Optimality of double bracket and generalized double bracket flows

Anthony M. Bloch* Arieh Iserles†

February 26, 2003

Abstract

In this paper we consider the optimal structure of double bracket flows and generalisations of these flows to more complex gradient flows. We discuss different notions of optimality and the relationship of the flows to the structure of convex polytopes.

1 Introduction

Double-bracket flows have been introduced by Brockett (1991) in the context of control theory and optimization and by Chu & Driessel (1991) in connection with the computation of certain problems in numerical linear algebra. Let $\mathrm{Sym}(n)$ denote the set of real, $n \times n$ symmetric matrices. A *double-bracket flow (DBF)* is the matrix differential equation

$$\boldsymbol{Y}' = [[\boldsymbol{N}, \boldsymbol{Y}], \boldsymbol{Y}], \quad t \geq 0, \qquad \boldsymbol{Y}(0) = \boldsymbol{Y}_0 \in \mathrm{Sym}(n), \tag{1.1}$$

where $\boldsymbol{N} \in \mathrm{Sym}(n)$ and $[\,\cdot\,,\,\cdot\,]$ is the usual matrix commutator. Without loss of generality, we assume that $\boldsymbol{N}$ is a diagonal matrix, $\boldsymbol{N} = \mathrm{diag}\,\boldsymbol{\lambda}$, say. Otherwise we may reduce (1.1) to this form: $\boldsymbol{Y}(t) = \boldsymbol{P}\tilde{\boldsymbol{Y}}(t)\boldsymbol{P}^\top$, where $\tilde{\boldsymbol{Y}}' = [[\boldsymbol{\Lambda}, \tilde{\boldsymbol{Y}}], \tilde{\boldsymbol{Y}}]$, $\tilde{\boldsymbol{Y}}(0) = \boldsymbol{P}^\top \boldsymbol{Y}_0 \boldsymbol{P}$, and $\boldsymbol{N} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^\top$, with $\boldsymbol{P} \in \mathrm{SO}(n)$ and $\boldsymbol{\Lambda}$ a diagonal matrix.

The Lie algebraic generalisation of this flow was analysed in (Bloch, Brockett & Ratiu 1990) and (Bloch, Brockett & Ratiu 1992).

Equations of the form (1.1) share a number of features that render them of interest both in the context of a number of applications and as mathematical objects in their own right. Related equations are discussed in (Brockett 1993) and (Helmke & Moore 1994) for example.

Firstly, (1.1) is a special instance of an *isospectral flow*

$$\boldsymbol{Y}' = [\boldsymbol{B}(t, \boldsymbol{Y}), \boldsymbol{Y}], \quad t \geq 0, \qquad \boldsymbol{Y}(0) = \boldsymbol{Y}_0 \in \mathrm{Sym}(n), \tag{1.2}$$

*Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA, email abloch@umich.edu, supported in part by the National Science Foundation.

†Department of Applied Mathematics and Theoretical Physics, Centre for Mathematical Sciences, University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, England, email A.Iserles@damtp.cam.ac.uk.

where $\boldsymbol{B} : \mathbb{R}_+ \times \mathrm{Sym}(n) \to \mathfrak{so}(n)$, the set $\mathfrak{so}(n)$ being the Lie algebra of $n \times n$ real, skew-symmetric matrices. It easily follows that

$$\boldsymbol{Y}(t) = \boldsymbol{Q}(t) Y_0 \boldsymbol{Q}^\top(t), \quad t \geq 0, \tag{1.3}$$

where

$$\boldsymbol{Q}' = \boldsymbol{B}(t, \boldsymbol{Q}\boldsymbol{Y}_0\boldsymbol{Q}^\top), \quad t \geq 0, \qquad \boldsymbol{Q}(0) = \boldsymbol{I}.$$

Since the latter is a Lie-group equation, we deduce that $\boldsymbol{Q}$ evolves in $\mathrm{SO}(n)$, therefore, by (1.3), $\boldsymbol{Y}$ is a similarity transformation of the initial value $\boldsymbol{Y}_0$ (Iserles, Munthe-Kaas, Nørsett & Zanna 2000). In other words, the eigenvalues of $Y(t)$ in (1.2) (and, by implication, in the DBF (1.1)) are invariants of the flow and do not vary as t increases. We note in passing that the equation (1.3) is the key to practical computation of the solution of (1.2), whilst respecting its invariants (Calvo, Iserles & Zanna 1997, Iserles 2002).

The second feature of the DBF (1.1) is that it is a *gradient system*, with a global Lyapunov function, therefore it is assured of convergence to a fixed point of the flow as $t \to \infty$ (Brockett 1991). More precisely, as shown in (Bloch et al. 1992) it is a gradient flow with respect to the so-called *normal metric* on an appropriate adjoint orbit of a compact Lie group. This feature, of critical importance in applications, is in general invalid for arbitrary isospectral flows (1.2): thus, for example, the solution of

$$\boldsymbol{Y}' = [\boldsymbol{B}, \boldsymbol{Y}], \quad t \geq 0, \qquad \boldsymbol{Y}(0) = \begin{bmatrix} a & b \\ b & c \end{bmatrix}, \qquad \text{where} \qquad \boldsymbol{B} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

is

$$\boldsymbol{Y}(t) = \begin{bmatrix} \frac{1}{2}(a+c) + b\sin 2t + \frac{1}{2}(a-c)\cos 2t & b\cos 2t - \frac{1}{2}(a-c)\sin 2t \\ b\cos 2t - \frac{1}{2}(a-c)\sin 2t & \frac{1}{2}(a+c) - b\sin 2t - \frac{1}{2}(a-c)\cos 2t \end{bmatrix},$$

an oscillatory matrix.

It is generally true in fact that integrable Hamiltonian systems are described by isospectral flows. As shown by Bloch (1990) however, in the case when the matrix $\boldsymbol{Y}$ in the double bracket flow is tridiagonal and the matrix $\boldsymbol{N}$ is the diagonal matrix $\mathrm{diag}(1, 2, \ldots n)$, the double bracket flow is both gradient and Hamiltonian on a level set of its integrals. The flow is the Toda lattice flow and the level set noncompact and diffeomorphic to a product of lines, unlike many Hamiltonian systems where the level set of the integrals is diffeomorphic to a torus. The seminal initial paper on the nonperiodic Toda flow is (Moser 1974) where its gradient nature is demonstrated. Other key early papers include (Symes 1980), (Symes 1982) and (Deift, Nanda & Tomei 1983).

Since the existence of a global Lyapunov function is fundamental to the subject matter of this paper, we herewith reproduce a proof, following that of Brockett (1991), as well as presenting a second, local Lyapunov function which is (to the best of our knowledge) new and helps to shed light on the asymptotic behaviour of $\boldsymbol{Y}(t)$. We define

$$h(t) = \mathrm{tr}\,[\boldsymbol{Y}(t) - \boldsymbol{N}]^2 = \mathrm{tr}\,\boldsymbol{Y}^2(t) + \mathrm{tr}\,\boldsymbol{N}^2 - \mathrm{tr}\,[\boldsymbol{Y}(t)\boldsymbol{N} + \boldsymbol{N}\boldsymbol{Y}(t)], \quad t \geq 0.$$

Both $\operatorname{tr} \boldsymbol{N}^2$ and $\operatorname{tr} \boldsymbol{Y}^2(t)$ are independent of t: the first since $\boldsymbol{N}$ itself is a constant matrix, the second because the trace is the sum of the eigenvalues and, since the eigenvalues of $\boldsymbol{Y}(t)$ stay put, so do also the eigenvalues of $\boldsymbol{Y}^2(t)$. We deduce that

$$h'(t) = -\frac{\mathrm{d}}{\mathrm{d}t} \operatorname{tr} [\boldsymbol{Y}(t)\boldsymbol{N} + \boldsymbol{N}\boldsymbol{Y}(t)], \quad t \geq 0.$$

We find that

$$h' = -2 \sum_{k=1}^{n} \sum_{i=1}^{n} (\lambda_k - \lambda_i)^2 Y_{k,i}^2. \tag{1.4}$$

Therefore, $h'(t) \leq 0$ and, unless $\boldsymbol{Y}(t)$ is a diagonal matrix, $h'(t) < 0$. Since $(\boldsymbol{Y} - \boldsymbol{N})^2$ is positive semidefinite, the function h is nonnegative and we deduce that h is a global Lyapunov function and that $\lim_{t\to\infty} h(t)$ exists. This implies $\lim_{t\to\infty} h'(t) = 0$, therefore $\hat{\boldsymbol{Y}} = \lim_{t\to\infty} \boldsymbol{Y}(t)$ exists and, because of our analysis, *is a diagonal matrix.*

Needless to say, the diagonal elements of $\hat{\boldsymbol{Y}}$ are the eigenvalues of $\boldsymbol{Y}_0$ (hence of $\boldsymbol{Y}(t)$, $t \geq 0$), which we denote by $d_1, d_2, \ldots, d_n$. We henceforth assume for the sake of simplicity (and to avoid a welter of special cases of no great intrinsic significance) that the $2n$ real numbers $\{d_1, d_2, \ldots, d_n\}$ and $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ are distinct. Since any diagonal matrix annihilates the vector field in (1.1), all matrices of the from $\operatorname{diag} \boldsymbol{d_\pi}$, where $\boldsymbol{\pi}$ is a permutation of $\{1, 2, \ldots, n\}$, are fixed points of DBF. To enquire of their attractivity, we let $\boldsymbol{Y}(t) = \operatorname{diag} \boldsymbol{d_\pi} + E(t)$, where $\|E(t)\|$ is small. Since diagonal matrices form a commutative algebra, we deduce that

$$\begin{aligned} E' = Y' &= [[N, \operatorname{diag} \boldsymbol{d_\pi} + E], \operatorname{diag} \boldsymbol{d_\pi} + E] = [[N, E], \operatorname{diag} \boldsymbol{d_\pi} + E] \\ &= [[N, E], \operatorname{diag} \boldsymbol{d_\pi}] + \mathcal{O}\big(\|E\|^2\big) . \end{aligned}$$

Therefore, for every $k, l = 1, 2, \ldots, n$,

$$\begin{aligned} E'_{k,l} &= (\lambda_k d_{\pi_l} - \lambda_l d_{\pi_l} - \lambda_k d_{\pi_k} + \lambda_l d_{\pi_k} E_{k,l} + \mathcal{O}\big(\|E\|^2\big) \\ &= -(\lambda_k - \lambda_l)(d_{\pi_k} - d_{\pi_l}) E_{k,l} + \mathcal{O}\big(\|E\|^2\big) . \end{aligned}$$

We deduce that each $E_{k,l}$ decays (at an exponential speed) if and only if

$$(\lambda_k - \lambda_l)(d_{\pi_k} - d_{\pi_l}) > 0, \qquad k, l = 1, 2, \ldots, n, \ k \neq l. \tag{1.5}$$

Therefore there is just one attractive fixed point $\hat{\boldsymbol{Y}}$, given by the permutation of $\boldsymbol{d}$ that arranges the eigenvalues in the same ordering as the $\boldsymbol{\lambda}s$.

Next, we construct a *local* Lyapunov function,

$$m(t) = \sum_{k=1}^{n} \sum_{\substack{l=1 \\ l\neq k}}^{n} Y_{k,l}^2(t) = \sum_{k=1}^{n} \sum_{l=1}^{n} Y_{k,l}^2(t) - \sum_{k=1}^{n} Y_{k,k}^2(t) = \|\boldsymbol{Y}(t)\|_{\mathsf{F}}^2 - \sum_{k=1}^{n} Y_{k,k}^2(t),$$

where $\| \cdot \|_{\mathsf{F}}$ is the *Frobenius norm.* Since $\|\boldsymbol{A}\|_{\mathsf{F}}^2$ equals the sum of squares of the eigenvalues of a real, square matrix $\boldsymbol{A}$, we use isospectrality to argue that $\|\boldsymbol{Y}(t)\|_{\mathsf{F}}^2$ is constant. We find that that

$$
\begin{aligned}
m'(t) &= -2\sum_{k=1}^{n} Y_{k,k}Y'_{k,k} = -4\sum_{k=1}^{n} Y_{k,k}\sum_{i=1}^{n}(\lambda_k-\lambda_i)Y_{k,i}^2\\
&= -2\sum_{k=1}^{n}\sum_{i=1}^{n}(\lambda_k-\lambda_i)Y_{k,k}Y_{k,i}^2 - 2\sum_{i=1}^{n}\sum_{k=1}^{n}(\lambda_i-\lambda_k)Y_{i,i}Y_{k,i}^2\\
&= -2\sum_{k=1}^{n}\sum_{i=1}^{n}(\lambda_k-\lambda_i)(Y_{k,k}-Y_{i,i})Y_{k,i}^2.
\end{aligned}
$$

We have already seen that, unless $\boldsymbol{Y}_0$ is a diagonal matrix, the solution $\boldsymbol{Y}$ tends to the diagonal matrix consistently with (1.5). Therefore, there exists $t^* \geq 0$ such that for every $t \geq t^*$ the vector $\mathrm{diag}\,\boldsymbol{Y}(t)$ lies in the closed, nonempty set

$$\{\boldsymbol{x} \in \mathbb{R}^n : (\lambda_k-\lambda_i)(x_k-x_i) \geq 0,\ k,l = 1,2\ldots,n\}.$$

As we have just demonstrated, within this neighbourhood of the attractive fixed point, $m \geq 0$ is a local Lyapunov function and the Euclidean norm of off-diagonal elements is decaying monotonically.

The third important feature of the DBF (1.1) is that its fixed point $\hat{\boldsymbol{Y}}$, a diagonal matrix with $\boldsymbol{d}_{\boldsymbol{\pi}}$ along the diagonal (cf. (1.5)) optimises distance to $\boldsymbol{N}$ in the Frobenius norm.

As a matter of fact, one can distinguish between three types of optimality:

1. *Discrete optimality:* The Frobenius distance from $\boldsymbol{N}$ is minimised at $\hat{\boldsymbol{Y}}$ among all the $n!$ diagonal matrices similar to $\boldsymbol{Y}_0$. This, relatively weak form of optimality is the one considered (and proved) by Brockett (1991). Indeed, in Brockett's terminology, the DBF (1.1) is an "analogue computer" for the calculation of eigenvalues, sorting of lists and the solution of linear programming problems, and this makes its asymptotic behaviour, hence discrete optimality, a natural concept.

2. *Isospectral optimality:* Given an arbitrary matrix norm $\|\cdot\|$, is it true that for all (without loss of generality) diagonal matrices $\boldsymbol{N}$ and all $\boldsymbol{Y}_0 \in \mathrm{Sym}(n)$ that

   $$\|\hat{\boldsymbol{Y}}-\boldsymbol{N}\| \leq \|\boldsymbol{X}-\boldsymbol{N}\|, \qquad \boldsymbol{X} \in \mathrm{Sym}(n), \quad \boldsymbol{X} \text{ similar to } \boldsymbol{Y}_0, \quad \|\boldsymbol{X}-\hat{\boldsymbol{Y}}\| \text{ small?}$$

   In the latter case we say that the norm is isospectrally optimal. Moreover, if the inequality is (for every $\boldsymbol{Y}_0$ and $\boldsymbol{N}$) sharp for $\boldsymbol{X} \neq \hat{\boldsymbol{Y}}$ in an arbitrarily small open neighbourhood, except perhaps for a lower-dimensional submanifold, we say that the norm optimality is *strong.* If for some $\boldsymbol{Y}_0$ and $\boldsymbol{N}$ we have equality in a proper neighbourhood of $\hat{Y}$ then optimality is *weak.*

3. *Global optimality:* This is the strongest possible form of optimality,

   $$\|\hat{\boldsymbol{Y}}-\boldsymbol{N}\| \leq \|\boldsymbol{X}-\boldsymbol{N}\|, \qquad \boldsymbol{X} \in \mathrm{Sym}(n), \quad \boldsymbol{X} \text{ similar to } \boldsymbol{Y}_0. \tag{1.6}$$

   Intriguingly, although never mentioned in (Brockett 1991), it is implicit in the analysis of that paper that the Frobenius norm is globally optimal. This follows at once from $h'(t) \leq 0$ (since $h(t) = \|Y(t) - \boldsymbol{N}\|_{\mathsf{F}}^2$), the fact that $\hat{Y}$ is the only fixed point for nondiagonal $\boldsymbol{Y}_0$s and discrete optimality. Given that all these facts are already present in (Brockett 1991), it is only fair to attribute the global optimality of the Frobenius norm to Brockett.

In this paper we wish to focus on isospectral optimality for different unitarily-invariant norms: a matrix norm $\|\cdot\|$ is *unitarily invariant* if $\|\boldsymbol{P}\boldsymbol{X}\boldsymbol{Q}\| = \|\boldsymbol{X}\|$ for every $\boldsymbol{P}, \boldsymbol{Q} \in \mathrm{SO}(n)$, $\boldsymbol{X} \in \mathrm{M}_{n,n}$ (Horn & Johnson 1991, p. 203). We commence Section 2 with the celebrated characterisation of unitarily-invariant norms in terms of symmetric gauges, due to von Neumann (1937). The most important family of unitarily-invariant norms are the so-called *Schatten p-norms,* $1 \leq p \leq \infty$. In the remainder of the section we determine the range of p such that a Schatten p-norm is isospectrally optimal.

We remark that the derivative of the Frobenius norm $h(t) = \|\boldsymbol{Y}(t) - \boldsymbol{N}\|_{\mathsf{F}}^2$ along the flow of the double bracket equations reduces to the derivative of $\mathrm{tr}\,\boldsymbol{Y}\boldsymbol{N}$ which is precisely the gradient function for the flow. In this case global optimality follows as in paragraph 3 above. The key difficulty in the current paper is that the more general norms we discuss here involve higher powers of $\boldsymbol{Y}$.

In Section 3 we discuss the case $n = 2$. Although the solution of (1.1) can be easily written down in an explicit fashion, isospectral optimality is far from trivial. We resolve this issue completely and characterise all isospectrally-optimal norms for $n = 2$. We also show that all Schatten p-norms are globally optimal for $n = 2$.

In Section 4 we move on to consider generalisations of the double bracket flow where the flow is the gradient of a higher-power trace. This leads a richer equilibrium and stability structure which we examine. We also discuss the role of the Schur–Horn polytope in the dynamics of these systems, generalising the work in (Bloch et al. 1992).

The result of Schur is as follows: Let A be a Hermitian matrix with eigenvalues λ_j, $j = 1, \ldots, n$. Denote by $\lambda = (\lambda_1, \ldots, \lambda_n)$ the vector of eigenvalues of A and by $A^0 = (A_{11}, \ldots, A_{nn})$ the diagonal of A. Then $A^0 \in (\widehat{\Sigma_n \lambda})$. Here the last notation denotes the convex hull under the action of the permutation group.

Conversely the result of Horn is: Let $\lambda \in R^n$, with components arranged in non-increasing order. If $A^0 \in (\widehat{\Sigma_n \lambda})$, there is a Hermitian matrix A with eigenvalues λ whose diagonal is A^0.

2 Schatten p-norms

A vector norm $|\!|\!| \cdot |\!|\!| : \mathbb{R}^n \to \mathbb{R}_+$ is said to be *symmetric* if $|\!|\!| \boldsymbol{P}\boldsymbol{x} |\!|\!| = |\!|\!| \boldsymbol{x} |\!|\!|$ for every $\boldsymbol{x} \in \mathbb{R}^n$ and every permutation matrix $\boldsymbol{P} \in \mathrm{M}_{n,n}$, and *absolute* if $|\!|\!| \boldsymbol{x} |\!|\!| = |\!|\!| |\boldsymbol{x}| |\!|\!|$, $\boldsymbol{x} \in \mathbb{R}^n$, where $|\boldsymbol{x}|$ is the vector of the absolute values of the elements of $\boldsymbol{x}$. If $|\!|\!| \cdot |\!|\!|$ is both symmetric and absolute, it is called a *symmetric gauge* (Horn & Johnson 1991, Lewis & Overton 1996). Thus, for example, every $\ell_p[\mathbb{R}^n]$ norm is a symmetric gauge, but no nontrivial weighted $\ell_p[\mathbb{R}^n]$ norm is symmetric and the norm $|\!|\!| [x_1, x_2]^\top |\!|\!| = [x_1^2 + x_2^2 + (x_1 - x_2)^2]^{1/2}$ is not positive.[1]

[1] We leave it as an elementary exercise to the reader to prove that this is indeed a norm. Perhaps the easiest proof follows by showing that its unit ball is convex.

Given $\boldsymbol{X} \in \mathrm{M}_{n,n}$, we denote by $\operatorname{sing}(\boldsymbol{X})$ the vector of *singular values* of $\boldsymbol{X}$: to ensure proper definition, we assume that the elements of $\operatorname{sing}(\boldsymbol{X})$ are arranged in increasing order, but this will play no further role in our discussion. Note that for $\boldsymbol{X} \in \mathrm{Sym}(n)$ the map $\operatorname{sing}(\boldsymbol{X})$ produces the absolute values of the eigenvalues of the matrix.

Theorem 1 (von Neumann (1937)) *A matrix norm $\|\cdot\|$ is unitarily invariant if and only if there exists a symmetric gauge $|\cdot|$ such that*

$$\|\boldsymbol{X}\| = |\operatorname{sing}(\boldsymbol{X})|, \qquad \boldsymbol{X} \in \mathrm{M}_{n,n}. \tag{2.1}$$

We have already mentioned that the standard $\ell_p[\mathbb{R}^n]$ norm, $1 \le p \le \infty$, is a symmetric gauge. Therefore, by Theorem 1, it gives rise to a unitarily-invariant norm, the *Schatten p-norm* $\|\cdot\|_p = |\operatorname{sing}(\cdot)|_p$ (Horn & Johnson 1991, p. 199). In particular, the special choices $p \in \{1, 2, \infty\}$ are important enough to deserve special names:

$\|\boldsymbol{X}\|_1 = \mathbf{1}^\top \operatorname{sing}(\boldsymbol{X})$:	the Ky Fan norm (Horn & Johnson 1991, p. 195),
$\|\boldsymbol{X}\|_2 = \|\boldsymbol{X}\|_{\mathsf{F}}$:	the Frobenius norm,
$\|\boldsymbol{X}\|_\infty = \max \dfrac{|\boldsymbol{X}\boldsymbol{v}|_2}{|\boldsymbol{v}|_2}$:	the (operator) Euclidean norm.

Noting in passing that the Schatten 2-norm is isospectrally optimal, we seek to characterise the range of $p \in [1, \infty]$ for which a Schatten p-norm shares this feature.

We can prove, leaving the details to (Bloch & Iserles 2003*b*), that

Theorem 2 *The Schatten p-norm is strongly isospectrally optimal for $1 < p \le \infty$ and weakly isospectrally optimal for $p = 1$.*

The idea of the proof is as follows: We consider matrices $\boldsymbol{Y}_\varepsilon \in \mathrm{Sym}(n)$ from an isospectral orbit of $\hat{\boldsymbol{Y}}$ (i.e., similar to $\hat{\boldsymbol{Y}}$) within distance $0 < \varepsilon \ll 1$ from $\hat{\boldsymbol{Y}}$. By construction, there exists a nonzero $\boldsymbol{\Omega} \in \mathfrak{so}(n)$ such that

$$\begin{aligned}\boldsymbol{Y}_\varepsilon &= \mathrm{e}^{\varepsilon \boldsymbol{\Omega}} \hat{\boldsymbol{Y}} \mathrm{e}^{-\varepsilon \boldsymbol{\Omega}} = \mathrm{e}^{\varepsilon \mathrm{ad}\, \boldsymbol{\Omega}} \hat{\boldsymbol{Y}} \\ &= \hat{\boldsymbol{Y}} + \varepsilon [\boldsymbol{\Omega}, \hat{Y}] + \tfrac{1}{2} \varepsilon^2 [\boldsymbol{\Omega}, [\boldsymbol{\Omega}, \hat{\boldsymbol{Y}}]] + \mathcal{O}\big(\varepsilon^3\big),\end{aligned}$$

where $\mathrm{ad}_{\boldsymbol{A}} \boldsymbol{B} = [\boldsymbol{A}, \boldsymbol{B}]$ is the adjoint operator of the Lie algebra $\mathfrak{so}(n)$.

We can show that

$$\|\boldsymbol{Y}_\varepsilon - \boldsymbol{N}\| = \|\hat{\boldsymbol{Y}} - \boldsymbol{N}\| + \varepsilon^2 \boldsymbol{\mu}^\top \boldsymbol{g}(\boldsymbol{\eta}) + \mathcal{O}\big(\varepsilon^3\big),$$

where the eigenvalues of the diagonal matrix $\hat{\boldsymbol{Y}} - \boldsymbol{N}$ are $\eta_k := d_{\pi_k} - \lambda_k$, $k = 1, 2, \ldots, n$,

$$\mu_k = \sum_{i \ne k} \frac{(d_{\pi_k} - d_{\pi_i})(\lambda_k - \lambda_i)}{d_{\pi_k} - d_{\pi_i} - \lambda_k + \lambda_i} \Omega_{k,i}^2,$$

and $\boldsymbol{g}$ is the gradient of the symmetric gauge,

$$g_k(\boldsymbol{x}) = \frac{\partial |\boldsymbol{x}|}{\partial x_k}, \qquad k = 1, 2, \ldots, n.$$

We deduce that the DBF (1.1) is isospectrally optimal if and only if

$$\boldsymbol{\mu}^\top \boldsymbol{g}(\boldsymbol{\eta}) \geq 0, \tag{2.2}$$

and the inequality is sharp for at least one $\boldsymbol{\Omega} \in \mathfrak{so}(n)$.

We can show this inequality holds strictly for the Schatten p-norms greater than 1 and non-strictly for $p = 1$.

3 Isospectral optimality for $n = 2$

Given (1.1) with $n = 2$, we have just three variables and, because the eigenvalues of $\boldsymbol{Y}(t)$ are invariant, have just one degree of freedom. No wonder, thus, that the solution can be easily represented explicitly. Further, the equations in this case are just Toda lattice equations in disguise, since a two by two matrix is of course tridiagonal. Letting

$$\boldsymbol{N} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \qquad \boldsymbol{Y}_0 = \begin{bmatrix} \alpha & \beta \\ \beta & \gamma \end{bmatrix},$$

we have $d_1 = \frac{1}{2}(\alpha+\gamma-\sigma)$, $d_2 = \frac{1}{2}(\alpha+\gamma+\sigma)$,where $\sigma = \sqrt{(\alpha-\gamma)^2 + 4\beta^2}$. Moreover, as can be easily verified by direct differentiation, the solution of (1.1) is

$$\begin{aligned} Y_{1,1}(t) &= \tfrac{1}{2}(\alpha+\gamma) + \tfrac{1}{2}\sigma \tanh\left(\sigma\delta t + \tfrac{1}{2}\log\frac{\sigma+\alpha-\gamma}{\sigma-\alpha+\gamma}\right), \\ Y_{1,2}(t) = Y_{2,1}(t) &= \frac{\sigma \mathrm{sgn}\beta}{2\cosh\left(\sigma\delta t + \frac{1}{2}\log\frac{\sigma+\alpha-\gamma}{\sigma-\alpha+\gamma}\right)}, \\ Y_{2,2}(t) &= \tfrac{1}{2}(\alpha+\gamma) - \tfrac{1}{2}\sigma \tanh\left(\sigma\delta t + \tfrac{1}{2}\log\frac{\sigma+\alpha-\gamma}{\sigma-\alpha+\gamma}\right), \end{aligned} \tag{3.1}$$

where $\delta = \lambda_1 - \lambda_2$.

Alternatively, we may express $\boldsymbol{Y}$ in terms of d_1 and d_2. A general matrix on the isospectral orbit is necessarily of the form

$$Y = \begin{bmatrix} \frac{1}{2}(d_1+d_2) - f & \pm\sqrt{\frac{1}{2}(d_1-d_2)^2 - f^2} \\ \pm\sqrt{\frac{1}{2}(d_1-d_2)^2 - f^2} & \frac{1}{2}(d_1+d_2) + f \end{bmatrix}, \tag{3.2}$$

where c is arbitrary, $|f| \leq \frac{1}{2}|d_1 - d_2|$. If $\boldsymbol{Y}$ is a function of t then so is f and substitution in (1.1) yields, regardless of the sign of $Y_{1,2} = Y_{2,1}$, the Riccati equation

$$f' = 2(\lambda_2 - \lambda_1)[\tfrac{1}{4}(d_2-d_1)^2 - f^2], \quad t \geq 0, \qquad f(0) = f_0, \tag{3.3}$$

where f_0 is given by $\boldsymbol{Y}_0$. Its solution is

$$f(t) = \tfrac{1}{2}|d_1 - d_2| \frac{\tilde{f} + \tanh[\sqrt{|\lambda_2 - \lambda_1| \cdot |d_2 - d_1|}t]}{1 + \tilde{f}\tanh[\sqrt{|\lambda_2 - \lambda_1| \cdot |d_2 - d_1|}t]}, \qquad t \geq 0,$$

where $\tilde{f} = 2f_0/|d_1 - d_2|$. Substitution in (3.2) results in an alternative formula for $\boldsymbol{Y}$.

We next explore global optimality for $n = 2$. We commence by noting from (3.2) that the eigenvalues of $\boldsymbol{Y}(t) - \boldsymbol{N}$ are

$$\kappa_\pm(t) = \tfrac{1}{2}(d_1 + d_2 - \lambda_1 - \lambda_2) \pm \sqrt{(d_2 - d_1)^2 + (\lambda_2 - \lambda_1)^2 - 4(\lambda_2 - \lambda_1) f(t)}.$$

Theorem 3 *In the case $n = 2$ all unitarily-invariant norms are globally optimal.*

Proof Let $q = (d_2 - d_1)^2 + (\lambda_2 - \lambda_1)^2$, $r = d_1 + d_2 - \lambda_1 - \lambda_2$, therefore

$$\kappa_\pm = \tfrac{1}{2}[r \pm \sqrt{q - 4(\lambda_2 - \lambda_1) f}]. \tag{3.4}$$

We let $\boldsymbol{\kappa} = [\kappa_+, \kappa_-]^\top$. By Theorem 1, any unitarily-invariant norm is a superposition of a symmetric gauge with the singular value map. Therefore, by symmetry of $\boldsymbol{Y} - \boldsymbol{N}$, it is true that

$$\|\boldsymbol{Y} - \boldsymbol{N}\| = |\boldsymbol{\kappa}|$$

for some symmetric gauge $|\cdot|$. Let $w : \mathbb{R}_+^2 \to \mathbb{R}_+$ be a function that coincides with the symmetric gauge in the nonnegative quadrant (thus, for the p-norm, $w(\boldsymbol{x}) = \sqrt{x_1^p + x_2^p}$). Assuming that w is differentiable for $\boldsymbol{x} \neq \boldsymbol{0}$ in the positive quadrant, we denote its derivatives by

$$w_1 = \frac{\partial w}{\partial x_1}, \qquad w_2 = \frac{\partial w}{\partial x_2},$$

and note that, $|\cdot|$ being a weakly-convex function (being a norm), it is necessarily true that

$$w_1(\boldsymbol{x}), w_2(\boldsymbol{x}), \frac{w_1(\boldsymbol{x}) - w_2(\boldsymbol{x})}{x_1 - x_2} \geq 0, \qquad x_1, x_2 \geq 0,\ x_1 + x_2 > 0,\ x_1 \neq x_2. \tag{3.5}$$

(In the case when $|\cdot|$ is not differentiable in the open positive quadrant, we may approximate it arbitrarily closely by a differentiable symmetric gauge.) Note that we do need to bother ourselves with the case $x_1 = x_2$, since, by (3.4), away from equilibria the eigenvalues $\kappa_\pm$ are distinct.

Since

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{Y} - \boldsymbol{N}\| = \frac{\mathrm{d}}{\mathrm{d}t}|\boldsymbol{\kappa}| = w_1 \frac{\mathrm{d}|\kappa_+|}{\mathrm{d}t} + w_2 \frac{\mathrm{d}|\kappa_-|}{\mathrm{d}t},$$

it is true for $\kappa_+\kappa_- \geq 0$ that

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{Y} - \boldsymbol{N}\| = -(w_1 - w_2)\frac{(\lambda_2 - \lambda_1) f'}{\sqrt{q - 4(\lambda_2 - \lambda_1) f}},$$

while in the case $\kappa_1\kappa_2 < 0$ we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\boldsymbol{Y} - \boldsymbol{N}\| = -(w_1 + w_2)\frac{(\lambda_2 - \lambda_1) f'}{\sqrt{q - 4(\lambda_2 - \lambda_1) f}}.$$

Because of (3.3), it follows that $(\lambda_2 - \lambda_1) f' \geq 0$. Hence, we can use (3.4) to argue that $\mathrm{d}\|\boldsymbol{Y} - \boldsymbol{N}\|/\mathrm{d}t \leq 0$. This implies global optimality. □

We turn our attention to isospectral optimality and distinguish, along the lines of the analysis in Section 2, between weak and strong optimality. Using our earlier terminology let

$$g_k(\boldsymbol{x}) = \frac{\partial |\boldsymbol{x}|}{\partial x_k}, \quad k = 1, 2, \qquad v(\boldsymbol{x}) = \frac{g_1(\boldsymbol{x}) - q_2(\boldsymbol{x})}{x_1 - x_2}.$$

We can show (see (Bloch & Iserles 2003*b*)):

Theorem 4 *Every unitarily-invariant norm is weakly isospectrally optimal, while if*

$$v(\boldsymbol{x}) > 0, \qquad x_1, x_2 > 0, \quad x_1 \neq x_2 \tag{3.6}$$

then the the norm is strongly isospectrally optimal.

4 Generalized Double Bracket Equations

4.1 The main equations

The remainder of the paper is further reported and expanded in (Bloch & Iserles 2003*a*). Let $\boldsymbol{Y}_0, \boldsymbol{N} \in \mathrm{Sym}(n)$ and suppose that $g : \mathrm{Sym}(n) \to \mathbb{R}$ is a smooth function. In particular, we are interested in the functions

$$g_m(\boldsymbol{X}) = \mathrm{tr}\, \boldsymbol{X}^m, \qquad m \in \mathbb{Z}_+.$$

We denote by $\boldsymbol{G} : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ the Jacobian of g. In particular,

$$[\boldsymbol{G}_m(\boldsymbol{X})]_{k,l} = \frac{\partial g_m(X)}{\partial \boldsymbol{X}_{k,l}}, \quad k, l = 1, 2, \ldots, n.$$

It is an easy matter to prove that

$$\boldsymbol{G}_m(X) = m\boldsymbol{X}^{m-1}, \qquad m \in \mathbb{Z}_+. \tag{4.7}$$

The *generalised double-bracket equation (GDBE)* is

$$\boldsymbol{Y}' = [\boldsymbol{Y}, [\boldsymbol{G}(\boldsymbol{Y} - \boldsymbol{N}), \boldsymbol{Y}]], \quad t \geq 0, \qquad \boldsymbol{Y}(0) = \boldsymbol{Y}_0. \tag{4.8}$$

In particular, we focus on $\boldsymbol{G} = \boldsymbol{G}_m$, in which case we drop the multiplicative factor m (which merely corresponds to reparametrisation) and consider

$$\boldsymbol{Y}' = [\boldsymbol{Y}, [(\boldsymbol{Y} - \boldsymbol{N})^{m-1}, \boldsymbol{Y}]], \quad t \geq 0, \qquad \boldsymbol{Y}(0) = \boldsymbol{Y}_0. \tag{4.9}$$

This, clearly, makes sense and is nontrivial only for $m \geq 2$.

We are interested in the stability and optimality of the equilibria of this generalized system. For $m \geq 3$ as we shall discuss below not all equilibria are diagonal. The diagonal equilibria are of great interest however and we begin by considering these.

4.2 Derivation of the linearised flow

Given

$$\boldsymbol{Y}' = [\boldsymbol{Y}, [\boldsymbol{Y}, (\boldsymbol{Y} - \boldsymbol{N})^m]]$$

the linearisation at $\boldsymbol{Y} = \boldsymbol{D} = \operatorname{diag} \boldsymbol{d}$ is

$$\begin{aligned}(\delta \boldsymbol{Y})' = {} & [\boldsymbol{D}, [\delta \boldsymbol{Y}, (\boldsymbol{D} - \boldsymbol{N})^m]] \\ & + [\boldsymbol{D}, [\boldsymbol{D}, \delta \boldsymbol{Y} (\boldsymbol{D} - \boldsymbol{N})^{m-1} + (\boldsymbol{D} - \boldsymbol{N}) \delta \boldsymbol{Y} (\boldsymbol{D} - \boldsymbol{N})^{m-2} + \cdots]].\end{aligned}$$

Writing this in components with $\eta_j = d_j - \lambda_j$ we have

$$(\delta y_{ij})' = \delta y_{ij} (\eta_j^m - \eta_i^m)(d_i - d_j) + (\delta y_{ij} \eta_j^{m-1} + \eta_i \delta y_{ij} \eta_j^{m-2} + \cdots)(d_i - d_j)^2$$

or

$$(\delta y_{ij})' = \delta y_{ij} \left[(\eta_j^m - \eta_i^m)(d_i - d_j) + \frac{\eta_j^m - \eta_i^m}{\eta_j - \eta_i} (d_i - d_j)^2 \right].$$

Hence for stability we require

$$(d_i - d_j)(\eta_j^m - \eta_i^m) \left(1 + \frac{d_i - d_j}{\eta_j - \eta_i} \right) < 0$$

or

$$(d_i - d_j)(\lambda_i - \lambda_j) \frac{\eta_j^m - \eta_i^m}{\eta_j - \eta_i} < 0.$$

5 Dynamics

5.1 The case $m = 3$

It is natural to define in this case $h(t) = \operatorname{tr} (\boldsymbol{Y} - \boldsymbol{N})^3$ and hope that this function is strictly monotone for all non-diagonal $\boldsymbol{Y}$. Numerical computations confirm this but, so far, analysis is *very* messy. The beauty of this approach is that, once we prove it, we can claim that the combinatorial problem has solution for every n, which is a pleasing, roundabout, homotopy argument for the solution of a discrete problem.

An alternative, rather weaker yet interesting, is to show that $h'(t) < 0$ for an open, nonempty set of initial values: at least this will imply the existence of an attractive fixed point, i.e. of a solution to the set of inequalities.

Let $\boldsymbol{\lambda} \in \mathbb{R}^n$ be given and assume that its components are distinct. We say that a vector $\boldsymbol{v} \in \mathbb{R}^n$ is *consistent with* $\boldsymbol{\lambda}$ if

$$(\lambda_k - \lambda_l)(v_k - v_l)(v_k + v_l - \lambda_k - \lambda_l) > 0, \qquad k, l = 1, 2, \ldots, n, \ k \neq l. \tag{5.1}$$

We denote the set of all such vectors by $\mathscr{C}_{\boldsymbol{\lambda}}$.

Lemma 5 *The set $\mathscr{C}_{\boldsymbol{\lambda}}$ is nonempty.*

Proof We commence by assuming that $\lambda_k > 0$, $k = 1, 2, \ldots, n$. Let $\rho > 0$ and set $v_k = \rho\lambda_k$, $k = 1, 2, \ldots, n$. Then

$$(\lambda_k - \lambda_l)(v_k - v_l)(v_k + v_l - \lambda_k - \lambda_l) = \rho(\rho - 1)(\lambda_k - \lambda_l)^2(\lambda_k + \lambda_l) > 0$$

and (5.1) is satisfied.

We next observe that, for any $\alpha \in \mathbb{R}$,

$$\boldsymbol{v} \in \mathscr{C}_{\boldsymbol{\lambda}} \qquad \Leftrightarrow \qquad \boldsymbol{v} + \alpha\mathbf{1} \in \mathscr{C}_{\boldsymbol{\lambda}+\alpha\mathbf{1}}.$$

We can thus always choose α so that $\lambda_k + \alpha > 0$, $k = 1, 2, \ldots, n$ and reduce everything to the framework of the previous paragraph. $\square$

The last lemma is less useful than it sounds, because we are not interested in arbitrary members of $\mathscr{C}_{\boldsymbol{\lambda}}$. Given $\boldsymbol{Y}_0 \in \mathrm{Sym}(n)$ and $\boldsymbol{N} = \mathrm{diag}\,\boldsymbol{\lambda}$, we interpret $\boldsymbol{v}$ as $\mathrm{diag}\,\boldsymbol{Y}_0$, therefore we wish $\boldsymbol{v}$ to live in the Horn–Schur polytope of $\boldsymbol{d} = \sigma(\boldsymbol{Y}_0)$: according to the Horn lemma, in that case, for every $\boldsymbol{v}$ in the polytope we can always find $\boldsymbol{Y}_0 \in \mathrm{Sym}(n)$ with $\sigma(\boldsymbol{Y}_0) = \boldsymbol{\lambda}$ and $\mathrm{diag}\,\boldsymbol{Y}_0 = \boldsymbol{v}$.

5.2 Stability Sectors in the Horn-Schur Polytope

What is going on in the Horn–Schur polytope? We start from the case $m = 3$, $n = 3$ and note that, since every point in the polytope is normal to $\mathbf{1}$, it lives in $\mathbb{R}^2$. Thus, plotting it, we can project it on a plane: Given $\boldsymbol{x} \in \mathbb{R}^3$ such that (as is the case in the HSP) $\boldsymbol{x}^\top\mathbf{1} = c$, where $c = \boldsymbol{d}^\top\mathbf{1}$, we plot the coordinates

$$\mathbb{R}^2 \ni \boldsymbol{z} = \begin{bmatrix} \frac{1}{2}(1+\sqrt{3}) & \frac{1}{2}(-1+\sqrt{3}) & 0 \\ \frac{1}{2}(-1+\sqrt{3}) & \frac{1}{2}(1+\sqrt{3}) & 0 \end{bmatrix} (\boldsymbol{x} - \tfrac{1}{3}c\mathbf{1}).$$

Our first collection of plots takes two vectors, $\boldsymbol{d}, \boldsymbol{\lambda} \in \mathbb{R}^3$ and colours differently the set of all points in the HS polytope according to the following scheme: Take a point $\boldsymbol{y}$ therein and check for which permutation $\boldsymbol{\eta} = \boldsymbol{\pi}(\boldsymbol{\lambda})$ of $\boldsymbol{\lambda}$ it is true that

$$(\eta_k - \eta_l)(y_k - y_l)(y_k + y_l - \eta_k - \eta_k) \geq 0, \qquad k, l = 1, 2, 3.$$

We use the following colour code, consistent with RGB colouring:

$$\begin{aligned} \boldsymbol{\pi} &= [1,2,3]: && \text{red,} \\ \boldsymbol{\pi} &= [1,3,2]: && \text{yellow,} \\ \boldsymbol{\pi} &= [2,3,1]: && \text{green,} \\ \boldsymbol{\pi} &= [3,2,1]: && \text{cyan,} \\ \boldsymbol{\pi} &= [3,1,2]: && \text{blue,} \\ \boldsymbol{\pi} &= [2,1,3]: && \text{magenta.} \end{aligned}$$

In monochrome, “colours” correspond to different shades of gray, ordered from the lightest to the darker as

$$\text{yellow} \prec \text{cyan} \prec \text{green} \prec \text{magenta} \prec \text{red} \prec \text{blue}.$$

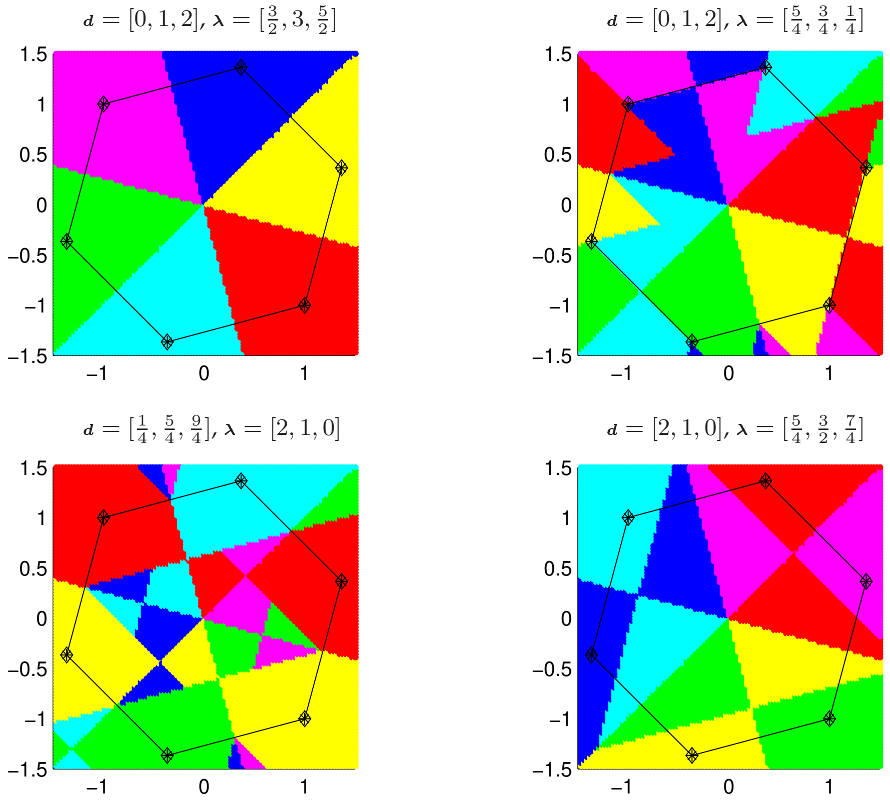

Figure 1: Colouring of the HSP.

This scheme of permutations is consistent with 'travelling' along the faces of the HS polytope.

The meaning of the experiment is as follows: the red points are arranged consistently with the stable attractor for $N = \operatorname{diag}[\lambda_1, \lambda_2, \lambda_3]$, the magenta points consistently with $N = \operatorname{diag}[\lambda_2, \lambda_1, \lambda_3]$ and so on. Of course, it does not mean (or, at least, we did not prove it, unlike the case $m = 2$, and it might well be false) that, once you solve the ODE with an initial condition in the red region (and with $N = \operatorname{diag}[\lambda_1, \lambda_2, \lambda_3]$) then you *must* tend to the corresponding vertex. All it means that if a vertex is bordered completely by a red region and we start near enough to it, attractivity of the fixed point will bring us to the vertex in question.

In Fig. 1 we display four such plots, for different choices of $\boldsymbol{d}$ and $\boldsymbol{\lambda}$.

Since the λ_ks are distinct, the colour changes once one of six equalities holds for $\boldsymbol{\eta}$,

$$\eta_k = \eta_l, \qquad \eta_k + \eta_l = \lambda_k + \lambda_l, \qquad 1 \le k < l \le 3.$$

Recall that everything has been rotated in Fig. 1 to a plane, hence we need to express the above equalities as linear equations in the (x, y) plane. Moreover, the one consequence of being in the HS polytope is that $\mathbf{1}^\top \boldsymbol{\eta} = c$ for $c = \mathbf{1}^\top \boldsymbol{d}$.

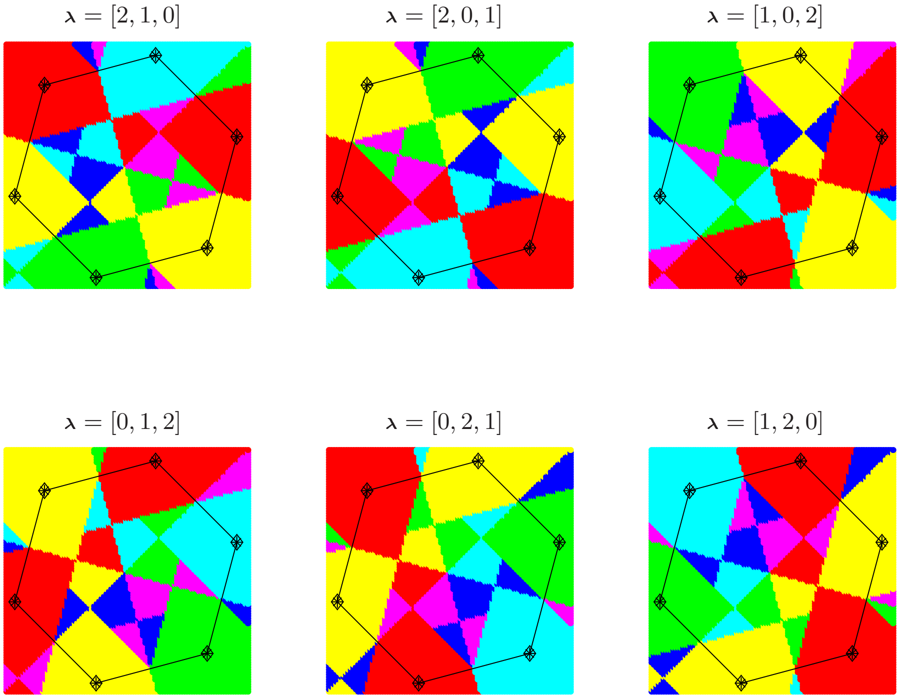

Figure 2: $\boldsymbol{d} = [\frac{1}{4}, \frac{5}{4}, \frac{9}{4}]$ and different permutations of $\lambda = [2, 1, 0]$.

For example, the condition $\eta_2 = \eta_3$, in tandem with $\eta_1+\eta_2+\eta_3 = c$, after rotation, yields

$$x = \tfrac{1}{4}[(3+\sqrt{3})\eta_1 - (1+3\sqrt{3})c], \qquad y = \tfrac{1}{4}[(-3+\sqrt{3})\eta_1 + (1-3\sqrt{3})c].$$

Eliminating η_1 from the above equations results in the line

$$y(x) = (\sqrt{3}-2)[x + 8(1+\sqrt{3})c].$$

In general, we obtain the following lines along which colour might change:

$$\begin{aligned}
\eta_1 = \eta_2 : &\qquad y(x) = x,\\
\eta_1 = \eta_3 : &\qquad y(x) = -(2+\sqrt{3})[x + 4(\sqrt{3}-1)c],\\
\eta_2 = \eta_3 : &\qquad y(x) = (\sqrt{3}-2)[x + 8(1+\sqrt{3})c],\\
\eta_1 + \eta_2 = \lambda_1 + \lambda_2 : &\qquad y(x) = -x - 2\sqrt{3}c + \sqrt{3}(\lambda_1+\lambda_2),\\
\eta_1 + \eta_3 = \lambda_1 + \lambda_3 : &\qquad y(x) = (2-\sqrt{3})[x - (6+2\sqrt{3})(\lambda_1+\lambda_3)],\\
\eta_2 + \eta_3 = \lambda_2 + \lambda_3 : &\qquad y(x) = (2+\sqrt{3})[x + (6-2\sqrt{3})(\lambda_2+\lambda_3)].
\end{aligned}$$

Note that the first three lines are always parallel to a face of the (rotated) HS polytope. Moreover, the *slopes* of all six lines are independent of the choice of $\boldsymbol{d}$ and $\boldsymbol{\lambda}$, all that changes is the affine shift.

In Fig. 2 we display the coloured HS polytope for fixed $\boldsymbol{d}$ and all permutations of $\boldsymbol{\lambda}$. Several symmetries stare us in the face. For example, any permutation of the form $(a, b, c) \mapsto (a, c, b)$ corresponds to red$\mapsto$magenta, blue$\mapsto$green and cyan$\mapsto$yellow. There are other symmetries, which might include both colour swaps and reflections (or rotations?) of the picture. Are they universal or special for this particular choice of $\boldsymbol{d}$ and $\boldsymbol{\lambda}$? The jury is out.

6 General Equilibria

As mentioned earlier, it is entirely possible for the matrices $\boldsymbol{Y} \in \mathrm{Sym}(n)$ and $\boldsymbol{N} = \mathrm{diag}\,\boldsymbol{\lambda}$ to be such that $[(\boldsymbol{Y}-\boldsymbol{N})^2, \boldsymbol{Y}] = O$, yet $(\boldsymbol{Y}-\boldsymbol{N})^2$ is *not* diagonal. The very important consequence if that *we might have fixed points – even attractive fixed points – inside the HS polytope!*

6.1 Nontrivial solutions: the solution for $n = 3$

We consider now solutions to the equilibrium condition

$$[\boldsymbol{N}^2, \boldsymbol{Y}] = [\boldsymbol{N}, \boldsymbol{Y}^2] \tag{6.1}$$

The idea is to commence by specifying the *diagonal* of $\boldsymbol{Y} \in \mathrm{Sym}(3)$. Thus, suppose that

$$\boldsymbol{Y} = \begin{bmatrix} \alpha & a & b \\ a & \beta & c \\ b & c & \gamma \end{bmatrix},$$

where α, β, γ are fixed and we seek to determine a, b, c so that (6.1) is satisfied.

(6.1) is equivalent to

$$(\lambda_k + \lambda_l) Y_{k,l} = (\boldsymbol{Y}^2)_{k,l}, \qquad 1 \le k < l \le n \tag{6.2}$$

(in our case $n = 3$) and the latter can be used to find the conditions on a, b, c: they are

$$\begin{aligned} (\lambda_1 + \lambda_2) a &= (\alpha + \beta) a + cb, \\ (\lambda_1 + \lambda_3) b &= (\alpha + \gamma) b + ac, \\ (\lambda_2 + \lambda_3) c &= (\beta + \gamma) c + ab. \end{aligned}$$

An obvious solution of the above equations is $a = b = c = 0$, which corresponds to a diagonal matrix $\boldsymbol{Y}$ (which, needless to say, obeys (6.1)). Being interested in nontrivial solutions, we exclude this case, whereby

$$a^2(\lambda_1 + \lambda_2 - \alpha - \beta) = b^2(\lambda_1 + \lambda_3 - \alpha - \gamma) = c^2(\lambda_2 + \lambda_3 - \beta - \gamma) = abc.$$

Note that in the present case $a, b, c \neq 0$. We obtain

$$a = (-1)^{k_a} b \sqrt{\frac{\lambda_1 + \lambda_3 - \alpha - \gamma}{\lambda_1 + \lambda_2 - \alpha - \beta}},$$

$$c = (-1)^{k_c} b \sqrt{\frac{\lambda_1 + \lambda_3 - \alpha - \gamma}{\lambda_2 + \lambda_3 - \beta - \gamma}},$$

where $k_a, k_b \in \{0,1\}$. Moreover,

$$abc = (\lambda_1 + \lambda_3 - \alpha - \gamma)b^2 = (-1)^{k_a+k_c} \frac{\lambda_1 + \lambda_3 - \alpha - \gamma}{\sqrt{(\lambda_1 + \lambda_2 - \alpha - \beta)(\lambda_2 + \lambda_3 - \beta - \gamma)}} b^3.$$

Therefore, we deduce that

$$\begin{aligned} a &= (-1)^{k_a} \sqrt{(\lambda_1 + \lambda_3 - \alpha - \gamma)(\lambda_2 + \lambda_3 - \beta - \gamma)}, \\ b &= (-1)^{k_a+k_c} \sqrt{(\lambda_1 + \lambda_2 - \alpha - \beta)(\lambda_2 + \lambda_3 - \beta - \gamma)}, \\ c &= (-1)^{k_c} \sqrt{(\lambda_1 + \lambda_2 - \alpha - \beta)(\lambda_1 + \lambda_3 - \alpha - \gamma)}. \end{aligned} \tag{6.3}$$

This exhausts all possible solutions, except that we must ensure that they are real. This will happen iff the three numbers

$$\lambda_1 + \lambda_2 - \alpha - \beta, \lambda_1 + \lambda_3 - \alpha - \gamma, \lambda_2 + \lambda_3 - \beta - \gamma$$

are all of the same sign. Therefore, (α, β, γ) must lie in one of two cones: either

$$\begin{aligned} \alpha + \beta &\geq \lambda_1 + \lambda_2, \\ \alpha + \gamma &\geq \lambda_1 + \lambda_3, \\ \beta + \gamma &\geq \lambda_2 + \lambda_3 \end{aligned}$$

or

$$\begin{aligned} \alpha + \beta &\leq \lambda_1 + \lambda_2, \\ \alpha + \gamma &\leq \lambda_1 + \lambda_3, \\ \beta + \gamma &\leq \lambda_2 + \lambda_3. \end{aligned}$$

Letting $\bar{\alpha} = \alpha - \lambda_1$, $\bar{\beta} = \beta - \lambda_2$, $\bar{\gamma} = \gamma - \lambda_3$, the first cone, $\mathcal{C}$,say, becomes

$$\bar{\alpha} + \bar{\beta} \geq 0, \quad \bar{\alpha} + \bar{\gamma} \geq 0, \quad \bar{\beta} + \bar{\gamma} \geq 0,$$

while the second cone is $-\mathcal{C}$. Therefore the condition for the existence of a nontrivial solution is that

$$\operatorname{diag} Y \in \boldsymbol{\lambda} + \mathcal{C} \quad \text{or} \quad \operatorname{diag} Y \in \boldsymbol{\lambda} - \mathcal{C}. \tag{6.4}$$

Of course, the whole point is that $\operatorname{diag} Y \in \mathcal{HS}$, the Horn–Schur polytope. Thus, nontrivial solutions exist iff

$$\text{either} \quad \mathcal{HS} \cap \boldsymbol{\lambda} + \mathcal{C} \neq \emptyset \quad \text{or} \quad \mathcal{HS} \cap \boldsymbol{\lambda} - \mathcal{C} \neq \emptyset.$$

Recall that all $\boldsymbol{x} \in \mathcal{HS}$ lie on the plane $\mathbf{1}^\top \boldsymbol{x} = \mathbf{1}^\top \boldsymbol{d}$. The intersection of this plane with the union of the two cones is always nonempty, but this is not to say that the intersection of the HS polytope with the union of the cones is nonempty. The easiest special case is when $\mathbf{1}^\top \boldsymbol{d} = \mathbf{1}^\top \boldsymbol{\lambda}$, whereby the intersection of the plane with the cones is just the point set $\{\boldsymbol{\lambda}\}$. Therefore, in that case, either $\boldsymbol{\lambda} \in \mathcal{HS}$ and we have a nontrivial solution,[2] or $\boldsymbol{\lambda} \notin \mathcal{HS}$ and all solutions are trivial. For all other values of $\mathbf{1}^\top \boldsymbol{d}$ the set of nontrivial solutions in the HS polytope is itself a polytope. Moreover, every point in this polytope corresponds to four different solutions, with different sign configurations.

[2] Recall our assumption that the spectra are distinct, hence $\boldsymbol{\lambda}$ cannot be a vertex of the HS polytope.

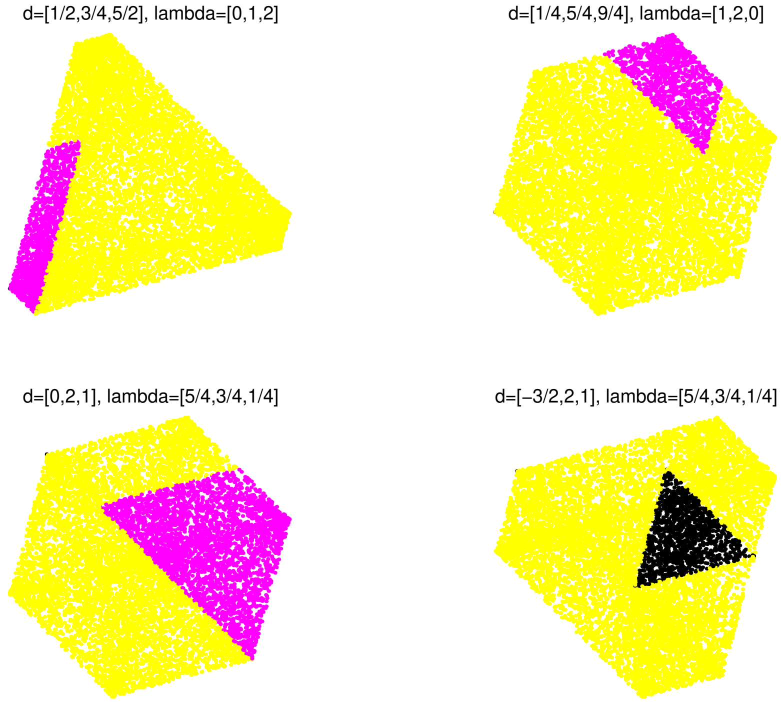

Figure 3: Different nontrivial solution regimes. Magenta denotes points in $\boldsymbol{\lambda} + \mathcal{C}$, black points are in $\boldsymbol{\lambda} - \mathcal{C}$ and yellow in neither cone. (In monochrome, "yellow" is the brightest and "black" the darkest.)

6.2 General $n \geq 2$

Let us commence by setting the stage for general n. Denoting $\operatorname{diag} \boldsymbol{Y} = \boldsymbol{x}$, it follows from (6.2) that

$$(\lambda_k+\lambda_l)Y_{k,l} = \sum_{i=1}^{k-1} Y_{i,k}Y_{i,l} + x_k Y_{k,l} + \sum_{i=k+1}^{l-1} Y_{k,i}Y_{i,l} + x_l Y_{k,l} + \sum_{i=l+1}^{n} Y_{k,i}Y_{l,i}, \quad 1 \leq k < l \leq n.$$

Therefore, letting

$$\mu_{k,l} = \lambda_k + \lambda_l - x_k - x_l, \qquad 1 \leq k < l \leq n,$$

we have

$$\mu_{k,l} Y_{k,l} = \sum_{i=1}^{k-1} Y_{i,k}Y_{i,l} + \sum_{i=k+1}^{l-1} Y_{k,i}Y_{i,l} + \sum_{i=l+1}^{n} Y_{k,i}Y_{l,i}, \quad 1 \leq k < l \leq n. \tag{6.5}$$

Consider now the general case. There is just a single solution for $n = 2$, namely a diagonal matrix ($Y_{1,2} = 0$). We denote the number of solutions in the $n \times n$ case by

ψ_n, therefore

$$\psi_2 = 1, \quad \psi_3 = 5, \quad \psi_4 = 25.$$

(It is tempting to conjecture that $\psi_n = 5^{n-2}$.)

Clearly, for $n \geq 3$ there are lower-dimensional problems hidden in the matrix and they can be obtained simply by taking rows and columns of off-diagonal zeros. Our real interest is in the solutions which cannot be obtained from lower-dimensional cases. We denote the number of such solutions by $\psi_n^\star$: $\psi_2^\star = 1$, $\psi_3^\star = 4$, $\psi_4^\star = 16$. It follows trivially that

$$\psi_n = \sum_{k=0}^{n-3} \binom{n}{k} \psi_{n-k}^\star + 1 = \sum_{k=3}^{n} \binom{n}{k} \psi_k^\star + 1 \tag{6.6}$$

(the "+1" corresponds to the diagonal solution). So, from now on, we concentrate only on the "truly n-dimensional" solutions.

We say that the symmetric matrix $E = (e_{k,l})_{k,l=1}^n$ is *tri-balanced* if

$$e_{k,l} \in \{0,1\}, \quad k,l = 1,\ldots,n \qquad \text{and} \qquad e_{i,k} + e_{i,l} = e_{k,l} \bmod 2, \quad k,l,1 = 1,\ldots,n$$

(note that automatically $e_{k,k} = 0$, but we will not use diagonal elements). Assuming that E is tri-balanced, we set

$$Y_{k,l} = (-1)^{e_{k,l}} \sqrt{(\sigma - \lambda_k + x_k)(\sigma - \lambda_l + x_l)}, \qquad k,l = 1,2,\ldots,n, \quad k \neq l, \tag{6.7}$$

where

$$\sigma = \frac{1}{n-2} \sum_{i=1}^{n} (\lambda_i - x_i) = \frac{1}{n-2} \mathrm{tr}\,(\boldsymbol{N} - \boldsymbol{Y}). \tag{6.8}$$

We claim that (6.7) represents a solution of (6.5), thereby of (6.1):

$$\begin{aligned}
\sum_{\substack{i=1 \\ i\neq k,l}}^{n} Y_{k,i} Y_{l,i} &= \sum_{\substack{i=1 \\ i\neq k,l}}^{n} (-1)^{e_{i,k}+e_{i,l}} (\sigma - \lambda_i + x_i) \sqrt{(\sigma - \lambda_k + x_k)(\sigma - \lambda_l + x_l)} \\
&= \sum_{\substack{i=1 \\ i\neq k,l}}^{n} (-1)^{e_{i,k}+e_{i,l}-e_{k,l}} (\sigma - \lambda_i + x_i) Y_{k,l} \\
&= \left[\sum_{i=1}^{n} (\sigma - \lambda_i + x_i) - 2\sigma + (\lambda_k + \lambda_l) - (x_k + x_l) \right] Y_{k,l} = \mu_{k,l} Y_{k,l}.
\end{aligned}$$

for all $k \neq l$. Therefore, for every tri-balanced matrix there corresponds a solution (6.7).

But how many tri-balanced matrices are there? Suppose that $e_{k,1} = e_{1,k}$ is given for all $k = 1,\ldots,n$ (of course, $e_{1,1} = 0$) and set

$$e_{k,l} = e_{1,k} + e_{1,l}, \qquad k,l = 1,2,\ldots,n.$$

It is easy to see that, with the above definition,

$$e_{k,i} + e_{i,l} = (e_{1,k} + e_{1,i}) + (e_{1,l} + e_{1,i}) = e_{1,k} + e_{1,l} = e_{k,l} \bmod 2,$$

hence we have a tri-balanced matrix. Moreover, no other choice of $e_{k,l}$ is tri-balanced. This is true because suppose that there exist k, l such that $e_{i,k} \neq e_{1,i}+e_{1,k} \bmod 2$. Then necessarily $e_{i,k} \neq e_{1,i} + e_{1,k} + 1 \bmod 2$, hence for every i

$$\begin{aligned}
& e_{i,k} + e_{i,l} = e_{1,k} + e_{1,l} \bmod 2 \\
\Rightarrow \quad & e_{i,l} = e_{1,l} - (e_{1,i} + e_{1,k} + 1) = e_{1,i} + e_{1,l} + 1 \bmod 2 \\
\overset{i=l}{\Rightarrow} \quad & 2e_{1,l} + 1 = 0 \bmod 2,
\end{aligned}$$

which is a contradiction. Therefore, we have precisely 2^{n-1} tri-balanced matrices and $\psi_n^\star \geq 2^{n-1}$.

We deduce that

$$\psi_n \geq 1 + \sum_{k=3}^{n} \binom{n}{k} 2^{k-1} = \tfrac{1}{2}(3^n + 1 - 2n^2).$$

A strange number, yet this upper bound coincides with ψ_n for $n = 1, 2, 3$. But does it represent *all* solutions of (6.1)?

Again, solutions are valid only if either

$$x_k \geq \lambda_k - \sigma, \qquad k = 1, 2, \ldots, n,$$

or

$$x_k \leq \lambda_k - \sigma, \qquad k = 1, 2, \ldots, n.$$

A cone and its negative, again.... Thus, we have a solution (given eigenvalues $\boldsymbol{d}$, rather than the diagonal $\boldsymbol{x}$) if and only and only if one of these cones intersects the HS polytope.

References

Bloch, A. M. (1990), 'Steepest descent, linear programming and Hamiltonian flows', *Contep. Maths AMS* **114**, 77–88.

Bloch, A. M. & Iserles, A. (2003*a*), Generalized double bracket equations, Technical Report to appear, DAMTP, University of Cambridge.

Bloch, A. M. & Iserles, A. (2003*b*), On the optimality of double-bracket flows, Technical Report to appear, DAMTP, University of Cambridge.

Bloch, A. M., Brockett, R. W. & Ratiu, T. (1990), 'A new formulation of the generalized Toda lattice equations and their fixed-point analysis via the moment map', *Bull. AMS* **23**, 447–456.

Bloch, A. M., Brockett, R. W. & Ratiu, T. (1992), 'Completely integrable gradient flows', *Comm. Math. Phys* **147**, 57–74.

Brockett, R. W. (1991), 'Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems', *Linear Algebra Appl.* **146**, 79–91.

Brockett, R. W. (1993), 'Differential geometry and the design of gradient algorithms', *Proc. Symp. Pure Maths* **54**, 69–92.

Calvo, M. P., Iserles, A. & Zanna, A. (1997), 'Numerical solution of isospectral flows', *Maths Comput.* **66**, 1461–1486.

Chu, M. T. & Driessel, K. R. (1991), 'Constructing symmetric nonnegative matrices with prescribed eigenvalues by differential equations', *SIAM J. Math. Anal.* **22**, 1372–1387.

Deift, P., Nanda, T. & Tomei, C. (1983), 'Differential equations for the symemtric eigenvalue problem', *SIAM J. Num. Anal.* **20**, 1–22.

Helmke, U. & Moore, J. (1994), *Optimization and Dynamical Systems*, Springer Verlag, New York.

Horn, R. A. & Johnson, C. R. (1991), *Topics in Matrix Analysis*, Cambridge University Press, Cambridge.

Iserles, A. (2002), 'On the discretization of double-bracket flows', *Found. Comp. Maths* **2**, 305–329.

Iserles, A., Munthe-Kaas, H. Z., Nørsett, S. P. & Zanna, A. (2000), 'Lie-group methods', *Acta Numerica* **9**, 215–365.

Lewis, A. S. & Overton, M. L. (1996), 'Eigenvalue optimization', *Acta Numerica* **5**, 149–190.

Moser, J. (1974), Finitely many mass points on the line under the influence of an exponential potential – an integrable system, *in* 'Dynamical Systems, Theory and Applications', number 38 *in* 'Springer Notes in Physics', Springer Verlag, Berlin, pp. 467–497.

Symes, W. W. (1980), 'Hamiltonian group actions and integrable systems', *Physica D* **1**, 339–374.

Symes, W. W. (1982), 'The QR algorithm and scattering for the nonperiodic Toda lattice', *Physica D* **4**, 275–280.

von Neumann, J. (1937), 'Some matrix inequalities and metrization of matrix space', *Tomsk University Review* **1**, 286–300. In *Collected Works*, Pergamon, Oxford, Volume IV (1962), 205–218.