 and the imposed genericity assumption on N, it is known that the double bracket equation (1.1) has only a finite number of equilibrium points. Thus H_k converges to a finite subset in $M(H_0)$. Moreover, $[H_k, N] \to 0$ as $k \to \infty$. Therefore $\|H_{k+1} - H_k\| \to 0$ as $k \to \infty$. This shows that (H_k) must converge to an equilibrium point, thus completing the proof of Part (d).

To establish exponential convergence, note that since α is constant, the map

$$H \mapsto e^{-\alpha[H,N]} H e^{\alpha[H,N]}$$

is a smooth recursion on all $M(H_0)$, and we may consider the linearization of this map at an equilibrium point $\pi' Q \pi$. The linearization of this recursion, expressed in terms of $\Xi_k \in T_{\pi' Q \pi} M(H_0)$, is

$$\Xi_{k+1} = \Xi_k - \alpha \left[(\Xi_k N - N \Xi_k) \pi' Q \pi - \pi' Q \pi (\Xi_k N - N \Xi_k) \right]. \quad (3.8)$$

Thus for the elements of $\Xi_k = (\xi_{ij})_k$ we have

$$(\xi_{ij})_{k+1} = \left[1 - \alpha \left(\lambda_{\pi(i)} - \lambda_{\pi(j)}\right)(\mu_i - \mu_j)\right] (\xi_{ij})_k, \quad \text{for } i, j = 1, \dots, n. \quad (3.9)$$

The tangent space $T_{\pi' Q \pi} M(H_0)$ at $\pi' Q \pi$ consists of those matrices $\Xi = [\pi' Q \pi, \Omega]$ where $\Omega \in \text{skew}(n)$, the class of skew symmetric matrices. Thus, the matrices Ξ are linearly parametrized by their components ξ_{ij}, where $i < j$, and $\lambda_{\pi(i)} \neq \lambda_{\pi(j)}$. As this is a linearly independent parametrisation, the eigenvalues of the linearization (3.8) can be read directly from the linearization (3.9), and are $1-\alpha\left(\lambda_{\pi(i)}-\lambda_{\pi(j)}\right)(\mu_i-\mu_j)$, for $i<j$ and $\lambda_{\pi(i)}\neq\lambda_{\pi(j)}$. From classical stability theory for discrete-time linear dynamical systems, (3.8) is asymptotically stable if and only if all eigenvalues have absolute value less than 1. Equivalently, (3.8) is asymptotically stable if and only if

$$0<\alpha\left(\lambda_{\pi(i)}-\lambda_{\pi(j)}\right)(\mu_i-\mu_j)<2$$

for all $i<j$ with $\lambda_{\pi(i)}\neq\lambda_{\pi(j)}$. This condition is only satisfied when $\pi=I$ and consequently $\pi'Q\pi=Q$. Thus the only possible stable equilibrium point for the recursion is $H_\infty=Q$. Certainly $(\lambda_i-\lambda_j)(\mu_i-\mu_j)<4\|N\|_2\|H_0\|_2$. Also since $\|N\|_2\|H_0\|_2<2\|N\|\|H_0\|$ we have $\alpha<\frac{1}{2\|N\|_2\|H_0\|_2}$. Therefore $\alpha(\lambda_i-\lambda_j)(\mu_i-\mu_j)<2$ for all $i<j$ which establishes exponential stability of (3.8). This completes the proof. ■

Remark 3.5 In the nongeneric case where N has multiple eigenvalues, the proof techniques for Parts (d) and (e) do not apply. All the results except convergence to a *single* equilibrium point remain in force. □

Remark 3.6 It is difficult to characterise the set of exceptional initial conditions, for which the algorithm converges to some unstable equilibrium point $H_\infty\neq Q$. However, in the continuous-time case it is known that the unstable basins of attraction of such points are of zero measure in $M(H_0)$, see Section 2.1. □

Remark 3.7 By using a more sophisticated bounding argument, a variable step size selection can be determined as

$$\alpha_k=\frac{1}{2\|[H_k,N]\|}\log\left(\frac{\|[H_k,N]\|^2}{\|H_0\|\cdot\|[N,[H_k,N]]\|}+1\right) \tag{3.10}$$

Rigorous convergence results are given for this selection in Moore et al. (1994). The convergence rate is faster with this selection. □

Recursive Flows on Orthogonal Matrices

The associated recursions on the orthogonal matrices corresponding to the gradient flows (1.25) are

$$\boxed{\Theta_{k+1}=\Theta_k e^{\alpha_k[\Theta_k' H_0\Theta_k,N]},\qquad \alpha=1/\left(4\|H_0\|\cdot\|N\|\right)} \tag{3.11}$$

where Θ_k is defined on $O(n)$ and α is a general step-size scaling factor. Thus $H_k=\Theta_k' H_0\Theta_k$ is the solution of the Lie-bracket recursion (3.1). Precise results on (3.11) are now stated and proved for generic H_0 and constant step-size selection, although corresponding results are established in Moore et al. (1994) for the variable step-size scaling factor (3.10).

Theorem 3.8 *Let $H_0 = H_0'$ be a real symmetric $n \times n$ matrix with distinct eigenvalues $\lambda_1 > \cdots > \lambda_n$. Let $N \in \mathbb{R}^{n\times n}$ be* $\operatorname{diag}(\mu_1, \ldots, \mu_n)$ *with $\mu_1 > \cdots > \mu_n$. Then the recursion* (3.11) *referred to as the* associated orthogonal Lie-bracket algorithm, *has the following properties:*

(a) A solution Θ_k, $k = 1, 2, \ldots$, to the associated orthogonal Lie-bracket algorithm remains orthogonal.

(b) Let $f_{N,H_0} : O(n) \to \mathbb{R}$, $f_{N,H_0}(\Theta) = -\frac{1}{2}\|\Theta' H_0 \Theta - N\|^2$ be a function on $O(n)$. Let Θ_k, $k = 1, 2, \ldots$, be a solution to the associated orthogonal Lie-bracket algorithm. Then $f_{N,H_0}(\Theta_k)$ is a strictly monotonically increasing function of $k \in \mathbb{N}$, as long as $[\Theta_k' H_0 \Theta_k, N] \neq 0$.

(c) Fixed points of the recursive equation are characterised by matrices $\Theta \in O(n)$ such that

$$[\Theta' H_0 \Theta, N] = 0.$$

There are exactly $2^n n!$ such fixed points.

(d) Let Θ_k, $k = 1, 2, \ldots$, be a solution to the associated orthogonal Lie-bracket algorithm, then Θ_k converges to an orthogonal matrix Θ_∞, satisfying $[\Theta_\infty' H_0 \Theta_\infty, N] = 0$.

(e) All fixed points of the associated orthogonal Lie-bracket algorithm are strictly unstable, except those 2^n points $\Theta_* \in O(n)$ such that

$$\Theta_*' H_0 \Theta_* = Q,$$

where $Q = diag(\lambda_1, \ldots, \lambda_n)$. Such points Θ_* are locally exponentially asymptotically stable and $H_0 = \Theta_* Q \Theta_*'$ is the eigenspace decomposition of H_0.

Proof 3.9 Part (a) follows directly from the orthogonal nature of $e^{\alpha[\Theta_k' H_0 \Theta_k, N]}$. Let $g : O(n) \to M(H_0)$ be the matrix valued function $g(\Theta) = \Theta' H_0 \Theta$. Observe that g maps solutions $(\Theta_k \mid k \in \mathbb{N})$ of (3.11) to solutions $(H_k \mid k \in \mathbb{N})$ of (3.1).

Consider the potential $f_{N,H}(\Theta_k) = -\frac{1}{2}\|\Theta_k' H_0 \Theta_k - N\|^2$, and the potential $f_N = -\frac{1}{2}\|H_k - N\|^2$. Since $g(\Theta_k) = H_k$ for all $k = 1, 2, \ldots$, then $f_{N,H_0}(\Theta_k) = f_N(g(\Theta_k))$ for $k = 1, 2, \ldots$. Thus $f_N(H_k) = f_N(g(\Theta_k)) = f_{N,H_0}(\Theta_k)$ is strictly monotonically increasing for $[H_k, N] = [g(\Theta_k), N] \neq 0$, and Part (b) follows.

If Θ_k is a fixed point of the associated orthogonal Lie-bracket algorithm with initial condition Θ_0, then $g(\Theta_k)$ is a fixed point of the Lie-bracket algorithm. Thus, from Theorem 3.3, $[g(\Theta_k), N] = [\Theta_k' H_0 \Theta_k, N] = 0$. Moreover, if $[\Theta_k' H_0 \Theta_k, N] = 0$ for some given $k \in \mathrm{N}$, then by inspection $\Theta_{k+l} = \Theta_k$ for $l = 1, 2, \dots$, and Θ_k is a fixed point of the associated orthogonal Lie-bracket algorithm. A simple counting argument shows that there are precisely $2^n n!$ such points and Part (c) is established.

To prove (d) note that since $g(\Theta_k)$ is a solution to the Lie-bracket algorithm, it converges to a limit point $H_\infty \in M(H_0)$, $[H_\infty, N] = 0$, by Theorem 3.3. Thus Θ_k must converge to the pre-image set of H_∞ via the map g. The genericity condition on H_0 ensures that the set generated by the pre-image of H_∞ is a finite disjoint set. Since $\|[g(\Theta_k), N]\| \to 0$ as $k \to \infty$, then $\|\Theta_{k+1} - \Theta_k\| \to 0$ as $k \to \infty$. From this convergence of Θ_k follows.

To prove Part (e), observe that, due to the genericity condition on H_0, the dimension of $O(n)$ is the same as the dimension of $M(H_0)$. Thus g is locally a diffeomorphism on $O(n)$, and taking a restriction of g to such a region, the local stability structure of the equilibria are preserved under the map g^{-1}. Thus, all fixed points of the associated orthogonal Lie-bracket algorithm are locally unstable except those that map via g to the unique locally asymptotically stable equilibrium of the Lie-bracket recursion. ∎

Simulations

A simulation has been included to demonstrate the recursive schemes developed. The simulation deals with a real symmetric 7×7 initial condition, H_0, generated by an arbitrary orthogonal similarity transformation of matrix $Q = \mathrm{diag}(1, 2, 3, 4, 5, 6, 7)$. The matrix N was chosen to be $\mathrm{diag}(1, 2, 3, 4, 5, 6, 7)$ so that the minimum value of f_N occurs at Q such that $f_N(Q) = 0$. Figure 3.2 plots the diagonal entries of H_k at each iteration and demonstrates the asymptotic convergence of the algorithm. The exponential behaviour of the curves appears at around iteration 30, suggesting that this is when the solution H_k enters the locally exponentially attractive domain of the equilibrium point Q. Figure 3.3 shows the evolution of the potential $f_N(H_k) = -\frac{1}{2}\|H_k - N\|^2$, demonstrating its monotonic increasing properties and also displaying exponential convergence after iteration 30.

Computational Considerations

It is worth noting that an advantage of the recursive Lie-bracket scheme over more traditional linear algebraic schemes for the same tasks, is the

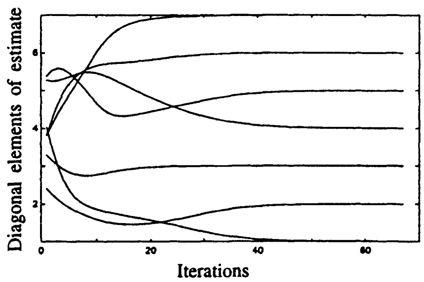

FIGURE 3.2. The recursive Lie-bracket scheme

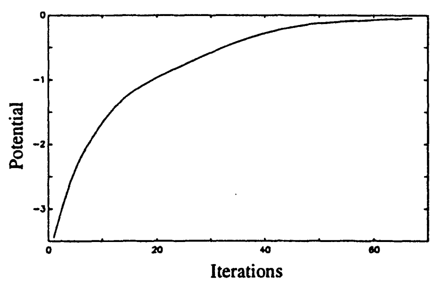

FIGURE 3.3. The potential function $f_N(H_k)$

presence of a step-size α and the arbitrary target matrix N. The focus in this section is only on the step size selection. The challenge remains to devise more optimal (possibly time-varying) target matrix selection schemes for improved performance. This suggests an application of optimal control theory.

It is also possible to consider alternatives of the recursive Lie-bracket scheme which have improved computational properties. For example, consider a $(1,1)$ Padé approximation to the matrix exponential

$$e^{\alpha[H_k,N]} \approx \frac{2I - \alpha[H_k,N]}{2I + \alpha[H_k,N]}.$$

Such an approach has the advantage that, as $[H_k,N]$ is skew symmetric, then the Padé approximation will be orthogonal, and will preserve the isospectral nature of the Lie-bracket algorithm. Similarly, an (n,n) Padé approximation of the exponential for any n will also be orthogonal.

Actually Newton methods involving second order derivatives can be devised to give local quadratic convergence. These can be switched in to the Lie-bracket recursions here as appropriate, making sure that at each iteration the potential function increases. The resulting schemes are then much more competitive with commercial diagonalization packages than the purely linear methods of this section. Of course there is the possibility that quadratic convergence can also be achieved using shift techniques, but we do not explore this fascinating territory here.

Another approach is to take just an Euler iteration,

$$H_{k+1} = H_k + \alpha[H_k,[H_k,N]],$$

as a recursive algorithm on $\mathbb{R}^{n\times n}$. A scheme such as this is similar in form to approximating the curves generated by the recursive Lie-bracket scheme by straight lines. The approximation will not retain the isospectral nature of the Lie-bracket recursion, but this fact may be overlooked in some applications, because it is computationally inexpensive. We cannot recommend this scheme, or higher order versions, except in the neighborhood of an equilibrium.

Main Points of Section

In this section we have proposed a numerical scheme for the calculation of double bracket gradient flows on manifolds of similar matrices. Step-size selections for such schemes has been discussed and results have been obtained on the nature of equilibrium points and on their stability properties. As a consequence, the schemes proposed in this section could be used as a computational tool with known bounds on the total time required to make a calculation. Due to the computational requirements of calculating matrix exponentials these schemes may not be useful as a direct numerical tool in traditional computational environments, however, they provide insight into discretising matrix flows such as is generated by the double bracket equation.

Notes for Chapter 2

As we have seen in this chapter, the isospectral set $M(Q)$ of symmetric matrices with fixed eigenvalues is a homogeneous space and the least squares distance function $f_N : M(Q) \to \mathbb{R}$ is a smooth Morse-Bott function. Quite a lot is known about the critical points of such functions and there is a rich mathematical literature on Morse theory developed for functions defined on homogeneous spaces.

If Q is a rank k projection operator, then $M(Q)$ is a Grassmannian. In this case the trace function $H \mapsto \operatorname{tr}(NH)$ is the classical example of a Morse function on Grassmann manifolds. See Milnor (1963) for the case where $k = 1$ and Wu (1965), Hangan (1968) for a slightly different construction of a Morse function for arbitrary k. For a complete characterization of the critical points of the trace functional on classical groups, the Stiefel manifold and Grassmann manifolds, with applications to the topology of these spaces, we refer to Frankel (1965) and Shayman (1982). For a complete analysis of the critical points and their Morse indices of the trace function on more general classes of homogeneous spaces we refer to Hermann (1962; 1963; 1964), Takeuchi (1965) and Duistermaat et al. (1983). For results on gradient flows of certain least squares functions defined on infinite dimensional homogeneous spaces arising in physics we refer to the important work of Atiyah and Bott (1982), Pressley (1982) and Pressley and Segal (1986). The work of Byrnes and Willems (1986) contains an interesting application of moment map techniques from symplectic geometry to total least squares estimation.

The geometry of varieties of isospectral Jacobi matrices has been studied by Tomei (1984) and Davis (1987). The set $\operatorname{Jac}(\lambda_1, \dots, \lambda_n)$ of Jacobi matrices with distinct eigenvalues $\lambda_1 > \cdots > \lambda_n$ is shown to be a smooth compact manifold. Furthermore, an explicit construction of the universal covering space is given. The report of Driessel (1987b) contains another elementary proof of the smoothness of $\operatorname{Jac}(\lambda_1, \dots, \lambda_n)$. Also the tangent space of $\operatorname{Jac}(\lambda_1, \dots, \lambda_n)$ at a Jacobi matrix L is shown to be the vector space of all Jacobi matrices ζ of the form $\zeta = L\Omega - \Omega L$, where Ω is skew symmetric. Closely related to the above is the work by de Mari and Shayman (1988) de Mari, Procesi and Shayman (1992) on Hessenberg varieties; i.e. varieties of invariant flags of a given Hessenberg matrix. Furthermore there are interesting connections with torus varieties; for this we refer to Gelfand and Serganova (1987).

The double bracket equation (1.1) and its properties were first studied by Brockett (1988); see also Chu and Driessel (1990). The simple proof of the Wielandt-Hoffman inequality via the double bracket equation is due to Chu and Driessel (1990). A systematic analysis of the double bracket flow (1.2) on adjoint orbits of compact Lie groups appears in Bloch, Brockett and Ratiu (1990; 1992). For an application to subspace learning see Brockett (1991a). In Brockett (1989b) it is shown that the double bracket equation can simulate any finite automaton. Least squares matching problems arising in computer vision and pattern analysis are tackled via double bracket-like equations in Brockett (1989a). An interesting connection exists between the double bracket flow (1.1) and a fundamental equation arising in micromagnetics. The Landau-Lifshitz equation on the two-sphere is a nonlinear diffusion equation which, in the absence of diffusion terms, becomes equivalent to the double bracket equation. Stochastic versions of the double bracket flow are studied by Colonius and Kliemann (1990).

A thorough study of the Toda lattice equation with interesting links to representation theory has been made by Kostant (1979). For the Hamiltonian mechanics interpretation of the QR-algorithm and the Toda flow see Flaschka (1974; 1975), Moser (1975), and Bloch (1990b). For connections of the Toda flow with scaling actions on spaces of rational functions in system theory see Byrnes (1978), Brockett and Krishnaprasad (1980) and Krishnaprasad (1979). An interesting interpretation of the Toda flow from a system theoretic viewpoint is given in Brockett and Faybusovich (1991); see also Faybusovich (1989). Numerical analysis aspects of the Toda flow have been treated by Symes (1980b; 1982), Chu (1984b; 1984a), Deift et al. (1983), and Shub and Vasquez (1987). Expository papers are Watkins (1984), Chu (1984a). The continuous double bracket flow $\dot{H} = [H, [H, \operatorname{diag} H]]$ is related to the discrete Jacobi method for diagonalization. For a phase portrait analysis of this flow see Driessel (1987a). See also Wilf (1981) and Golub and Van Loan (1989), Section 8.4 "Jacobi methods", for a discussion on the Jacobi method.

For the connection of the double bracket flow to Toda flows and flows on Grassmannians much of the initial work was done by Bloch (1990b; 1990a) and then by Bloch, Brockett and Ratiu (1990) and Bloch, Flaschka and Ratiu (1990). The connection to the Riccati flow was made explicit in Helmke (1991) and independently observed by Faybusovich. The paper of Faybusovich (1992b) contains a complete description of the phase portrait of the Toda flow and the corresponding QR algorithm, including a discussion of structural stability properties. In Faybusovich (1989) the relationship between QR-like flows and Toda-like flows is described. Monotonicity properties of the Toda flow are discussed in Lagarias (1991). A VLSI type implementation of the Toda flow by a nonlinear lossless electrical network is given by Paul, Hüper and Nossek (1992).

Infinite dimensional versions of the Toda flow with applications to sorting of function values are in Brockett and Bloch (1989), Deift, Li and Tomei (1985), Bloch, Brockett, Kodama and Ratiu (1989). See also the closely related work by Bloch (1985a; 1987) and Bloch and Byrnes (1986).

Numerical integration schemes of ordinary differential equations on manifolds are presented by Crouch and Grossman (1991), Crouch, Grossman and Yan (1992a; 1992b) and Ge-Zhong and Marsden (1988). Discrete-time versions of some classical integrable systems are analyzed by Moser and Veselov (1991). For complexity properties of discrete integrable systems see Arnold (1990) and Veselov (1992). The recursive Lie-bracket diagonalization algorithm (3.1) is analyzed in detail in Moore et al. (1994). Related results appear in Brockett (1993) and Chu (1992a). Step-size selections for discretizing the double bracket flow also appear in the recent PhD thesis of Smith (1993).