Available online at www.sciencedirect.com
SCIENCE DIRECT®
LINEAR ALGEBRA AND ITS APPLICATIONS
ELSEVIER
Linear Algebra and its Applications 379 (2004) 353–379
www.elsevier.com/locate/laa

On computing canonical forms using flows

Kenneth R. Driessel

Department of Mathematics, Colorado State University, Fort Collins, CO, USA

Received 1 November 2002; accepted 21 July 2003

Submitted by T.Y. Tam

Abstract

Let G be a Lie group acting on a vector space $\mathscr{V}$. I say that a vector field $u : \mathscr{V} \to \mathscr{V}$ is *orbital* if for all p in $\mathscr{V}$, $u(p)$ is tangent to the orbit of p at p. Now let the vector space $\mathscr{V}$ and the vector space tangent to G at the identity have inner products. In this setting I define a simple map (which I call *quasi-projection*) which transforms any vector field on $\mathscr{V}$ into an orbital one. I use the quasi-projection map to define flows which compute canonical forms.

AMS classification: 15A21

Keywords: Canonical forms; Matrix group; Group action; Orbit; QR algorithm; QR flow; Toda flow; Dynamical system; Ordinary differential equation; Vector field; Differential geometry; Quasi-projection; Orthogonal similarity; Similarity; Orthogonal equivalence; Equivalence; Congruence; Simultaneous congruence

1. Introduction

We consider the following task: Compute canonical forms using flows. We begin our analysis of this task by clarifying the meaning of the phrases "canonical forms" and "computing using flows".

Let X be a set and let G be a group. A map $\mu : G \times X \to X$ is an *action* of G on X if it satisfies the following conditions for all $g, h \in G$ and $x \in X$:

(i) $\mu(g, \mu(h, x)) = \mu(gh, x)$,
(ii) $\mu(e, x) = x$,

E-mail address: driessel@math.colostate.edu (K.R. Driessel).

0024-3795/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.laa.2003.07.003

where e is the identity element of G. When μ can be easily inferred from the context I usually write $g * x$ in place of $\mu(g, x)$.

Let $\mathscr{G}(X)$ denote the group of invertible maps $\phi : X \to X$ of the set X into itself. If the map $\mu : G \times X \to X$ is a group action then the map $G \to \mathscr{G}(X) : g \mapsto (x \mapsto g * x)$ is a group homomorphism and the action map μ can easily be recovered from this homomorphism. As a consequence the phrase "transformation group" is sometimes used in place of the phrase "group action". (For example, Olver [15] uses "transformation group".)

We say that two points $x, y \in X$ are *equivalent* if there exists a group element that maps one to the other; in symbols, $\exists g \in G$, $g * x = y$. The equivalence class determined by an element $x \in X$ is called the *orbit* of x. It is easy to see that this orbit is the following set:

$$G * x := \{g * x : g \in G\}.$$

Olver [15, p. 58] defines "canonical form" as follows:

> In this context, a *canonical form* of an element $x \in X$ just means a distinguished, "simple" representative $x_0 \in G * x$ of the orbit containing x. Thus, a complete list of canonical forms can be identified with a list of orbits of the group, since each orbit must contain one (and, in an irredundant list, only one) canonical form, which thereby serves to distinguish the orbit. Of course, there is no uniquely specified canonical form, and some choice, usually based on one's aesthetic judgment of "simplicity", must be exercised.

(Birkhoff and Maclane [2] require that the canonical form of x be uniquely determined by x. This uniqueness requirement is often too stringent.)

We should also clarify the meaning of "flow". (Here I follow Ref. [11].) Let $\mathscr{V}$ be a vector space, let W be an open subset of $\mathscr{V}$ and let $f : W \to \mathscr{V}$ be a continuously differentiable function which may be called a *vector field* on W. We associate with this vector field the following (ordinary) differential equation: $x' = f(x)$. For each $y \in W$ there is a unique solution $\phi(t)$ with initial condition $\phi(0) = y$ defined on a maximal open interval $J(y) \subseteq \mathbb{R}$. We write $\phi(t, y)$ to indicate the dependence of the solution on the initial condition. Let $\Omega := \{(t, y) \in \mathbb{R} \times W : t \in J(y)\}$. Then the map $\Omega \to W : (t, y) \mapsto \phi(t, y)$ is called the *flow* of the differential equation. The set Ω is open in $\mathbb{R} \times W$ and the map $\phi : \Omega \to W$ is continuous.

Since flows are continuous, we cannot use them to compute arbitrary canonical forms. The map from elements to their canonical forms must be continuous. Consequently, we shall limit our attention to Lie group actions. (Here I follow Refs. [15,18].)

A *Lie group* G is a group that also carries the structure of a differentiable manifold such that the product map $G \times G \to G : (g, h) \mapsto gh$ and the inverse map $G \to G : g \mapsto g^{-1}$ are smooth (that is, infinitely differentiable). All of the examples of Lie groups that we shall discuss are well-known matrix groups. I use the language of Lie groups and manifolds only as an abstract framework to guide our discussion.

Let X be a smooth manifold and let G be a Lie group. A map $\mu : G \times X \to X$ is a *Lie group action* of G on X if μ is smooth and is a group action. The phrase "Lie transformation group" is also used in place of "Lie group action". (For example, Olver [15] uses the former phrase.)

Let $\mu : G \times X \to X$ be a Lie group action. Consider the orbit $G * p$ determined by $p \in X$. Note that for any $h \in G$ the map $G * p \to G * p : g * p \to h * (g * p)$ is a diffeomorphism of the orbit. Furthermore, given any two points of the orbit there is such a diffeomorphism taking the one point to the other. (In other words, the action of G on the orbit $G * p$ is "transitive".) Thus, if we view the orbit as a manifold then all its points look the same. We say that the manifold is "homogeneous". (For more on this topic see [18, Chapter 3, Section 5] and [8].) Later when we add more structure (namely a Riemannian 2-form) to the orbit we shall want to retain this homogeneity. We shall generally lose homogeneity when we introduce a vector field.

Here is a summary of the contents of this report. The next section which is entitled "Examples of canonical forms" contains a review of some examples of canonical forms. These examples will be treated further in a later section.

The section entitled "Quasi-projection" is the principal section of this report. I define the quasi-projection map there. Let me briefly preview that definition. Let G be a Lie group acting on a vector space $\mathscr{V}$. Then I say that a vector field $u : \mathscr{V} \to \mathscr{V}$ is *orbital* if for all p in $\mathscr{V}$, $u(p)$ is tangent to the orbit $G * p$ at p. Now let the vector space $\mathscr{V}$ and the space tangent to G at the identity have inner products. In this setting I define the quasi-projection map Π which transforms an arbitrary vector field $v : \mathscr{V} \to \mathscr{V}$ into a vector field $\Pi v : \mathscr{V} \to \mathscr{V}$ which is orbital. It follows that solutions of the differential equation $x' = \Pi v(x)$ stay on orbits.

In the section entitled "Applications of quasi-projection" I further treat the examples discussed earlier. In particular, I use the quasi-projection map to define flows which compute canonical forms.

Throughout the report I use some elementary facts about the Frobenius inner product. For the convenience of the reader I review this inner product in Appendix A.

2. Examples of canonical forms

Let us consider some examples. We shall limit our attention to classical subgroups of the group of invertible real matrices. (Curtis [4] discusses most of the matrix groups that we consider here.)

Example ($\mathrm{O}(n)$ *acting on* $\mathbb{R}^n$). Consider the vector space $\mathbb{R}^n$ of n-tuples of real numbers with the usual Euclidean inner product $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ which is defined by $\langle x, y \rangle := \sum x_i y_i$. Let $\mathrm{O}(n)$ denote the group of orthogonal linear transformations of this space. Recall that $\mathrm{O}(n)$ consists of the n-by-n real matrices Q with the property that the transpose Q^{T} of Q equals the inverse Q^{-1} of Q; in symbols, $\mathrm{O}(n) = \{Q \in \mathbb{R}^{n\times n} : QQ^{\mathrm{T}} = I\}$ where $\mathbb{R}^{n\times n}$ denotes the set of n-by-n real matrices. We use the natural multiplicative action of $\mathrm{O}(n)$ on $\mathbb{R}^n$:

$$\mathrm{O}(n) \times \mathbb{R}^n \to \mathbb{R}^n : (Q, x) \mapsto Qx.$$

Consider the orbit of a point p in $\mathbb{R}^n$:

$$\mathrm{O}(n) * p = \{Qp : Q \in \mathrm{O}(n)\}.$$

This orbit is the sphere centered at the origin with radius equal to the length $\|p\|$ of p. Let e be any unit vector. Then for a canonical form we can take $\|p\|e$. Note that the canonical forms are irredundant in this example.

Example ($\mathrm{O}(m)$ *acting on* $\mathbb{R}^{m\times n}$). In the previous example we had the orthogonal group acting on single vectors. In this example, the orthogonal group acts simultaneously on a number of vectors, namely, the columns of a matrix. Let $\mathbb{R}^{m\times n}$ denote the vector space of m-by-n matrices where $m \geqslant n$. Again we use the natural multiplicative action of $\mathrm{O}(m)$ on $\mathbb{R}^{m\times n}$:

$$\mathrm{O}(m) \times \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : (Q, M) \mapsto QM.$$

Consider the orbit of a matrix M in $\mathbb{R}^{m\times n}$:

$$\mathrm{O}(m) * M = \{QM : Q \in \mathrm{O}(m)\}.$$

Recall that any m-by-n matrix M with rank n can be written as the product of an orthogonal matrix Q and an upper triangular matrix $R : M = QR$. (This factoring corresponds to the Gram–Schmidt ortho-normalization process. See, for example, [16] or [14].) We can take upper triangular matrices as canonical forms. Note that these canonical forms provide a redundant list since the diagonal matrices $\mathrm{diag}(\pm 1, \pm 1, \ldots, \pm 1)$ are orthogonal.

Example ($\mathrm{Upper}(n)$ *acting on* $\mathbb{R}^n$). Let $\mathrm{Upper}(n)$ denote the group of invertible upper triangular matrices. We use the natural multiplicative action of $\mathrm{Upper}(n)$ on $\mathbb{R}^n$:

$$\mathrm{Upper}(n) \times \mathbb{R}^n \to \mathbb{R}^n : (R, x) \mapsto Rx.$$

Consider the orbit of a point p in $\mathbb{R}^n$:

$$\mathrm{Upper}(n) * p = \{Rp : R \in \mathrm{Upper}(n)\}.$$

Define the integer-valued function κ on $\mathbb{R}^n$ as follows:

$$\kappa := \mathbb{R}^n \to \{1, 2, \ldots, n\} : x \mapsto \min\{j : x_j \neq 0\}.$$

Note that the orbit of p consists of the elements $x \in \mathbb{R}^n$ satisfying the following condition: $x_{\kappa(p)} \neq 0$ and $\forall i < \kappa(p), x_i = 0$. For a canonical form of p we can take the coordinate vector $e_{\kappa(p)}$ where e_i is the vector which has its ith entry equal to 1 and its other entries equal to 0. Note that the canonical form for $e_1 + \varepsilon e_2$ is e_2 for all $\varepsilon \neq 0$. It follows that the canonical form map $p \mapsto e_{\kappa(p)}$ is not continuous.

Example (*Uni-Lower*(n) *acting on* $\mathbb{R}^{n\times n}$). Let Uni-Lower(n) denote the group of lower triangular n-by-n matrices with ones on the main diagonal. We use the natural multiplicative action of Uni-Lower(n) on $\mathbb{R}^{n\times n}$:

$$\text{Uni-Lower}(n) \times \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : (L, M) \mapsto LM.$$

Consider the orbit of a matrix M in $\mathbb{R}^{n\times n}$:

$$\text{Uni-Lower}(n) * M = \{LM : L \in \text{Uni-Lower}(n)\}.$$

Recall that most square matrices can be written as a product LR of a uni-lower triangular matrix and an upper triangular matrix R and that this factoring is usually unique. This factoring corresponds to Gaussian elimination. (See, for example, [16] or [14].) Consequently, we can usually take upper triangular matrices as canonical forms.

Example (O(n) *acting on* Sym(n), *orthogonal similarity*). Let Sym(n) denote the vector space of n-by-n symmetric matrices. We use the "similarity" action of the orthogonal group O(n) on the symmetric matrices

$$\mathrm{O}(n) \times \mathrm{Sym}(n) \to \mathrm{Sym}(n) : (Q, A) \mapsto QAQ^{\mathrm{T}}.$$

Consider the orbit of a symmetric matrix A:

$$\mathrm{O}(n) * A = \{QAQ^{\mathrm{T}} : Q \in \mathrm{O}(n)\}.$$

Recall that every symmetric matrix A has an independent set of orthonormal eigenvectors. This result is called the "spectral" theorem. (See, for example, [9,16] or [14].) We obtain the decomposition $A = QDQ^{\mathrm{T}}$ where Q is orthogonal and D is diagonal. The diagonal entries of D are the eigenvalues of A. Consequently, we can take diagonal matrices as canonical forms. Note that these canonical forms provide a redundant list since the eigenvalues can be rearranged.

Example (O$(m) \times$ O(n) *acting on* $\mathbb{R}^{m\times n}$, *orthogonal equivalence*). Let O$(m) \times$ O(n) denote the direct product of the orthogonal groups O(m) and O(n). We use the following action of this direct product on $\mathbb{R}^{m\times n}$:

$$\mathrm{O}(m) \times \mathrm{O}(n) \times \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : U \times V \times M \mapsto UMV^{\mathrm{T}}.$$

(When discussing a direct product, I sometimes write $a \times b$ instead of (a, b) to denote an ordered pair.) Consider the orbit of a matrix:

$$(\mathrm{O}(m) \times \mathrm{O}(n)) * M = \{UMV^{\mathrm{T}} : U \in \mathrm{O}(m), V \in \mathrm{O}(n)\}.$$

Recall that every matrix has a decomposition $M = UDV^{\mathrm{T}}$ where U and V are orthogonal and D is diagonal. This result is called the "singular value decomposition". (See, for example, [16] or [14].) The diagonal entries of D are called the "singular values" of M. Clearly we can take diagonal matrices as canonical forms.

Example (Gl(m) $\times$ Gl(n) *acting on* $\mathbb{R}^{m\times n}$, *equivalence*). Let Gl(m) denote the general linear group of m-by-m, invertible, real matrices. We use the following action of the direct product Gl(m) $\times$ Gl(n) on $\mathbb{R}^{m\times n}$:

$$\mathrm{Gl}(m)\times\mathrm{Gl}(n)\times\mathbb{R}^{m\times n}\to\mathbb{R}^{m\times n}:G\times H\times M\mapsto GMH^{-1}.$$

Consider the orbit of a matrix:

$$(\mathrm{Gl}(m)\times\mathrm{Gl}(n))*M=\{GMH^{-1}:G\in\mathrm{Gl}(m),\,H\in\mathrm{Gl}(n)\}.$$

Recall that every matrix M has a decomposition $M=GDH^{-1}$ where G and H are invertible matrices and D is a diagonal matrix with ones and zeros on its main diagonal. The number of ones equals the rank of M. Clearly we can take such diagonal matrices as canonical forms. However, note that the map $M\mapsto \operatorname{rank} M$ is not continuous. (For example, consider diagonal matrices with small diagonal values.)

Example (Gl(n) *acting on* Sym(n), *congruence*). We use the following action of Gl(n) on Sym(n):

$$\mathrm{Gl}(n)\times\mathrm{Sym}(n)\to\mathrm{Sym}(n):(G,M)\mapsto GMG^{\mathrm{T}}.$$

Consider the orbit of a symmetric matrix:

$$\mathrm{Gl}(n)*M=\{GMG^{\mathrm{T}}:G\in\mathrm{Gl}(n)\}.$$

Recall that every symmetric matrix M has a decomposition $M=GDG^{\mathrm{T}}$ where G is an invertible matrix and D is a diagonal matrix with diagonal entries equal to +1, 0 or -1. This result is associated with Sylvester's law of inertia and the signature of a symmetric matrix. (See, for example, [1,13] or [12].) Clearly we can take such diagonal matrices as canonical forms. However note that the map from symmetric matrices to such diagonal matrices is not continuous. (Consider, for example, diagonal matrices with small diagonal entries.)

Example (Gl(n) *acting on* Pos(n) $\times$ Sym(n), *simultaneous congruence*). Let Pos(n) denote the set of positive definite symmetric n-by-n matrices. Note that this set is open in the space of symmetric matrices. We use the following action of Gl(n) on the direct product Pos(n) $\times$ Sym(n):

$$\mathrm{Gl}(n)\times\mathrm{Pos}(n)\times\mathrm{Sym}(n)\to\mathrm{Pos}(n)\times\mathrm{Sym}(n):G\times A\times B\mapsto GAG^{\mathrm{T}}\times GBG^{\mathrm{T}}.$$

Consider the orbit of a pair:

$$\mathrm{Gl}(n)*(A\times B)=\{GAG^{\mathrm{T}}\times GBG^{\mathrm{T}}:G\in\mathrm{Gl}(n)\}.$$

Recall that for every pair $A\times B\in\mathrm{Pos}(n)\times\mathrm{Sym}(n)$ there exists an invertible matrix G and a diagonal matrix D such that $A=GG^{\mathrm{T}}$ and $B=GDG^{\mathrm{T}}$. (See, for example, [1, Chapter 4, Section 12] or [13, Chapter 2, Section 15].) Clearly we can take the pairs $I\times D$ where I is the identity matrix and D is diagonal as canonical forms.

Example (Gl(n) *acting on* $\mathbb{R}^{n\times n}$, *similarity*). We use the following "similarity" action of the general linear group Gl(n) on n-by-n real matrices:

$$\mathrm{Gl}(n) \times \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : (G, M) \mapsto GMG^{-1}.$$

Consider an orbit:

$$\mathrm{Gl}(n) * M = \{GMG^{-1} : G \in \mathrm{Gl}(n)\}.$$

Let $a := (a_1, a_2, \ldots, a_n) \in \mathbb{R}^n$ be an n-tuple of real numbers. Then the *companion matrix* $C(a)$ associated with a is defined as follows:

$$C(a) := \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_1 \\ 1 & 0 & 0 & \cdots & 0 & -a_2 \\ 0 & 1 & 0 & \cdots & 0 & -a_3 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -a_{n-1} \\ 0 & 0 & 0 & \cdots & 1 & -a_n \end{pmatrix}.$$

Recall that most matrices M (for example, matrices with distinct eigenvalues) are similar to a companion matrix; in other words, we have the decomposition $M = GC(a)G^{-1}$. (See, for example, [12] or [13] or [2].) Consequently, we can usually take companion matrices as canonical forms.

3. Quasi-projection

Let G be a Lie group, let $\mathscr{V}$ be an inner product space and let $\mu : G \times \mathscr{V} \to \mathscr{V}$ be a Lie group action. Recall that I say a vector field $u : \mathscr{V} \to \mathscr{V}$ is *orbital* if it satisfies the following condition: at every point p in $\mathscr{V}$, the vector $u(p)$ is tangent to the orbit of p; in symbols, $\forall p \in \mathscr{V}, u(p) \in \mathrm{Tan}\cdot(G * p)\cdot p$, where $\mathrm{Tan}\cdot(G * p)\cdot p$ is the space tangent to the orbit $G * p$ at p. There are several ways to transform an arbitrary vector field on $\mathscr{V}$ into an orbital one. Orthogonal projection is the standard way. (Thorpe [17] uses orthogonal projection.) I want to describe another way which I call "quasi-projection". But first let us review orthogonal projection.

Let $A : \mathscr{U} \to \mathscr{V}$ be a linear map between two inner product spaces. Let $A^* : \mathscr{V} \to \mathscr{U}$ be the linear map adjoint to A. (Halmos [9] uses this notation.) Recall that the *adjoint* map A^* is defined as follows: For $y \in \mathscr{V}$, A^*y is the unique element of $\mathscr{U}$ satisfying $\forall x \in \mathscr{U}, \langle x, A^*y\rangle = \langle Ax, y\rangle$. A linear map $A : \mathscr{V} \to \mathscr{V}$ on an inner product space is *self-adjoint* if it equals its adjoint: $A = A^*$.

The following result is well-known. (See, for example, [14, Section 5.5] or [16, Section 3.2].)

Proposition. *If* $A : \mathscr{U} \to \mathscr{V}$ *is an injective linear map from one inner product space to another then the "normal equation"* $A^*Ax = A^*b$ *has a unique solution which is given by* $(A^*A)^{-1}A^*b$. *Furthermore, the linear map* $P := A(A^*A)^{-1}A^*$ *on* $\mathscr{V}$ *has the following properties*:

(1) *The range of* P *equals the range of* A: Range P = Range A.
(2) *The kernel of* P *equals the orthogonal complement of the range of* A: Kernel P = (Range $A)^{\perp}$.
(3) *The map* P *is the projection of* $\mathscr{V}$ *onto* Range A *along* (Range $A)^{\perp}$ *which corresponds to the decomposition* $\mathscr{V}$ = Range $A \oplus$ (Range $A)^{\perp}$. *In particular*, $P^2 = P$ *and* $P^* = P$.

The map P is called the *projection map associated with the least squares problem* $Ax \sim b$. The projection vector Pb is the element of the range of A which is closest to b in the least squares sense.

It is often difficult to directly use the projection map $P := A(A^*A)^{-1}A^*$. If A is not injective then the inverse of A^*A does not exist. Even when A is injective it is often difficult to compute $(A^*A)^{-1}$—for example, if A is ill-conditioned or the dimension of the vector space $\mathscr{V}$ is large. We can often avoid these difficulties by using the linear map $AA^* : \mathscr{V} \to \mathscr{V}$ instead of the projection map $A(A^*A)^{-1}A : \mathscr{V} \to \mathscr{V}$. For the map AA^* we have the following analog of the last proposition.

Proposition. *Let* $A : \mathscr{U} \to \mathscr{V}$ *be a linear map between two inner product spaces. Then the map* $AA^* : \mathscr{V} \to \mathscr{V}$ *has the following properties*:

(1) *The range of* AA^* *equals the range of* A: Range(AA^*) = Range A.
(2) *The kernel of* AA^* *equals the orthogonal complement of the range of* A: Kernel(AA^*) = (Range $A)^{\perp}$.
(3) *The map* AA^* *is self-adjoint.*

Proof. Here is a proof of property (1). It is obvious that Range $AA^* \subseteq$ Range A. We want to see the other inclusion. Consider any element Ax in the range of A. Let $x = u + v$ where $u \in$ (Kernel $A)^{\perp}$ and $v \in$ Kernel A. Since Range A^* = (Kernel $A)^{\perp}$ we have $Ax = Au$ is an element of the range of AA^*. Here is a proof of property (3):

$$\langle u, AA^*v\rangle = \langle A^*u, A^*v\rangle = \langle A^{**}A^*u, v\rangle = \langle AA^*u, v\rangle \text{ since } A^{**} = A.$$

Finally we consider property (2). By (1) and (3) we have

$$(\text{Range } A)^{\perp} = (\text{Range}(AA^*))^{\perp} = \text{Kernel}(AA^*)^* = \text{Kernel}(AA^*). \qquad \square$$

I call the map $AA^* : \mathscr{V} \to \mathscr{V}$ the *quasi-projection map associated with the least squares problem* $Ax \sim b$.

Remark. We can compare the projection $P := A(A^*A)^{-1}A^*$ and the quasi-projection AA^* as follows. Since the restriction AA^* : Range $A \to$ Range A of AA^* to the range of A is self-adjoint, we can find an orthonormal basis of Range A consisting of eigenvectors: $AA^*v_i = \lambda_i v_i$ for $i = 1, 2, \ldots, m$ where m is the dimension of Range A. For any $v \in \mathscr{V}$ let $v = r + s$ where $r \in$ Range A and $s \in$ (Range $A)^{\perp}$. We have $Pv = r = \sum\langle r, v_i\rangle v_i$ and $AA^*v = AA^*r = \sum\langle r, v_i\rangle\lambda_i v_i$. Thus we see that

AA^* is a projection followed by an eigenvalue–eigenvector scaling. (Also note that $AA^*P = PAA^* = AA^*$.)

We now apply these ideas. As before, let $\mu : G \times \mathscr{V} \to \mathscr{V}$ be a Lie group acting on an inner product space $\mathscr{V}$. We want to transform an arbitrary vector field $v : \mathscr{V} \to \mathscr{V}$ on $\mathscr{V}$ into an orbital one. Let $p \in \mathscr{V}$ and consider the following map F_p of G onto the orbit of p:

$$F_p := G \to G * p : g \mapsto g * p.$$

Note that this map is smooth. (We regard the orbit as a smooth immersed submanifold of $\mathscr{V}$.) In particular, we have the differential of F_p at the identity e:

$$L_p := \mathrm{d}(F_p)e : \mathrm{Tan}\cdot G\cdot e \to \mathrm{Tan}\cdot(G * p)\cdot p.$$

This differential is a linear map from the space tangent to the Lie group G at the identity onto the space tangent to the orbit $G * p$ at p. The latter space may be regarded as a subspace of $\mathscr{V}$. In particular we can project $v(p)$ onto this tangent space.

I prefer to quasi-project rather than project. In order to do so I need to assume that the space $\mathrm{Tan}\cdot G\cdot e$ tangent to G at e has an inner product. Then we can regard the differential $L_p = \mathrm{d}(F_p)e$ as a linear map from one inner product space—namely, $\mathrm{Tan}\cdot G\cdot e$—into another inner product space—namely, $\mathscr{V}$, which is the ambient space of $\mathrm{Tan}\cdot(G * p)\cdot p$. We can then quasi-project the vector $v(p)$ onto the range—namely, $\mathrm{Tan}\cdot(G * p)\cdot p$—of the linear differential map. In this way we obtain the following orbital vector field on $\mathscr{V}$:

$$\Pi\cdot v := \mathscr{V} \to \mathscr{V} : p \mapsto (L_p \circ L_p^*)v(p).$$

I call this vector field the *quasi-projection of the vector field* v *induced by the Lie group action*. Note that the quasi-projection of v has the required tangential property:

$$\forall p \in \mathscr{V}, \quad \Pi\cdot v(p) \in \mathrm{Tan}\cdot(G * p)\cdot p.$$

Let us consider, in particular, the quasi-projection of a gradient vector field. Let $f : \mathscr{V} \to \mathbb{R}$ be a twice continuously differentiable real-valued (“objective”) function which we want to maximize. Then the *gradient vector field* $\nabla f : \mathscr{V} \to \mathscr{V}$ determined by f is defined by the derivative $\mathrm{D}f$ of f : $\langle \nabla f p, h\rangle := \mathrm{D}fph$. We consider the gradient flow determined by the objective function f:

$$x' = \nabla f x.$$

Recall that f is increasing on solutions of this differential equation since

$$(f \circ x)' = \mathrm{D}fxx' = \langle \nabla f x, x'\rangle = \langle \nabla f x, \nabla f x\rangle \geqslant 0.$$

Let us compare the quasi-projected gradient flow, or simply *quasi-gradient flow*, determined by f:

$$x' = \Pi(\nabla f)x.$$

We see that f is also increasing on solutions of this differential equation since

$$\begin{aligned}(f \circ x)' &= \langle \nabla f x, x' \rangle = \langle \nabla f x, \Pi(\nabla f)x \rangle \\ &= \langle \nabla f x, L_x L_x^* \nabla f x \rangle = \langle L_x^* \nabla f x, L_x^* \nabla f x \rangle \geqslant 0.\end{aligned}$$

4. Applications of quasi-projection

We can use quasi-projections to compute canonical forms. Let us consider some examples.

Example (O(n) *acting on* $\mathbb{R}^n$). As before we use the natural multiplicative action of O(n) on $\mathbb{R}^n$ defined by $(Q, x) \mapsto Qx$. Recall that orbits $\mathrm{O}(n) * p$ are spheres. Note that we have the following tangent spaces:

$$\begin{aligned}&\mathrm{Tan}\cdot\mathrm{O}(n)\cdot I = \mathrm{Skew}(n), \\ &\mathrm{Tan}\cdot(\mathrm{O}(n) * p)\cdot p = p^{\perp} = \mathrm{Skew}(n) * p := \{Kp : K \in \mathrm{Skew}(n)\},\end{aligned}$$

where Skew(n) denotes the vector space of n-by-n skew-symmetric matrices.

Let $p \in \mathbb{R}^n$ and consider the map

$$F_p := \mathrm{O}(n) \to \mathrm{O}(n) * p : Q \mapsto Qp.$$

For the differential of this map at I, we have

$$\mathrm{d}F_p I = \mathrm{Skew}(n) \to \mathrm{Skew}(n) * p : K \mapsto Kp.$$

Let L_p denote the linear map obtained from this differential by expanding the codomain space to $\mathbb{R}^n$; in symbols,

$$L_p := \mathrm{Skew}(u) \to \mathbb{R}^n : K \mapsto Kp.$$

If we use the Frobenius inner product on Skew(n) and the Euclidean inner product on $\mathbb{R}^n$, then

$$L_p^* = \mathbb{R}^n \to \mathrm{Skew}(n) : x \mapsto (1/2)(xp^{\mathrm{T}} - (xp^{\mathrm{T}})^{\mathrm{T}})$$

since, for any skew-symmetric matrix K, we have

$$\langle Kp, x \rangle = \langle K, xp^{\mathrm{T}} \rangle = \langle K, (1/2)(xp^{\mathrm{T}} - (xp^{\mathrm{T}})^{\mathrm{T}}) \rangle.$$

In other words, if we regard $\mathrm{d}F_p I$ as a map from Skew(n) into $\mathbb{R}^n$, we have

$$(\mathrm{d}F_p I)^* \cdot x = (1/2)(xp^{\mathrm{T}} - (xp^{\mathrm{T}})^{\mathrm{T}}).$$

We compute the composition of $\mathrm{d}F_p I$ and its adjoint:

$$\begin{aligned}((\mathrm{d}F_p I) \circ (\mathrm{d}F_p I)^*)x &= \mathrm{d}F_p I((1/2)(xp^{\mathrm{T}} - (xp^{\mathrm{T}})^{\mathrm{T}})) \\ &= (1/2)(xp^{\mathrm{T}} - px^{\mathrm{T}})p = (1/2)(\langle p, p \rangle x - \langle p, x \rangle p).\end{aligned}$$

This result should be compared with the usual projection onto the tangent space:

$$\mathrm{Proj}\cdot(p^{\perp})\cdot x = x - \frac{\langle x, p \rangle}{\langle p, p \rangle} p.$$

The quasi-projection Πv of any vector field $v : \mathbb{R}^n \to \mathbb{R}^n$ is the following map:

$$\mathbb{R}^n \to \mathbb{R}^n : p \mapsto (\langle p, p\rangle v(p) - \langle p, v(p)\rangle p).$$

Recall that for the canonical forms we can take the vectors on the ray determined by any unit vector e. We can easily compute these canonical forms by means of the following quasi-gradient flow. We define the objective function f by means of the distance to e:

$$f := \mathbb{R}^n \to \mathbb{R} : x \mapsto (1/2)\langle x - e, x - e\rangle.$$

We easily calculate that the derivative of f at p is the following linear function:

$$\mathrm{D}fp = \mathbb{R}^n \to \mathbb{R} : h \mapsto \langle p - e, h\rangle.$$

The gradient vector field ∇f determined by f is

$$\nabla f = \mathbb{R}^n \to \mathbb{R}^n : p \mapsto p - e.$$

We quasi-project this vector field to get the following orbital quasi-gradient vector field:

$$\Pi(\nabla f) = \mathbb{R}^n \to \mathbb{R}^n : p \mapsto \langle p, e\rangle p - \langle p, p\rangle e$$

since $\Pi(\nabla f)p = (1/2)(\langle p, p\rangle(p - e) - \langle p, p - e\rangle p) = (1/2)(\langle p, e\rangle p - \langle p, p\rangle e)$. We compute by means of the negative quasi-gradient; that is, we use the following differential equation (in which the constant $1/2$ is omitted):

$$x' = \langle x, x\rangle e - \langle x, e\rangle x.$$

The only equilibrium points of this differential equation are multiples of e. Clearly the stable equilibrium points are in canonical form.

Example ($\mathrm{O}(m)$ *acting on* $\mathbb{R}^{m\times n}$). As before we use the natural multiplicative action of $\mathrm{O}(m)$ on $\mathbb{R}^{m\times n}$ defined by $(Q, X) \mapsto QX$. Note that we have the following tangent spaces:

$$\mathrm{Tan}\cdot\mathrm{O}(m)\cdot I = \mathrm{Skew}(m),$$
$$\mathrm{Tan}\cdot(\mathrm{O}(m) * M)\cdot M = \mathrm{Skew}(m) * M.$$

Let $P \in \mathbb{R}^{m\times n}$ and consider the map

$$F_P := \mathrm{O}(m) \to \mathrm{O}(m) * P : Q \mapsto QP.$$

For the differential of this map at I, we have

$$\mathrm{d}F_P I = \mathrm{Skew}(m) \to \mathrm{Skew}(m) * P : K \mapsto KP.$$

Let L_P denote the linear map obtained from this differential by expanding the codomain space to $\mathbb{R}^{m\times n}$; in symbols,

$$L_P := \mathrm{Skew}(m) \to \mathbb{R}^{m\times n} : K \mapsto KP.$$

If we use the Frobenius inner product on $\mathrm{Skew}(m)$ and $\mathbb{R}^{m\times n}$, then

$$L_P^* = \mathbb{R}^{m\times n} \to \mathrm{Skew}(m) : X \mapsto (1/2)(XP^{\mathrm{T}} - (XP^{\mathrm{T}})^{\mathrm{T}})$$

since

$$\langle KP, X\rangle = \langle K, XP^{\mathrm{T}}\rangle = \langle K, (1/2)(XP^{\mathrm{T}} - (XP^{\mathrm{T}})^{\mathrm{T}})\rangle.$$

In other words, if we regard $\mathrm{d}F_P I$ as a map from $\mathrm{Skew}(m)$ into $\mathbb{R}^{m\times n}$, we have

$$(\mathrm{d}F_P I)^*{\cdot}X = (1/2)(XP^{\mathrm{T}} - (XP^{\mathrm{T}})^{\mathrm{T}}).$$

We compute the composition of $\mathrm{d}F_P I$ and its adjoint:

$$\begin{aligned}((\mathrm{d}F_P I)\circ(\mathrm{d}F_P I)^*)X &= \mathrm{d}F_P I((1/2)(XP^{\mathrm{T}} - PX^{\mathrm{T}}))\\ &= (1/2)(XP^{\mathrm{T}} - PX^{\mathrm{T}})P.\end{aligned}$$

The quasi-projection $\Pi{\cdot}v$ of any vector field $v : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is the following map:

$$\mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : P \mapsto (1/2)(v(P)P^{\mathrm{T}} - P(v(P))^{\mathrm{T}})P.$$

Recall that for the canonical forms, we can take upper triangular matrices. We can compute the canonical forms by means of the following quasi-gradient flow. We define the objective function f by means of the size of the upper triangular part:

$$f := \mathbb{R}^{m\times n} \to \mathbb{R} : X \mapsto (1/2)\langle X_{\mathrm{u}}, X_{\mathrm{u}}\rangle,$$

where X_{u} is the upper triangular part of X (including the diagonal of X). For example, when $m > n$, we take

$$X_{\mathrm{u}} := \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n}\\ 0 & x_{22} & \cdots & x_{2n}\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & x_{nn}\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots\\ 0 & 0 & \cdots & 0 \end{pmatrix}.$$

We easily calculate that the derivative of f at P is the following linear function:

$$\mathrm{D}f P = \mathbb{R}^{m\times n} \to \mathbb{R} : H \mapsto \langle P_{\mathrm{u}}, H\rangle.$$

The gradient vector field ∇f is

$$\nabla f = \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : P \mapsto P_{\mathrm{u}}.$$

We quasi-project this vector field to get the following orbital, quasi-gradient vector field:

$$\Pi(\nabla f) = \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : P \mapsto (1/2)(P_{\mathrm{u}}P^{\mathrm{T}} - PP_{\mathrm{u}}^{\mathrm{T}})P.$$

We compute by means of the negative quasi-gradient; that is, we use the following differential equation (in which the constant $1/2$ is omitted):

$$X' = (XX_{\mathrm{u}}^{\mathrm{T}} - X_{\mathrm{u}}X^{\mathrm{T}})X.$$

Note that upper triangular matrices are equilibrium points of this flow. Clearly they are stable. I do not know if any of the other equilibrium points are stable.

Remark. Since we are using the Frobenius inner product on $\mathbb{R}^{m\times n}$, we can view any orbit $\mathrm{O}(m) * M$ as a Riemannian manifold. The orbit thus viewed remains homogeneous since, for all $Q \in \mathrm{O}(m)$ and $J, K \in \mathrm{Skew}(m)$, we have

$$\langle JM, KM\rangle = \langle QJQ^{\mathrm{T}}QM, QKQ^{\mathrm{T}}QM\rangle.$$

Example (Upper(n) *acting on* $\mathbb{R}^n$). As before we use the natural multiplicative action of Upper(n) on $\mathbb{R}^n$ defined by $(R, x) \mapsto Rx$. Note that we have the following tangent spaces:

$$\mathrm{Tan}\cdot\mathrm{Upper}(n)\cdot I = \mathrm{upper}(n),$$
$$\mathrm{Tan}\cdot\mathrm{Upper}(n) * p\cdot p = \mathrm{upper}(n) * p := \{Xp : X \in \mathrm{upper}(n)\},$$

where upper(n) denotes the vector space of all n-by-n upper triangular matrices.

Let $p \in \mathbb{R}^n$ and consider the map

$$F_p := \mathrm{Upper}(n) \to \mathrm{Upper}(n) * p : R \mapsto Rp.$$

For the differential of this map at I we have

$$\mathrm{d}F_pI = \mathrm{upper}(n) \to \mathrm{upper}(n) * p : X \to Xp.$$

Let L_p denote the linear map obtained from this differential by expanding the codomain space to $\mathbb{R}^n$; in symbols,

$$L_p := \mathrm{upper}(n) \to \mathbb{R}^n : X \mapsto Xp.$$

If we use the Frobenius inner product in upper(n) and the euclidean inner product in $\mathbb{R}^n$ then

$$L_p^* = \mathbb{R}^n \to \mathrm{upper}(n) : x \mapsto (xp^{\mathrm{T}})_{\mathrm{u}},$$

where Y_{u} denotes the upper triangular part of the matrix Y. Note that

$$(xp^{\mathrm{T}})_{\mathrm{u}} = \begin{pmatrix} x_1p_1 & x_1p_2 & \cdots & x_1p_n \\ 0 & x_2p_2 & \cdots & x_2p_n \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_np_n \end{pmatrix}.$$

In other words, if we regard $\mathrm{d}F_pI$ as a map from upper(n) into $\mathbb{R}^n$, we have

$$(\mathrm{d}F_pI)^*\cdot x = (xp^{\mathrm{T}})_{\mathrm{u}}.$$

We compute the composition of $\mathrm{d}F_p I$ and its adjoint:

$$((\mathrm{d}F_p I)\circ(\mathrm{d}F_p I)^*)x = \mathrm{d}F_p I((xp^{\mathrm{T}})_{\mathrm{u}})$$
$$= (xp^{\mathrm{T}})_{\mathrm{u}} p = \begin{pmatrix} x_1(p_1^2+p_2^2+\cdots+p_n^2) \\ x_2(p_2^2+\cdots+p_n^2) \\ \vdots \\ x_n p_n^2 \end{pmatrix}.$$

We remarked before that, in this setting, the map from points to their canonical forms is not continuous. Consequently, we do not discuss the quasi-projection of any vector fields in this example.

Example (*Uni-Lower*(n) *acting on* $\mathbb{R}^{n\times n}$). As before let Uni-Lower(n) denote the group of lower triangular matrices with ones on the main diagonal. As before we use the natural multiplicative action of this group on $\mathbb{R}^{n\times n}$ defined by $(L, M)\mapsto LM$. Note that we have the following tangent spaces:

$$\text{Tan}\cdot\text{Uni-Lower}(n)\cdot I = \text{uni-lower}(n),$$
$$\text{Tan}\cdot(\text{Uni-Lower}(n)*M)\cdot M = \text{uni-lower}(n)*M := \{XM : X \in \text{uni-lower}(n)\},$$

where uni-lower(n) denotes the vector space of n-by-n strictly lower triangular matrices (which have zeros on the main diagonal).

Let $P \in \mathbb{R}^{n\times n}$ and consider the following map:

$$F_P := \text{Uni-Lower}(n) \to \text{Uni-Lower}(n) * P : L \mapsto LP.$$

For the differential of this map, we have

$$\mathrm{d}F_P I = \text{uni-lower}(n) \to \text{uni-lower}(n) * P : X \to XP.$$

Let L_P denote the linear map obtained from this differential by expanding the co-domain space to $\mathbb{R}^{n\times n}$; in symbols,

$$L_P := \text{uni-lower}(n) \to \mathbb{R}^{n\times n} : X \mapsto XP.$$

If we use the Frobenius inner product on uni-lower(n) and $\mathbb{R}^{n\times n}$ then

$$L_P^* = \mathbb{R}^{n\times n} \to \text{uni-lower}(n) : Y \mapsto (YP^{\mathrm{T}})_{\mathrm{l}}$$

since $\langle XP, Y\rangle = \langle X, YP^{\mathrm{T}}\rangle = \langle X, (YP^{\mathrm{T}})_{\mathrm{l}}\rangle$ where Z_{l} denotes the strictly lower triangular part of a matrix Z. In other words, if we regard $\mathrm{d}F_P I$ as a map from uni-lower(n) into $\mathbb{R}^{n\times n}$, we have

$$(\mathrm{d}F_P I)^*\cdot Y = (YP^{\mathrm{T}})_{\mathrm{l}}.$$

We compute the composition of $\mathrm{d}F_P I$ and its adjoint:

$$((\mathrm{d}F_P I)\circ(\mathrm{d}F_P I)^*)Y = \mathrm{d}F_P I((YP^{\mathrm{T}})_{\mathrm{l}}) = (YP^{\mathrm{T}})_{\mathrm{l}}P.$$

The quasi-projection $\Pi\cdot v$ of any vector field $v : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ is the following map:

$$\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : P \mapsto (v(P)P^{\mathrm{T}})_1 P.$$

Recall that for the canonical form we can take the upper triangular matrices. We can compute the canonical form by means of the following quasi-gradient flow. We define the objective function f by means of the size of the strictly lower triangular part:

$$f := \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : Y \mapsto (1/2)\langle Y_1, Y_1\rangle.$$

We want to minimize this objective function. We easily calculate that the derivative of f at P is the following linear function:

$$\mathrm{D}f\,P = \mathbb{R}^{n\times n} \to \mathbb{R} : H \mapsto \langle P_1, H\rangle.$$

The gradient vector field ∇f is

$$\nabla f = \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : P \mapsto P_1.$$

We quasi-project this vector field to get the following quasi-gradient vector field:

$$\Pi(\nabla f) = \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : P \mapsto (P_1 P^{\mathrm{T}})_1 P.$$

We compute by means of the negative quasi-gradient; that is, we use the following differential equation:

$$X' = -(X_1 X^{\mathrm{T}})_1 X.$$

Note that upper triangular matrices are equilibrium points of this flow. Clearly they are stable. I do not know if any of the other equilibrium points are stable.

Example ($\mathrm{O}(n)$ *acting on* $\mathrm{Sym}(n)$). As before we use the similarity action of $\mathrm{O}(n)$ on the space $\mathrm{Sym}(n)$ of symmetric matrices defined by $(Q, X) \mapsto QXQ^{\mathrm{T}}$. Recall that the orbits are isospectral surfaces. Note that we have the following tangent spaces:

$$\mathrm{Tan}\cdot\mathrm{O}(n)\cdot I = \mathrm{Skew}(n),$$
$$\mathrm{Tan}\cdot(\mathrm{O}(n) * A)\cdot A = \{[K, A] : K \in \mathrm{Skew}(n)\}.$$

Let $M \in \mathrm{Sym}(n)$ and consider the map

$$F_M := \mathrm{O}(n) \to \mathrm{Sym}(n) : Q \mapsto QMQ^{\mathrm{T}}.$$

For the differential of this map at I, we have

$$\mathrm{d}F_M I = \mathrm{Tan}\cdot\mathrm{O}(n)\cdot I \to \mathrm{Tan}\cdot(\mathrm{O}(n) * M)\cdot M : K \mapsto [K, M].$$

Let L_M denote the linear map obtained from this differential by expanding the codomain space to $\mathrm{Sym}(n)$. We use the Frobenius inner product on $\mathrm{Skew}(n)$ and $\mathrm{Sym}(n)$. Then

$$L_M^* = \mathrm{Sym}(n) \to \mathrm{Skew}(n) : X \mapsto [X, M]$$

since

$$\langle [K, M], X\rangle = \langle K, [X, M^{\mathrm{T}}]\rangle = \langle K, [X, M]\rangle.$$

In other words, if we regard the differential $\mathrm{d}F_M I$ as a map from Skew(n) into Sym(n), then we have

$$(\mathrm{d}F_M I)^* \cdot X = [X, M].$$

We compute the composition of the differential and its adjoint:

$$((\mathrm{d}F_M I) \circ (\mathrm{d}F_M I)^*)X = \mathrm{d}F_M I[X, M] = [[X, M], M].$$

The quasi-projection $\Pi \cdot v$ of any vector field $v : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ is the following map:

$$\mathrm{Sym}(n) \to \mathrm{Sym}(n) : M \mapsto [[v(M), M], M].$$

Recall that for the canonical forms we can take diagonal matrices. We can compute the canonical forms by means of the following quasi-gradient flows.

We can use an objective function f defined by means of the size of the diagonal of a matrix. For $X \in \mathrm{Sym}(n)$ let X_d be the diagonal part of X; in symbols,

$$X_\mathrm{d} := \begin{pmatrix} x_{11} & 0 & \cdots & 0 \\ 0 & x_{22} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & x_{nn} \end{pmatrix}.$$

We define f as follows:

$$f := \mathrm{Sym}(n) \to \mathbb{R} : X \mapsto (1/2)\langle X_\mathrm{d}, X_\mathrm{d} \rangle.$$

We want to maximize f. Note that all the matrices on an orbit have the same Frobenius norm. It follows that maximizing the size $\langle X_\mathrm{d}, X_\mathrm{d} \rangle$ of the diagonal part is equivalent to minimizing the size $\langle X - X_\mathrm{d}, X - X_\mathrm{d} \rangle$ of the off-diagonal part; that is, maximizing f on an orbit is equivalent to finding the matrix on the orbit closest to the subspace of diagonal matrices.

We easily calculate that the derivative of f at M is the following linear function:

$$\mathrm{D}fM = \mathrm{Sym}(n) \to \mathbb{R} : H \mapsto \langle M_\mathrm{d}, H \rangle.$$

In particular, note that $\langle M_\mathrm{d}, H \rangle = \langle M_\mathrm{d}, H_\mathrm{d} \rangle$. The gradient vector field ∇f determined by f is

$$\nabla f = \mathrm{Sym}(n) \to \mathrm{Sym}(n) : M \mapsto M_\mathrm{d}.$$

We quasi-project this vector field to get the following orbital, quasi-gradient vector field:

$$\Pi(\nabla f) = \mathrm{Sym}(n) \to \mathrm{Sym}(n) : M \mapsto [[M_\mathrm{d}, M], M].$$

We compute by means of the quasi-gradient; that is, we use the following differential equation:

$$X' = [[X_\mathrm{d}, X], X].$$

Since $\langle [[E_\mathrm{d}, E], E], E_\mathrm{d} \rangle = \langle [E_\mathrm{d}, E], [E_\mathrm{d}, E] \rangle$, we see that a matrix E is an equilibrium point iff $[E_\mathrm{d}, E] = 0$. I classified these equilibrium points a number of years ago. (See [5,6].) In particular, the only stable equilibrium points are diagonal matrices.

We can use an alternative objective function f defined by means of the distance to a fixed symmetric matrix D:

$$f := \mathrm{Sym}(n) \to \mathbb{R} : X \mapsto (1/2)\langle X - D, X - D\rangle.$$

We easily calculate the derivative of f at M; we get the following linear map:

$$\mathrm{D} f M = \mathrm{Sym}(n) \to \mathbb{R} : H \mapsto \langle M - D, H\rangle.$$

The gradient vector field ∇f determined by f is

$$\nabla f = \mathrm{Sym}(n) \to \mathrm{Sym}(n) : M \mapsto M - D.$$

We quasi-project this vector field to get the following orbital quasi-gradient vector field:

$$\Pi(\nabla f) = \mathrm{Sym}(n) \to \mathrm{Sym}(n) : M \mapsto [[-D, M], M]$$

since $[\nabla f M, M] = [M - D, M] = -[D, M]$. We compute by means of the negative quasi-gradient; that is, we use the following differential equation:

$$X' = [[D, X], X].$$

We see that E is an equilibrium point iff $[D, E] = 0$ since $\langle[[D, E], E], D\rangle = \langle[D, E], [D, E]\rangle$. If D is a diagonal matrix with distinct diagonal entries then the equilibrium points are diagonal matrices.

Remark. Chu and Driessel [3] derived the last differential equation but did so by means of a more complicated procedure. Compare also Helmke and Moore [10].

Remark. We can use the Frobenius inner product to view any orbit $\mathrm{O}(n) * A$ as a Riemannian manifold. The orbit thus viewed remains homogeneous since, for all $Q \in \mathrm{O}(n)$ and $J, K \in \mathrm{Skew}(n)$ we have

$$\begin{aligned}\langle [J, M], [K, M]\rangle &= \langle Q[J, M]Q^{\mathrm{T}}, Q[K, M]Q^{\mathrm{T}}\rangle \\ &= \langle [QJQ^{\mathrm{T}}, QMQ^{\mathrm{T}}], [QKQ^{\mathrm{T}}, QMQ^{\mathrm{T}}]\rangle.\end{aligned}$$

Example ($\mathrm{O}(m) \times \mathrm{O}(n)$ *acting on* $\mathbb{R}^{m\times n}$). As before we use the action of $\mathrm{O}(m) \times \mathrm{O}(n)$ on $\mathbb{R}^{m\times n}$ defined by $U \times V \times M \mapsto UMV^{\mathrm{T}}$. Note that we have the following tangent spaces:

$$\mathrm{Tan}\cdot(\mathrm{O}(m) \times \mathrm{O}(n))\cdot(I \times I) = \mathrm{Skew}(m) \times \mathrm{Skew}(n),$$
$$\mathrm{Tan}\cdot((\mathrm{O}(m) \times \mathrm{O}(n)) * M)\cdot M = \{JM - MK : J \in \mathrm{Skew}(m), K \in \mathrm{Skew}(n)\}.$$

Let $M \in \mathbb{R}^{m\times n}$ and consider the map

$$F_M := \mathrm{O}(m) \times \mathrm{O}(n) \to \mathbb{R}^{m\times n} : U \times V \mapsto UMV^{\mathrm{T}}.$$

For the differential of this map at $I \times I$, we have

$$\mathrm{d}F_M(I \times I) = \mathrm{Tan}\cdot(\mathrm{O}(m) \times \mathrm{O}(n))\cdot(I \times I) \to \mathrm{Tan}\cdot((\mathrm{O}(m) \times \mathrm{O}(n)) * M)\cdot M : \\ J \times K \mapsto JM - MK.$$

Let L_M denote the linear map obtained from this differential by expanding the codomain space to $\mathbb{R}^{m\times n}$. We use the Frobenius inner product on Skew(m), Skew(n) and $\mathbb{R}^{m\times n}$. Then

$$L_M^* : \mathbb{R}^{m\times n} \to \mathrm{Skew}(m) \times \mathrm{Skew}(n) : X \mapsto (1/2)(XM^{\mathrm{T}} - MX^{\mathrm{T}}) \times (X^{\mathrm{T}}M - M^{\mathrm{T}}X)$$

since

$$\begin{aligned}\langle JM - MK, X\rangle &= \langle J, XM^{\mathrm{T}}\rangle + \langle K, -M^{\mathrm{T}}X\rangle \\ &= \langle J, (1/2)(XM^{\mathrm{T}} - (XM^{\mathrm{T}})^{\mathrm{T}})\rangle \\ &\quad + \langle K, -(1/2)(M^{\mathrm{T}}X - (M^{\mathrm{T}}X)^{\mathrm{T}})\rangle \\ &= \langle J \times K, (1/2)((XM^{\mathrm{T}} - MX^{\mathrm{T}}) \times (X^{\mathrm{T}}M - M^{\mathrm{T}}X))\rangle.\end{aligned}$$

In other words, if we regard the differential $\mathrm{d}F_M(I \times I)$ as a map from Skew$(m) \times$ Skew(n) into $\mathbb{R}^{m\times n}$ then we have

$$(\mathrm{d}F_M(I \times I))^*\cdot X = (1/2)((XM^{\mathrm{T}} - MX^{\mathrm{T}}) \times (X^{\mathrm{T}}M - M^{\mathrm{T}}X)).$$

We compute the composition of the differential and its adjoint:

$$\begin{aligned}&((\mathrm{d}F_M(I \times I)) \circ (\mathrm{d}F_M(I \times I))^*)X \\ &\quad = \mathrm{d}F_M(I \times I)((1/2)((XM^{\mathrm{T}} - MX^{\mathrm{T}}) \times (X^{\mathrm{T}}M - M^{\mathrm{T}}X))) \\ &\quad = (1/2)((XM^{\mathrm{T}} - MX^{\mathrm{T}})M - M(X^{\mathrm{T}}M - M^{\mathrm{T}}X)) \\ &\quad = (1/2)(XM^{\mathrm{T}}M + MM^{\mathrm{T}}X - 2MX^{\mathrm{T}}M).\end{aligned}$$

The quasi-projection $\Pi\cdot v$ of any vector field $v : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is the following map:

$$\mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : M \mapsto (1/2)(v(M)M^{\mathrm{T}}M + MM^{\mathrm{T}}v(M) - 2Mv(M)^{\mathrm{T}}M).$$

Recall that for the canonical forms we can take diagonal matrices. We can compute the canonical forms by means of the following quasi-gradient flows.

We can use an objective function f defined by means of the size of the diagonal part of a matrix. For $X \in \mathbb{R}^{m\times n}$ let X_{d} be the diagonal part of X. We define f as follows:

$$f := \mathbb{R}^{m\times n} \to \mathbb{R} : X \mapsto \langle X_{\mathrm{d}}, X_{\mathrm{d}}\rangle.$$

We want to maximize f. Note that all the matrices on an orbit have the same Frobenius norm. It follows that maximizing the size $\langle X_{\mathrm{d}}, X_{\mathrm{d}}\rangle$ of the diagonal part is equivalent to minimizing the size of the off-diagonal part; that is, maximizing f on an orbit is equivalent to finding the matrix on the orbit closest to the subspace of diagonal matrices.

We easily compute that the derivatives of f at M is the following linear function:

$$\mathrm{D}fM = \mathbb{R}^{m\times n} \to \mathbb{R} : H \mapsto \langle M_\mathrm{d}, H\rangle.$$

In particular, note that $\langle M_\mathrm{d}, H\rangle = \langle M_\mathrm{d}, H_\mathrm{d}\rangle$. The gradient vector field ∇f determined by f is

$$\nabla f = \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : M \mapsto M_\mathrm{d}.$$

We quasi-project this vector field to get the following orbital, quasi-gradient vector field:

$$\Pi(\nabla f) = \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : M \mapsto (1/2)(M_\mathrm{d}M^\mathrm{T}M + MM^\mathrm{T}M_\mathrm{d} - 2MM_\mathrm{d}^\mathrm{T}M).$$

We compute by means of the quasi-gradient; that is, we use the following differential equation (in which the constant 1/2 is omitted):

$$X' = (X_\mathrm{d}X^\mathrm{T}X + XX^\mathrm{T}X_\mathrm{d} - 2XX_\mathrm{d}^\mathrm{T}X).$$

We see that E is an equilibrium point iff $E_\mathrm{d}^\mathrm{T}E = E^\mathrm{T}E_\mathrm{d}$ and $E_\mathrm{d}E^\mathrm{T} = EE_\mathrm{d}^\mathrm{T}$ since

$$\begin{aligned}
&2\langle (E_\mathrm{d}E^\mathrm{T} - EE_\mathrm{d}^\mathrm{T})E + E(E^\mathrm{T}E_\mathrm{d} - E_\mathrm{d}^\mathrm{T}E), E_\mathrm{d}\rangle\\
&\quad = 2\langle E_\mathrm{d}E^\mathrm{T} - EE_\mathrm{d}^\mathrm{T}, E_\mathrm{d}E^\mathrm{T}\rangle + 2\langle E^\mathrm{T}E_\mathrm{d} - E_\mathrm{d}^\mathrm{T}E, E^\mathrm{T}E_\mathrm{d}\rangle\\
&\quad = \|E_\mathrm{d}E^\mathrm{T} - EE_\mathrm{d}^\mathrm{T}\|^2 + \|E^\mathrm{T}E_\mathrm{d} - E_\mathrm{d}^\mathrm{T}E\|^2.
\end{aligned}$$

I derived this differential equation a number of years ago but by a more complicated procedure. (See [7].) I also classified the equilibrium points then. In particular, the only stable equilibrium points are diagonal matrices.

We can use an alternative objective function f defined by means of the distance to a fixed diagonal matrix D:

$$f := \mathbb{R}^{m\times n} \to \mathbb{R} : X \mapsto (1/2)\langle X - D, X - D\rangle.$$

We easily calculate the derivative of f at M; we get

$$\mathrm{D}fM = \mathbb{R}^{m\times n} \to \mathbb{R} : H \mapsto \langle M - D, H\rangle.$$

The gradient vector field ∇f determined by f is

$$\nabla f = \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : M \mapsto M - D.$$

We quasi-project this vector field to get the following orbital, quasi-gradient vector field:

$$\Pi(\nabla f) = \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : M \mapsto MD^\mathrm{T}M - (1/2)(DM^\mathrm{T}M + MM^\mathrm{T}D)$$

since

$$\begin{aligned}
&(1/2)((M - D)M^\mathrm{T}M + MM^\mathrm{T}(M - D) - 2M(M - D)^\mathrm{T}M)\\
&\quad = (1/2)(2MD^\mathrm{T}M - DM^\mathrm{T}M - MM^\mathrm{T}D).
\end{aligned}$$

We compute by means of the quasi-gradient; that is, we use the following differential equation (in which the constant $1/2$ is omitted):

$$X' = 2XD^\mathrm{T}X - DX^\mathrm{T}X - XX^\mathrm{T}D.$$

We see that E is an equilibrium point iff $ED^{\mathrm{T}} = DE^{\mathrm{T}}$ and $D^{\mathrm{T}}E = E^{\mathrm{T}}D$ since

$$\begin{aligned}&2\langle(ED^{\mathrm{T}} - DE^{\mathrm{T}})E + E(D^{\mathrm{T}}E - E^{\mathrm{T}}E), D\rangle\\&\quad = 2\langle ED^{\mathrm{T}} - DE^{\mathrm{T}}, DE^{\mathrm{T}}\rangle + 2\langle D^{\mathrm{T}}E - E^{\mathrm{T}}D, E^{\mathrm{T}}D\rangle\\&\quad = -(\|ED^{\mathrm{T}} - DE^{\mathrm{T}}\|^2 + \|D^{\mathrm{T}}E - E^{\mathrm{T}}D\|^2).\end{aligned}$$

Chu and Driessel [3] derived the last differential equation but by a more complicated procedure.

Example (Gl(m) × Gl(n) *acting on* $\mathbb{R}^{m\times n}$). As before we use the action of Gl(m) × Gl(n) on $\mathbb{R}^{m\times n}$ defined by $G \times H \times M \mapsto GMH^{-1}$. Note that we have the following tangent spaces:

$$\begin{aligned}&\mathrm{Tan}\cdot(\mathrm{Gl}(m) \times \mathrm{Gl}(n))\cdot(I \times I) = \mathbb{R}^{m\times m} \times \mathbb{R}^{n\times n},\\&\mathrm{Tan}\cdot((\mathrm{Gl}(m) \times \mathrm{Gl}(n)) * M = \{XM - MY : X \in \mathbb{R}^{m\times m}, Y \in \mathbb{R}^{n\times n}\}.\end{aligned}$$

Let $M \in \mathbb{R}^{m\times n}$ and consider the map

$$F_M := \mathrm{Gl}(m) \times \mathrm{Gl}(n) \to (\mathrm{Gl}(m) \times \mathrm{Gl}(n)) * M : (G \times H) \mapsto GMH^{-1}.$$

For the differential of this map at $(I \times I)$, we have

$$\begin{aligned}\mathrm{d}F_M(I\times I) = \mathrm{Tan}\cdot(\mathrm{Gl}(m)\times\mathrm{Gl}(n))\cdot(I\times I) \to \mathrm{Tan}\cdot((GL(m)\times\mathrm{Gl}(n))*M)\cdot M :\\ (X \times Y) \mapsto XM - MY.\end{aligned}$$

Let L_M denote the linear map obtained from this differential by expanding the codomain space to $\mathbb{R}^{m\times n}$; in symbols,

$$L_M := \mathbb{R}^{m\times m} \times \mathbb{R}^{n\times n} \to \mathbb{R}^{m\times n} : (X \times Y) \mapsto XM - MY.$$

We use the Frobenius inner product on $\mathbb{R}^{m\times m}$, $\mathbb{R}^{n\times n}$ and $\mathbb{R}^{m\times n}$. Then

$$L_M^* = \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} \times \mathbb{R}^{n\times n} : Z \mapsto ZM^{\mathrm{T}} \times (-M^{\mathrm{T}}Z)$$

since

$$\langle XM - MY, Z\rangle = \langle X, ZM^{\mathrm{T}}\rangle + \langle Y, -M^{\mathrm{T}}Z\rangle = \langle X \times Y, ZM \times (-M^{\mathrm{T}}Z)\rangle.$$

In other words, if we regard the differential $\mathrm{d}F_M(I \times I)$ as a map from $\mathbb{R}^{m\times m} \times \mathbb{R}^{n\times n}$ into $\mathbb{R}^{m\times n}$ then we have

$$(\mathrm{d}F_M(I \times I))^*\cdot Z = ZM \times (-M^{\mathrm{T}}Z).$$

We compute the composition of the differential and its adjoint:

$$\begin{aligned}((\mathrm{d}F_M(I \times I)) \circ (\mathrm{d}F_M(I \times I))^*)Z &= \mathrm{d}F_M(I \times I)(ZM^{\mathrm{T}} \times (-M^{\mathrm{T}}Z))\\&= ZM^{\mathrm{T}}M + MM^{\mathrm{T}}Z.\end{aligned}$$

The quasi-projection $\Pi\cdot v$ of a vector field $v : \mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n}$ is the following map:

$$\mathbb{R}^{m\times n} \to \mathbb{R}^{m\times n} : M \mapsto v(M)MM^{\mathrm{T}} + MM^{\mathrm{T}}v(M).$$

We remarked before that, in this setting, the map from points to their standard canonical forms (associated with rank) is not continuous. Consequently, we do not discuss any specific vector fields in this example.

Example (Gl(n) *acting on* Sym(n)). As before we use the action of Gl(n) on Sym(n) defined by $(G, M) \mapsto GMG^{\mathrm{T}}$. Note that we have the following tangent spaces:

$$\begin{aligned}&\mathrm{Tan}{\cdot}\mathrm{Gl}(n){\cdot}I = \mathbb{R}^{n\times n},\\ &\mathrm{Tan}{\cdot}(\mathrm{Gl}(n) * M){\cdot}M = \{XM + MX^{\mathrm{T}} : X \in \mathbb{R}^{n\times n}\}.\end{aligned}$$

Let $M \in \mathrm{Sym}(n)$ and consider the map

$$F_M := \mathrm{Gl}(n) \to \mathrm{Gl}(n) * M : G \mapsto GMG^{\mathrm{T}}.$$

For the differential of this map at I, we have

$$\mathrm{d}F_M I = \mathrm{Tan}{\cdot}\mathrm{Gl}(n){\cdot}I \to \mathrm{Tan}{\cdot}(\mathrm{Gl}(n) * M){\cdot}M : X \mapsto XM + MX^{\mathrm{T}}.$$

Let L_M denote the linear map obtained from this differential by expanding the codomain space to Sym(n); in symbols,

$$L_M := \mathbb{R}^{n\times n} \to \mathrm{Sym}(n) : X \mapsto XM + MX^{\mathrm{T}}.$$

We use the Frobenius inner product on $\mathbb{R}^{n\times n}$ and Sym(n). Then

$$L_M^* = \mathrm{Sym}(n) \to \mathbb{R}^{n\times n} : Y \mapsto 2YM$$

since

$$\begin{aligned}\langle XM + MX^{\mathrm{T}}, Y\rangle &= \langle X, YM\rangle + \langle X^{\mathrm{T}}, MY\rangle = \langle X, YM\rangle + \langle X, (MY)^{\mathrm{T}}\rangle\\ &= \langle X, 2YM\rangle.\end{aligned}$$

In other words, if we regard the differential $\mathrm{d}F_M I$ as a map from $\mathbb{R}^{n\times n}$ to Sym(n), then we have

$$(\mathrm{d}F_M I)^*{\cdot}Y = 2YM.$$

We compute the composition of the differential and its adjoint:

$$\begin{aligned}((\mathrm{d}F_M I) \circ (\mathrm{d}F_M I)^*)Y &= \mathrm{d}F_M I(2YM) = 2YMM + M(2YM)^{\mathrm{T}}\\ &= 2(YM^2 + M^2Y).\end{aligned}$$

The quasi-projection $\Pi{\cdot}v$ of any vector field $v : \mathrm{Sym}(n) \to \mathrm{Sym}(n)$ is the following map:

$$\mathrm{Sym}(n) \to \mathrm{Sym}(n) : M \to 2(v(M)M^2 + M^2v(M)).$$

We remarked before that, in this setting, the map from points to their standard canonical forms (associated with signature) is not continuous. Consequently, we do not discuss any specific vector fields in this example.

Example (Gl(n) *action on* Pos(n) $\times$ Sym(n)). As before we use the action of Gl(n) on the direct product Pos(n) $\times$ Sym(n) of the positive definite and the symmetric matrices defined by $(G, A \times B) \mapsto GAG^{\mathrm{T}} \times GBG^{\mathrm{T}}$. Note that we have the following tangent spaces:

$$\mathrm{Tan}{\cdot}\mathrm{Gl}(n){\cdot}I = \mathbb{R}^{n\times n},$$
$$\mathrm{Tan}{\cdot}(\mathrm{Gl}(n) * (A\times B)){\cdot}(A\times B) = \{(XA+AX^{\mathrm{T}})\times(XB+BX^{\mathrm{T}}) : X \in \mathbb{R}^{n\times n}\}.$$

For $P \in \mathrm{Pos}(n)$ and $M \in \mathrm{Sym}(n)$, consider the following map:

$$F_{P\times M} := \mathrm{Gl}(n) \to \mathrm{Pos}(n) \times \mathrm{Sym}(n) : G \mapsto GPG^{\mathrm{T}} \times GMG^{\mathrm{T}}.$$

For the differential of this map at I, we have

$$\mathrm{d}F_{P\times M}{\cdot}I = \mathrm{Tan}{\cdot}\mathrm{Gl}(n){\cdot}I \to \mathrm{Tan}{\cdot}(\mathrm{Gl}(n) * (P \times M)){\cdot}(P \times M) : \\ X \mapsto (XP + PX^{\mathrm{T}}) \times (XM + MX^{\mathrm{T}}).$$

Let $L_{P\times M}$ denote the map obtained from this differential by expanding the codomain space to Sym(n) $\times$ Sym(n); in symbols,

$$L_{P\times M} := \mathbb{R}^{n\times n} \to \mathrm{Sym}(n) \times \mathrm{Sym}(n) : X \mapsto (XP + PX^{\mathrm{T}}) \times (XM + MX^{\mathrm{T}}).$$

We use the Frobenius inner product on $\mathbb{R}^{n\times n}$ and Sym(n) $\times$ Sym(n). Then

$$L^*_{P\times M} = \mathrm{Sym}(n) \times \mathrm{Sym}(n) \to \mathbb{R}^{n\times n} : Y \times Z \mapsto 2(YP + ZM)$$

since, for every pair $Y \times Z$ of symmetric matrices, we have

$$\begin{aligned}
&\langle (XP + PX^{\mathrm{T}}) \times (XM + MX^{\mathrm{T}}), Y \times Z\rangle \\
&\quad = \langle XP + PX^{\mathrm{T}}, Y\rangle + \langle XM + MX^{\mathrm{T}}, Z\rangle \\
&\quad = \langle XP, Y\rangle + \langle PX^{\mathrm{T}}, Y\rangle + \langle XM, Z\rangle + \langle MX^{\mathrm{T}}, Z\rangle \\
&\quad = \langle X, YP^{\mathrm{T}}\rangle + \langle X^{\mathrm{T}}, P^{\mathrm{T}}Y\rangle + \langle X, ZM^{\mathrm{T}}\rangle + \langle X^{\mathrm{T}}, M^{\mathrm{T}}Z\rangle \\
&\quad = \langle X, YP\rangle + \langle X^{\mathrm{T}}, PY\rangle + \langle X, ZM\rangle + \langle X^{\mathrm{T}}, MZ\rangle \\
&\quad = \langle X, YP\rangle + \langle X, (PY)^{\mathrm{T}}\rangle + \langle X, ZM\rangle + \langle X, (MZ)^{\mathrm{T}}\rangle \\
&\quad = \langle X, YP + Y^{\mathrm{T}}P\rangle + \langle X, ZM + Z^{\mathrm{T}}M\rangle \\
&\quad = \langle X, 2YP\rangle + \langle X, 2ZM\rangle.
\end{aligned}$$

In other words, if we regard the differential $\mathrm{d}F_{P\times M}{\cdot}I$ as a map from $\mathbb{R}^{n\times n}$ to the product Sym(n) $\times$ Sym(n)then we have $(\mathrm{d}F_{P\times M}{\cdot}I)^*(Y \times Z) = 2(YP + ZM)$. We compute the composition of the differential and its adjoint:

$$\begin{aligned}
&((\mathrm{d}F_{P\times M}{\cdot}I) \circ (\mathrm{d}F_{P\times M}{\cdot}I)^*){\cdot}(Y \times Z) = \mathrm{d}F_{P\times M}{\cdot}I\,(2(YP + ZM)) \\
&\quad = 2((YP+ZM)P+P(YP+ZM)^{\mathrm{T}})\times 2((YP+ZM)M+M(YP+ZM)^{\mathrm{T}}) \\
&\quad = 2(YP^2 + ZMP + P^2Y + PMZ) \times 2(YPM + ZM^2 + MPY + M^2Z) \\
&\quad = 2(YP^2 + P^2Y + ZMP + PMZ)\times 2(YPM + MPY + ZM^2 + M^2Z).
\end{aligned}$$

Let v_1 and v_2 be two maps of Sym(n) $\times$ Sym(n) into Sym(n). Then the quasi-projection $\Pi{\cdot}v$ of any vector field $v : \mathrm{Sym}(n) \times \mathrm{Sym}(n) \to \mathrm{Sym}(n) \times \mathrm{Sym}(n)$ defined by $v(P \times M) := v_1(P \times M) \times v_2(P \times M)$ is given as follows:

$$
\begin{aligned}
\Pi\cdot v(P\times M) = {} & (v_1(P\times M)P^2 + P^2 v_1(P\times M) \\
& + v_2(P\times M)MP + PMv_2(P\times M)) \\
& \times 2(v_1(P\times M)PM + MPv_1(P\times M) \\
& + v_2(P\times M)M^2 + M^2 v_2(P\times M)).
\end{aligned}
$$

Recall that for the canonical form we can take pairs $I \times D$ where D is diagonal. We shall define a quasi-gradient flow which approaches these canonical forms. We can use an objective function f defined by means of the distance to pairs in canonical form. We define f as follows:

$$f := \mathrm{Sym}(n)\times\mathrm{Sym}(n)\to\mathbb{R} : P\times M \mapsto (1/2)(\langle P-I, P-I\rangle + \langle M_o, M_o\rangle)$$

where $X_o := X - X_d$ is the off-diagonal part of a matrix X. We easily calculate that the derivative of f at $P \times M$ is the following linear function:

$$\mathrm{D}f(P\times M) = \mathrm{Sym}(n)\times\mathrm{Sym}(n) \to \mathbb{R} : Y\times Z \mapsto \langle P-I, Y\rangle + \langle M_o, Z\rangle.$$

The gradient vector field ∇f is

$$\nabla f = \mathrm{Sym}(n)\times\mathrm{Sym}(n) \to \mathrm{Sym}(n)\times\mathrm{Sym}(n) : P\times M \mapsto (P-I)\times M_o.$$

We quasi-project this vector field to get the following orbital, quasi-gradient vector field:

$$
\begin{aligned}
\Pi(\nabla f) = {} & \mathrm{Sym}(n)\times\mathrm{Sym}(n) \to \mathrm{Sym}(n)\times\mathrm{Sym}(n) : P\times M \\
& \mapsto 2((P-I)P^2 + P^2(P-I) + M_o MP + PMM_o) \\
& \times 2((P-I)PM + MP(P-I) + M_o M^2 + M^2 M_o).
\end{aligned}
$$

We compute by means of the quasi-gradient; that is, we use the following system of differential equations (in which the constant 2 is omitted):

$$
\begin{aligned}
P' &= (P-I)P^2 + P^2(P-I) + M_o MP + PMM_o, \\
M' &= (P-I)PM + MP(P-I) + M_o M^2 + M^2 M_o.
\end{aligned}
$$

Note that canonical pairs $I \times D$ where D is diagonal are equilibrium points of this flow. Clearly they are stable. I do not know if any of the other equilibrium points are stable.

Example (Gl(n) *acting on* $\mathbb{R}^{n\times n}$). As before we use the similarity action of Gl(n) on $\mathbb{R}^{n\times n}$ which is defined by $(G, A) \mapsto GAG^{-1}$. Note that we have the following tangent spaces:

$$
\begin{aligned}
&\mathrm{Tan}\cdot\mathrm{Gl}(n)\cdot I = \mathbb{R}^{n\times n}, \\
&\mathrm{Tan}\cdot(\mathrm{Gl}(n) * A)\cdot A = \{[X, A] : X \in \mathbb{R}^{n\times n}\}.
\end{aligned}
$$

Let M be an n-by-n matrix and consider the following map:

$$F_M := \mathrm{Gl}(n) \to \mathbb{R}^{n\times n} : G \mapsto GMG^{-1}.$$

For the differential of this map at I, we have

$$\mathrm{d}F_M \cdot I = \mathrm{Tan}\cdot\mathrm{Gl}(n)\cdot I \mapsto \mathrm{Tan}\cdot(\mathrm{Gl}(n) * M)\cdot M : X \mapsto [X, M].$$

Let L_M denote the linear map obtained from this differential by expanding the co-domain to $\mathbb{R}^{n\times n}$; in symbols,

$$L_M := \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : X \mapsto [X, M].$$

We use the Frobenius inner product on $\mathbb{R}^{n\times n}$. Then

$$L_M^* = \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : Y \mapsto [Y, M^{\mathrm{T}}]$$

since $\langle [X, M], Y\rangle = \langle X, [Y, M^{\mathrm{T}}]\rangle$. In other words, if we regard the differential $\mathrm{d}F_M \cdot I$ as a map on $\mathbb{R}^{n\times n}$, then we have $(\mathrm{d}F_M\cdot I)^*\cdot Y = [Y, M^{\mathrm{T}}]$.

We compute the composition of the differential and its adjoint:

$$((\mathrm{d}F_M\cdot I) \circ (\mathrm{d}F_M\cdot I)^*)Y = \mathrm{d}F_M\cdot I[Y, M^{\mathrm{T}}] = [[Y, M^{\mathrm{T}}], M].$$

The quasi-projection $\Pi\cdot v$ of any vector field $v : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}$ is the following map:

$$\mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : M \mapsto [[v(M), M^{\mathrm{T}}], M].$$

Recall that for the canonical form we can take companion matrices. We shall define a quasi-gradient flow which approaches these canonical forms. We can use an objective function f defined by means of the distance to companion matrices. Let Z be the *lower shift matrix*:

$$Z := \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}.$$

For any matrix X let X_{c} denote the matrix obtained from X by zeroing the last column of X; in symbols,

$$X_{\mathrm{c}} := \begin{pmatrix} x_{1,1} & \cdots & x_{1,n-1} & 0 \\ x_{2,1} & \cdots & x_{2,n-1} & 0 \\ \vdots & \vdots & & \vdots \\ x_{n,1} & \cdots & x_{n,n-1} & 0 \end{pmatrix}.$$

We define the objective function f as follows:

$$f := \mathbb{R}^{n\times n} \to \mathbb{R} : X \mapsto (1/2)\langle X_{\mathrm{c}} - Z, X_{\mathrm{c}} - Z\rangle.$$

We want to minimize f. We easily calculate that the derivative of f at M is the following function:

$$\mathrm{D}f M = \mathbb{R}^{n\times n} \to \mathbb{R} : X \mapsto \langle M_{\mathrm{c}} - Z, X\rangle.$$

In particular, note that $\langle M_c - Z, X\rangle = \langle M_c - Z, X_c\rangle$. The gradient vector field ∇f determined by the objective function f is the following one:

$$\nabla f = \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : M \mapsto M_c - Z.$$

We quasi-project this vector field to get the following orbital, quasi-gradient vector field:

$$\Pi(\nabla f) = \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n} : M \mapsto [[M_c - Z, M^{\mathrm{T}}], M].$$

We compute by means of the quasi-gradient; that is, we use the following differential equation:

$$X' = [[X_c - Z, X^{\mathrm{T}}], X].$$

Note that

$$\langle [[X_c - Z, X^{\mathrm{T}}], X], X_c - Z\rangle = \langle [X_c - Z, X^{\mathrm{T}}], [X_c - Z, X^{\mathrm{T}}]\rangle.$$

Hence a matrix E is an equilibrium point of the differential equation iff $[E_c - Z, E^{\mathrm{T}}] = 0$. Note that companion matrices are equilibrium points of this flow. Clearly, they are stable. I do not know if any of the other equilibrium points are stable.

Acknowledgements

I wrote this report during the first half of the calendar year 2000 while visiting Iowa State University. I wish to thank the people who made this pleasant visit possible: Max Gunzburger (chair) and Irvin R. Hentzel.

I talked about this work at several seminars during 2000–2002. I lectured in the linear algebra seminar at Iowa State University (organized by Leslie Hogbin). I lectured in the algebra seminar at the University of Wyoming (organized by Sylvia Hobart); I visited the University of Wyoming during the 2000–2001 academic year; Bryan Shader arranged this visit. I gave a special series of lectures at Auburn University; I visited Auburn University in June, 2002; Peter Nylen arranged this visit.

In addition to people mentioned above, I wish to thank colleagues who spoke with me about this research and offered encouragement: Fred Furtado (University of Wyoming), Eric Moorhouse (University of Wyoming), Peter Olver (University of Minnesota) and Jack Conn (University of Minnesota).

I also with to thank Ruth De Boor (Iowa State University) for typing the original version of this report.

I revised this report in 2002 while affiliated with Colorado State University. I thank Rick Miranda for arranging this affiliation. In particular, I made a number of minor corrections. At the end of the section “Applications of quasi-projection”, I added the following two examples: “Gl(n) acting on Pos(n) $\times$ Sym(n)” and “ Gl(n) acting on $\mathbb{R}^{n\times n}$”.

Appendix A. The Frobenius inner product

Let $\mathbb{R}^{m\times n}$ denote the space of m-by-n real matrices. Recall that the *Frobenius inner product* on this space is defined by $\langle X, Y\rangle := \mathrm{Trace}(XY^{\mathrm{T}})$. (I use a superscript T to denote the transpose operation on matrices.) In terms of matrix entries we have $\langle X, Y\rangle = \sum X_{ij}Y_{ij}$. The following adjointness properties of this inner product are used repeatedly in the main part of this report.

Proposition. *Let $X, Y \in \mathbb{R}^{m\times n}$.*

(1) *If $P \in \mathbb{R}^{m\times m}$ then $\langle PX, Y\rangle = \langle X, P^{\mathrm{T}}Y\rangle$,*
(2) *If $P \in \mathbb{R}^{n\times n}$ then $\langle XP, Y\rangle = \langle X, YP^{\mathrm{T}}\rangle$.*

Proof. Here is a proof of (1):

$$\langle PX, Y\rangle = \mathrm{Trace}(PXY^{\mathrm{T}}) = \mathrm{Trace}(XY^{\mathrm{T}}P) = \mathrm{Trace}(X(P^{\mathrm{T}}Y)^{\mathrm{T}}) = \langle X, P^{\mathrm{T}}Y\rangle.$$

Here is an even easier proof of (2):

$$\langle XP, Y\rangle = \mathrm{Trace}(XPY^{\mathrm{T}}) = \mathrm{Trace}(X(YP^{\mathrm{T}})^{\mathrm{T}}) = \langle X, YP^{\mathrm{T}}\rangle. \qquad \square$$

Corollary. *Let $X, Y, P \in \mathbb{R}^{n\times n}$. Then*

(1) $\langle [P, X], Y\rangle = \langle X, [P^{\mathrm{T}}, Y]\rangle$,
(2) $\langle [X, P], Y\rangle = \langle X, [Y, P^{\mathrm{T}}]\rangle$.

Proof. Here is a proof of (1):

$$\langle XP - PX, Y\rangle = \langle X, YP^{\mathrm{T}} - P^{\mathrm{T}}Y\rangle.$$

The proof of (2) is similar. $\square$

References

[1] R. Bellman, Introduction to Matrix Analysis, McGraw-Hill Book Company, 1960.
[2] G. Birkhoff, S. MacLane, A Survey of Modern Algebra, Macmillan, 1953.
[3] M.T. Chu, K.R. Driessel, The projected gradient method for least squares matrix approximations with spectral constraints, SIAM J. Numer. Anal. 27 (1990) 1050–1060.
[4] M.L. Curtis, Matrix Groups, second ed., Springer-Verlag, 1984.
[5] K.R. Driessel, On isospectral gradient flows—solving matrix eigenproblems using differential equations, in: J.R. Cannon, U. Horung (Eds.), Inverse Problems, Birkäuser, 1986, pp. 69–91, ISNM77.
[6] K.R. Driessel, On finding the eigenvalues and eigenvectors of a matrix by means of an isospectral flow, Technical Report #541, Department of Mathematical Sciences, Clemson University, 1987.
[7] K.R. Driessel, On finding the singular values and singular vectors of a matrix by means of an iso singular gradient flow, Technical Report #87-01, Department of Mathematics, Idaho State University, Pocatello, Idaho, 1987.

[8] K.R. Driessel, Some homogeneous spaces of matrices, Technical Report 91-1, Department of Mathematics, Idaho State University, Pocatello, Idaho, 1991.
[9] P.R. Halmos, Finite-Dimensional Vector Spaces, second ed., D. Van Nostrand Company, Inc., 1958.
[10] U. Helmke, J.B. Moore, Optimization and Dynamical Systems, Springer-Verlag, 1994.
[11] M.W. Hirsch, S. Smale, Differential Equations, Dynamical Systems, and Linear Algebra, Academic Press, 1974.
[12] R.A. Horn, C.A. Johnson, Matrix Analysis, Cambridge University Press, 1985.
[13] P. Lancaster, Theory of Matrices, Academic Press, 1969.
[14] S.L. Leon, Linear Algebra with Applications, Macmillan, 1986.
[15] P. Olver, Classical Invariant Theory, Cambridge University Press, 1999.
[16] G. Strang, Linear Algebra and Its Applications, second ed., Academic Press, 1980.
[17] J.A. Thorpe, Elementary Topics in Differential Geometry, Springer-Verlag, 1979.
[18] F.W. Warner, Foundations of Differentiable Manifolds and Lie Groups, Scott, Foresman and Company, 1971.