Submitted to the 46th IEEE CDC on 9 Mar 2007

Design of continuous-time flows on intertwined orbit spaces

P.-A. Absil   C. Lageman   J. H. Manton

***Abstract*— Consider a space M endowed with two or more Lie group actions. Under a certain condition on the orbits of the Lie group actions, we show how to construct a flow on M that projects to prescribed flows on the orbit spaces of the group actions. Hence, in order to design a flow that converges to the intersection of given orbits, it suffices to design flows on the various orbit spaces that display convergence to the desired orbits, and then to lift these flows to M using the proposed procedure. We illustrate the technique by creating a flow for principal component analysis. The flow projects to a flow on the Grassmann manifold that achieves principal subspace analysis and to a flow on the "shape" manifold that converges to the set of orthonormal matrices.**

I. INTRODUCTION

Given a symmetric positive-definite $n \times n$ matrix A, we say that a flow ϕ on the set $\mathbb{R}^{n\times p}$ of all the $n \times p$ real matrices achieves *principal subspace analysis (PSA)* if, for almost all initial conditions $X \in \mathbb{R}^{n\times p}$, the column space of the solution $\phi(t, X)$ converges to the p-dimensional invariant subspace (or *eigenspace*) of A associated with the largest eigenvalues. If moreover the columns of the solution $\phi(t, X)$ converge to the p principal eigenvectors of A, i.e., those corresponding to the largest eigenvalues, then the flow is said to achieve *principal component analysis (PCA)*.

There is a vast literature on continuous-time flows that achieve computational tasks, spanning several areas of computational science. This includes, but is not limited to, linear programming [7], [8], [9], [17], continuous nonlinear optimization [16], [24], discrete optimization [21], [22], [38], [5], signal processing [6], [14], [11], model reduction [20], [39] and automatic control [20], [28], [19]. Applications in linear algebra, and especially in eigenvalue and singular value problems, are particularly abundant. Important advances in the area have come from the work on isospectral flows in the early 1980s [37], [13], [31]. Interest for studying continuous-time flows stems in part from the works of Ljung [25] and Kushner and Clark [23] relating the behavior of learning algorithms to the one of associated differential equations; see, e.g., Oja and Karhunen [34] for an application. As discussed in [2], another reason for considering continuous-time systems is that it is easier to enforce certain qualitative features on continuous-time systems than on discrete-time systems. In order to obtain a suitable discrete-time algorithm, it is thus often advantageous to first produce a continuous-time algorithm, then attempt to discretize it correctly. Continuous-time systems are also of particular interest in tracking problems, when problem parameters change continuously over time. The task of the algorithm is here to follow a solution $y(t)$ of the problem $P(t)$ as time evolves.

Several continuous-time dynamical systems on matrix spaces (also called *matrix flows*) have been proposed in the literature that achieve PSA or even PCA; see [20], [10], [12], [29], [35], [27], [30] and the many references therein. Early analyses of PSA and PCA flows have focused on local stability issues without addressing the problem of global convergence in a mathematically satisfactory way. A breakthrough came with the analysis by Yoshizawa *et al.* [40] of the flow

$$\dot{Y} = AYN - YNY^TAY, \; Y \in \mathbb{R}^{n\times p}. \tag{1}$$

This flow was studied by Brockett [9] in the case where Y belongs to the set of orthonormal $n \times n$ matrices; for the choice $N = I$ it yields the well-known Oja flow [33]. Assuming that A is positive definite, Yoshizawa *et al.* [40] show that (1) is a gradient flow for a certain cost function on $\mathbb{R}^{n\times p}$ endowed with a well-chosen Riemannian metric. Using Łojasiewicz's theorem [26], they show that all solutions of (1) converge to a single equilibrium point. A difficulty with gradient-based approaches, however, is the absence of a systematic procedure to detect whether a flow can be expressed as a gradient flow and to determine the corresponding cost function and metric.

In [1], a constructive procedure was proposed that yields a matrix flow with PSA and PCA properties. The key observation was that a flow ϕ on $\mathbb{R}^{n\times p}$ achieves PCA if and only if, for almost all initial points Y_0, the solution $Y(t) := \phi(t, Y_0)$ satisfies the following three conditions (we assume that the eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ of A satisfy $\lambda_p > \lambda_{p+1}$): (i) the column space of $Y(t)$ converges to the dominant invariant subspace of A; (ii) $Y(t)$ converges to the set of orthonormal matrices or, equivalently stated, $\lim_{t\to\infty} Y^T(t)Y(t) = I_p$; (iii) $Y^T(t)AY(t)$ converges to the set of diagonal matrices with nonincreasing diagonal entries. With a view to constructing a flow that satisfies these three conditions, it was shown in [1] that any dynamical system $\dot{Y} = X(Y)$ on the set $\mathbb{R}^{n\times p}_*$ ($p < n$) of all the $n \times p$ real matrices with full column rank, can be decomposed as

$$\dot{Y} = X_1(Y) + X_2(Y) + X_3(Y),$$

where the terms $X_2(Y)$ and $X_3(Y)$ are tangent to the submanifold $Y\mathrm{GL}(p) := \{YM : M \in \mathrm{GL}(p)\}$ and the terms $X_1(Y)$ and $X_3(Y)$ are tangent to the submanifold

P.-A. Absil is with the Department of Mathematical Engineering, Université catholique de Louvain, Belgium www.inma.ucl.ac.be/∼absil

C. Lageman is with the Centre for Mathematics and its Applications, Australian National University, Australia

J. H. Manton is with the Department of Information Engineering, Australian National University, Australia

$O(n)Y := \{UY : Y \in O(n)\}$. Note that $Y\mathrm{GL}(p)$ is the set of all the matrices that have the same column space as Y and $O(n)Y$ is the set of all the matrices that have the same "shape" as Y. By carefully selecting $X_1(Y)$ and $X_2(Y)$, it was possible to obtain a flow that, for almost all initial points, converges to the set of matrices whose column space is the dominant subspace of A and whose shape is orthonormal. This achives conditions (i) and (ii) above. Finally, the term $X_3(Y)$ in $T_Y(Y\mathrm{GL}(p) \cap O(n)Y)$ was chosen to take care of condition (iii).

In this paper, we extend this constructive procedure to the more general setting of several (say, k) proper, principal Lie group actions on a manifold M. The condition that the Lie group actions be proper and principal ensures that the quotients of the manifold by the group actions are smooth manifolds. This makes it possible to first address the task of creating suitable vector fields $X_1, \ldots, X_k$ on the k quotient manifolds independently. The key issue is then to be able to build a vector field X on the manifold M that "projects" to the vector fields $X_1, \ldots, X_k$. We give a sufficient condition on the group actions for such a vector field X to exist regardless of $X_1, \ldots, X_k$. Under this sufficient condition, we provide a procedure to construct X from any given vector fields $X_1, \ldots, X_k$ on the quotient manifolds.

Note that the obtained vector field X is in general not invariant by the Lie group actions. An invariance condition would be too restrictive. Instead, we use the more general concept of partial symmetries from Nijmeijer and van der Schaft [32]. Stating that a group action is a partial symmetry for a vector field amounts to saying that the vector field induces a unique vector field on the quotient by the group action.

The paper is organized as follows. After introducing notation in Section II, we develop in Section III a construction procedure for vector fields with several partial symmetries. This procedure is illustrated on the PCA problem in Section IV. Conclusions are drawn in Section V.

II. Notation

In this section, we introduce some terminology regarding manifolds and group actions on manifolds. We refer, e.g., to [15] for a detailed account.

Let M, N be smooth manifolds and $f : M \to N$ be a smooth function. We let TM and TN denote the tangent bundles and T_xM, T_yN denote the tangent spaces at specific points $x \in M$, $y \in N$. The notation $Tf : TM \to TN$ is used for the tangent map of f, and $T_xf : T_xM \to T_{f(x)}N$ for the tangent map at a point $x \in M$.

Let G be a Lie group. An *action* of G on M is a smooth mapping $\psi : G \times M \to M$ with $\psi(g, \psi(h, x)) = \psi(gh, x)$ and $\psi(e, x) = x$, e the identity of G. For shorter notation we usually let gx stand for $\psi(g, x)$ and ψ_g for the map $x \mapsto \psi(g, x)$. We call the action *proper* if the map $(g, x) \mapsto (x, \psi(g, x))$ is proper, i.e., preimages of compact sets are compact. The action is called *free* if $\psi(g, x) = x$ for a $x \in M$ implies $g = e$. The *orbit* Gx of $x \in M$ is the set $\{\psi(g, x) \mid g \in G\}$. The *isotropy group* (or *stabilizer* or *symmetry group*) G_x of an $x \in M$ is the subgroup $\{g \mid \psi(g, x) = x\}$. Following [18], we say that the action is *principal* if the isotropy groups of all $x \in M$ are conjugate, i.e., for all $x, y \in M$ there is an $h \in G$ with $G_x = h^{-1}G_yh$. Free actions are principal as the isotropy groups equal the trivial group $\{e\}$. The *orbit space* (or *quotient space*) with respect to the group operation is the quotient space of M with respect to the equivalence relation $x \sim y \Leftrightarrow \exists g \in G\colon \psi(x, g) = y$. A map f between manifolds M, N with group actions of a Lie group G is called *equivariant* if it commutes with the group actions, i.e., $gf(x) = f(gx)$ for all $x \in M$, $g \in G$. In particular, a vector field X is called equivariant if $X(gx) = T\psi_g X(x)$ for all $x \in M$.

For a general Lie group action on a manifold M, the orbit space need not be a manifold. In fact, it can be very complicated. For example, any complete, smooth dynamical system can be viewed as a smooth action of $\mathbb{R}$ on the underlying space. If we restrict ourselves to proper, principal group actions, then the orbit space is a smooth manifold [15]. In this case, the manifold M has the structure of a fiber bundle over the orbit space. The fiber through a point $x \in M$ is isomorphic to the homogeneous space G/G_x. For free group actions, one even has a principal fiber bundle, i.e., a neighborhood of a fiber is equivariant diffeomorphic to the product $G \times U$, for an open set $U \subset M/G$. For principal actions, fibers have a neighborhood equivariantly diffeomorphic to $G/G_x \times U$ [15]. In the remainder of this paper, we will only consider proper, principal group actions.

III. Construction of vector fields with partial symmetries

Assume that a manifold M with proper and principal group actions of Lie groups $G_1, \ldots, G_k$ is given. We will present a construction method which derives a vector field on M from vector fields $X_1, \ldots, X_k$ on the orbit spaces M/G_i. The methods aims to lift the convergence properties from the vector fields on the orbit spaces to the new one on the main manifold. For this purpose the constructed vector field will have the partial symmetries $G_1, \ldots, G_k$ and its projections on the orbit spaces will coincide with $X_1, \ldots, X_k$.

Before introducing our construction method, we recall the notions of horizontal and vertical distributions of group actions as these are an essential tool for our approach. The natural projection π of M onto the orbit space M/G is a submersion and defines the *vertical distribution* V by $V(x) = \ker T_x\pi$. Note that the vertical distribution consists of the tangent spaces to the orbits of the group action. For lifting curves and vector fields from the orbit space to the manifold M, we need another distribution, a *horizontal distribution* H, satisfying the following conditions

- $H(x) \oplus V(x) = T_xM$ and
- $T_x\psi_g H(x) = H(gx)$ for all $g \in G$, $x \in M$ (equivariance).

Unlike the vertical distribution, the horizontal one is usually not unique. Note that by the slice theorem [15] there is always an equivariant Riemannian metric h on M, i.e., $h(T\psi_g(x)v, T\psi_g(x)w) = h(v, w)$ for all $g \in G$, $x \in M$,

$v, w \in T_xM$. Hence, choosing the orthogonal complement of V with respect to h gives us always a horizontal distribution on M. Given a vector field X on M and a horizontal distribution H we can decompose X uniquely into a *horizontal part* $X_H \in H$ and *vertical part* $X_V \in V$ with $X = X_H + X_V$. Furthermore, for a vector field $\hat{X}$ on M/G, there exists a unique *horizontal lift*, depending on the horizontal distribution H, to a vector field $\hat{X}_H$ on M with $\hat{X}_H(x) \in H(x)$ and $T\pi\hat{X}_H = \hat{X}$. Unless stated otherwise, the vector fields we consider are always smooth.

We now recall the definition of partial symmetries for a vector field X on M [32].

Definition 3.1: Let M be a manifold and G a principal, proper group action. The action (or G) is a *partial symmetry* of a vector field X on M if there exists a horizontal distribution H such that

$$X_H(gx) = T\psi_g(x)X_H(x)$$

where X_H denotes the horizontal part of X.

Proposition 3.2: The action of G is a partial symmetry of X if and only if there exists a vector field $\tilde{X}$ on M/G such that the diagram

$$\begin{array}{ccc} M & \xrightarrow{X} & TM \\ \pi\downarrow & & T\pi\downarrow \\ M/G & \xrightarrow{\tilde{X}} & T(M/G) \end{array}$$

commutes, $\pi : M \to M/G$ the canonical projection.

Proof: See [32]. ∎

Remark 3.3: As a consequence of Proposition 3.2, we can use the horizontal space to lift a vector field $\tilde{X}$ on the orbit space to a vector field X on M with partial symmetry G. In fact, as T_xM is the direct sum of $V(x) = \ker T_x\pi$ and $H(x)$, there is a unique preimage $X(x)$ of $\tilde{X}(\pi(x))$ in $H(x)$. By the Proposition this vector field has the partial symmetry G.

Recall that a vector field on M is *complete* if its integral curves are defined for all $t \in \mathbb{R}$. Likewise a flow ϕ on M is called *complete* if it is defined for all $t \in \mathbb{R}$.

Proposition 3.4: Let X be a complete vector field on M with partial symmetry G. Then the flow ϕ of X induces a complete flow $\tilde{\phi}$ on M/G such that

$$\pi \circ \phi(x, t) = \tilde{\phi}(\pi(x), t)$$

for all $x \in M$, $t \in \mathbb{R}$, $\pi : M \to M/G$ the canonical projection. The flow $\tilde{\phi}$ is the flow of the induced vector field on M/G from Proposition 3.2.

Proof: Immediate. ∎

By Remark 3.3 vector fields on the orbit spaces can be lifted to vector fields on the main manifold by using horizontal distributions for the group actions. However, as we want to construct a single vector field from vector fields on several different orbit spaces, we will need some additional compatibility conditions on the vertical and horizontal distributions.

Let M be a manifold with principal and proper group actions of the groups $G_1, \ldots, G_k$. A *local coordinate system* τ for integrable distributions $V_1, \ldots, V_m$ on M is a diffeomorphism $\tau : U \to N_1 \times \ldots \times N_k$, $U \subset M$ open, N_i manifolds such that

$$A_i = \cap_{j \in I(i)} \tau_j^{-1}(c_j)$$

is an integral manifold of V_i for all $c_j \in N_j$, $j \in I(i)$ for suitable index sets $I(i)$. Following [36] we call the vertical distributions V_i *simultaneously integrable*, if for any $x \in M$ we have a neighborhood $U(x)$ with local coordinates $x_1, \ldots, x_n$ such that the integral manifolds of a V_i have the form

$$x_l \quad = \quad \text{const} \quad l \in I(i)$$

for suitable index sets $I(i) \subset \{1, \ldots, n\}$ with $\dim V_i$ elements. Note that simultaneous integrability implies the simultaneous integrability of sums of arbitrary intersections of the distributions.

For our construction method we will not require the simultaneous integrability of the vertical distributions directly. Instead we use a stronger transversality condition on the horizontal distributions.

Condition 3.5: Let $G_1, \ldots, G_k$ be groups acting on the manifold M with horizontal distributions $H_1, \ldots, H_k$. We will use the following condition:

- (H) The horizontal distributions H_i are contained in the vertical distributions of the other actions, i.e.

$$H_i \subset V_j \quad \text{for } i \neq j.$$

With this condition we propose the following construction method.

Construction 3.6: Let $G_1, \ldots, G_k$ be groups with principal and proper group actions on the manifold M and with horizontal distributions $H_1, \ldots, H_k$ satisfying (H). Then for arbitrary vector fields $X_1, \ldots, X_k$ on the orbit spaces M/G_i and a vector field $Y \in \cap_{i=1}^k V_i$ we construct a vector field X on M by

$$X = Y + \sum_{i=1}^{k} X_{H_i},$$

where X_{H_i} denotes the horizontal lift of X_i.

The following proposition ensures that this construction makes sense. Furthermore, the condition (H) is both necessary and sufficient for our construction.

Proposition 3.7: Let $G_1, \ldots, G_k$ be groups with principal and proper group actions on the manifold M and with horizontal distributions $H_1, \ldots, H_k$. Then the following are equivalent

1) (H) holds.
2) For arbitrary vector fields $X_1, \ldots, X_k$ on the quotient spaces M/G_i the vector field

$$X = \sum_{i=1}^{k} X_{H_i}$$

satisfies $T\pi_i X = X_i$ for $i = 1, \ldots, k$, $\pi_i : M \to M/G_i$ the canonical projections, X_{H_i} the horizontal lift of X_i.

Under this condition the group actions G_i are partial symmetries of X. Furthermore, if Y is a vector field with

$$Y \in \bigcap_{i=1}^{k} V_i,$$

then $X+Y$ has partial symmetries $G_1, \ldots, G_k$ and $T\pi_i(X+Y) = X_i$.

Proof: That (H) implies the condition on X follows directly from the definition of the horizontal lift and the vertical distributions. Assume that (H) does not hold. We can restrict ourselves to the case of two group actions G_1, G_2 on M. W.l.o.g. $H_1 \not\subset V_2$. Then there exists an $x \in M$ and $v \in T_xM$, with $v \in H_1(x) \setminus V_2(x)$, in particular $v \neq 0$. Furthermore, there is a vector field X_1 on M/G_1 with $X_1(\pi_1(x)) = T_x\pi_1 v$. Its lift X_{H_1} has the value v at x. On M/G_2 we choose the vector field $X_2 \equiv 0$ with lift $X_{H_2} \equiv 0$. The projection of $X = X_{H_1} + X_{H_2}$ to M/G_2 at the point $\pi_2(x)$ is $\neq 0$ as $\ker T_x\pi_2 = V_2(x)$. Thus $T_x\pi_2 X(x) \neq X_2(x)$ and (H) is equivalent to condition 2. The remaining claims follow directly from the definitions. ∎

Remark 3.8: If we would construct a vector field with symmetries, i.e., an equivariant one, instead of the one with partial symmetries, then the construction would be much more complicated. We would have to use compatibility conditions on the X_i restricting our possible choices to a large extent. Take for example the additive group $\mathbb{R}$ operating on $\mathbb{R}^2$ by translations ϕ_1, ϕ_2 of the first and the second coordinate. With partial symmetries we can choose arbitrary vector fields on the orbit spaces, here in both cases $\mathbb{R}$, to construct a vector field on $\mathbb{R}^2$. If we would require that the constructed vector field is equivariant with respect to ψ_1 and ψ_2 then the only possible choice for vector fields on $\mathbb{R}$ would be the constant ones.

By construction, the sum X induces the given vector fields X_i on the orbit spaces. We now consider the structure of X in local coordinates. For this we need some technical lemmas.

Lemma 3.9: Let $G_1, \ldots, G_k$ be groups acting principal and proper group actions on the manifold M, satisfying (H). We denote by Q the distribution $\cap_{j=1}^{k} V_j$. Then

$$TM = Q \oplus \bigoplus_{j=1}^{k} H_j,$$

where $\oplus$ denotes the direct sum of distributions.

Proof: Let $L_i = \cap_{j=1}^{i} V_j$ for $i \in \{1, \ldots, k\}$. We show inductively that

$$TM = L_i \oplus \bigoplus_{j=1}^{i} H_j.$$

As $L_1 = V_1$ we get from the definition of the vertical and horizontal distributions that $L_1 \oplus H_1 = TM$. Assume that the claim is shown for $i \in \{1, \ldots, k-1\}$. By (H) we have that $\oplus_{j=1}^{i} H_j \subset V_{i+1}$. Thus $L_i + V_{i+1} = TM$. From

$$\begin{aligned} V_{i+1} &= V_{i+1} \cap (L_i \oplus \bigoplus_{j=1}^{i} H_j) \\ &= (V_{i+1} \cap L_i) \oplus \bigoplus_{j=1}^{i} H_j = L_{i+1} \oplus \bigoplus_{j=1}^{i} H_j \end{aligned}$$

and $H_{i+1} \oplus V_{i+1} = TM$ we get that

$$L_{i+1} \oplus \bigoplus_{j=1}^{i+1} H_j = TM.$$

This proves our claim. ∎

Lemma 3.10: Let $G_1, \ldots, G_k$ be groups with principal and proper group actions on the manifold M, satisfying (H). We denote by Q the distribution $\cap_{j=1}^{k} V_j$. Then Q is integrable. Furthermore given local coordinates ρ for Q, the function $\mu\colon x \mapsto (\pi_1(x), \ldots, \pi_k(x), \rho(x))$ is a local coordinate system for the V_i. In particular, the distributions V_i are simultaneously integrable.

Proof: We denote by $\hat{\mu}$ the map $x \mapsto (\pi_1(x), \ldots, \pi_k(x))$. As $\ker T_x\pi_i(x) = V_i(x)$ for $i = 1, \ldots, k$ we have that $\ker T_x\hat{\mu} = \cap_{i=1}^{k} V_i(x) = Q(x)$. By Lemma 3.9 the equation

$$\begin{aligned} \dim M &= \dim Q + \sum_{i=1}^{k} \dim H_i \\ &= \dim Q + \sum_{i=1}^{k} \dim(M/G_i) \qquad (2) \end{aligned}$$

holds. This implies that $\operatorname{rank} T_x\hat{\mu} = \sum_{i=1}^{k} \dim(M/G_i)$. Thus $\hat{\mu}$ is a submersion and $Q(x)$ is integrable. We consider now the function μ given in the Lemma. As $\mu = (\hat{\mu}, \rho)$ and $\ker T_x\hat{\mu} = Q(x)$ we have that $\ker T_x\mu = \{0\}$. From equation (2) if follows that T_xM is an isomorphism. Thus μ must be a local diffeomorphism. Furthermore, the sets $\pi_i^{-1}(x_i)$, $x_i \in M/G_i$ are the integral manifolds of the distributions V_i. Hence, μ is a local coordinate system for the V_i. ∎

Proposition 3.11: Let X be the vector field from Construction 3.6. Then X is decoupled with respect to the group actions. That is, in suitable local coordinates, we have

$$X = \begin{pmatrix} f_1(x_1) \\ \vdots \\ f_k(x_k) \\ f_{k+1}(x_1, \ldots, x_{k+1}) \end{pmatrix},$$

where x_{k+1} are the coordinates of $\cap_{i=1}^{k} V_i$ and x_i, $i = 1, \ldots, k$ are the remaining coordinates given by π_i.

Proof: This is a direct consequence of Proposition 3.7 and Lemma 3.10. ∎

IV. Construction of a new PCA flow

In [1], a PCA flow was obtained using an ad hod constructive approach. In this section, we revisit the results of [1] in the context of the general theory developed in Section III.

We consider two actions on $\mathbb{R}_*^{n\times p}$:

- the action $\psi_1(M,Y) = YM$ of $\mathrm{GL}(p)$ by right multiplication
- and the action $\psi_2(U,Y) = UY$ of $O(n)$ by left multiplication.

The actions are both proper and ψ_1 is free. The isotropy groups of ψ_2 have the form

$$G_Y = Y(Y^TY)^{-1}Y^T + Y_\perp O(n-p)Y_\perp^T$$

with $Y_\perp$ an orthonormal $n\times(n-p)$ matrix, $Y^TY_\perp = 0$. Given an SVD $\Theta\Sigma\Omega = Y$ of Y, we see that G_Y can be written as

$$G_Y = \Theta\begin{pmatrix} I_p & 0 \\ 0 & O(n-p)\end{pmatrix}\Theta^T$$

where I_p is the $p\times p$ identity matrix. Hence, ψ_2 is principal.

The quotient space of ψ_1 is the *Grassmann manifold* $\mathrm{Grass}(p,n)$ of p-dimensional subspaces of $\mathbb{R}^n$ as an orbit consists of all $n\times p$ matrices with the same span (see [3] for details). For ψ_2 we call the quotient space the *shape manifold* as the orbits of ψ_2 are given by matrices of the same shape, i.e., related by rotations.

For the construction of a flow on $\mathbb{R}_*^{n\times p}$ we have to choose the horizontal distributions such that the condition (H) holds.

The vertical distributions of the actions ψ_1 and ψ_2 are

$$V_1(Y) = \{YM \mid M\in\mathfrak{gl}(p)\},$$

where $\mathfrak{gl}(p) = \mathbb{R}^{p\times p}$ is the Lie algebra of $p\times p$ matrices, and

$$\begin{aligned} V_2(Y) &= \{\Omega Y \mid \Omega^T = -\Omega\} \\ &= \{Y_\perp K + Y(Y^TY)^{-1}\Omega \mid K\in\mathbb{R}^{(n-p)\times p}, \\ &\qquad\qquad \Omega^T = -\Omega\in\mathbb{R}^{p\times p}\}. \end{aligned}$$

We choose as horizontal distributions

$$H_1(Y) = \{W \mid Y^TW = 0\} = \{Y_\perp K \mid K\in\mathbb{R}^{(n-p)\times p}\}$$

and

$$H_2(Y) = \{Y(Y^TY)^{-1}S \mid S^T = S\}.$$

It is easily checked that these distributions are indeed horizontal and satisfy (H).

We choose now the flows on the Grassmann and shape manifold with a view to lifting them to $\mathbb{R}_*^{n\times p}$. As our new flow should have PCA properties we choose a PSA flow on the Grassmann manifold. Let f be the real-valued function on $\mathrm{Grass}(p,n)$ induced by the generalized Rayleigh quotient $Y\mapsto \mathrm{trace}((Y^TY)^{-1}Y^TAY)$. The function f is a Morse-Bott function and the gradient ascent flow with respect to the canonical Riemannian metric has PSA properties, see [29]. (Note that the authors consider only the compact Stiefel manifold, but the Grassmann manifold is just a quotient of the compact Stiefel and the function on the Stiefel manifold induces the function above on the Grassmann manifold. Hence, the results hold there, too.) The equilibria are the p-dimensional eigenspaces of A and the eigenspaces of the largest eigenvalues constitute an asymptotically stable manifold. Furthermore, all other manifolds of equilibria are unstable. Lifting this vector field to $\mathbb{R}_*^{n\times p}$ gives us (see [4])

$$X_1(Y) = (I - Y(Y^TY)^{-1}Y^T)AY. \tag{3}$$

On the shape manifold we choose the vector field whose lift on $\mathbb{R}_*^{n\times p}$ is

$$X_2(Y) = Y(Y^TY)^{-1}(I_p - Y^TY). \tag{4}$$

This vector field is hyperbolic with one asymptotically stable equilibrium, the equivalence class of orthogonal matrices. To see this, consider the function v on the shape manifold induced by $\mathrm{trace}((I_p - Y^TY)(I_p - Y^TY))$. One easily checks that $\dot v = -4v$ and $v(t)$ is strictly decreasing. Furthermore $v(\pi_2(Y))\to\infty$ for $\mathrm{trace}(Y^TY)\to\infty$, π_2 the canonical projection on the orbit space of ψ_2. Let $Y = \Omega\,\mathrm{diag}(\sigma_1,\ldots,\sigma_p)\Theta$ be an SVD of $Y\in\mathbb{R}_*^{n\times p}$ with $\sigma_1\geq\ldots\geq\sigma_p > 0$ and $\mathrm{diag}(\sigma_1,\ldots,\sigma_p)$ the $n\times p$ matrix with $\sigma_1,\ldots,\sigma_p$ on the diagonal. A simple calculation shows that

$$X_2(Y) = \Omega\,\mathrm{diag}(\sigma_1^{-1}-1,\ldots,\sigma_p^{-1}-1)\Theta. \tag{5}$$

Denote by $\mathbb{R}_*^p$ the set $\{x\in\mathbb{R}^p \mid x = (x_1,\ldots,x_p), x_i\neq 0 \text{ for } i = 1,\ldots,p\}$. Using the over-parametrization $M = SO(n)\times\mathbb{R}_*^p\times SO(p)$ we can lift X_2 locally around Y to M by setting $\hat X_2(\Omega,\sigma_1,\ldots,\sigma_p,\Theta) = (0,\sigma_1^{-1}-1,\ldots,\sigma_p^{-1}-1,0)$. If $\sigma_i < 1$ then the σ_i values are locally increasing along the integral curve of $\hat X_2$ through $(\Omega,\sigma_1,\ldots,\sigma_p,\Theta)$. As $\hat X_2$ projects to X_2, the integral curves of $\hat X_2$ are mapped on integral curves of X_2. Thus, if an singular value σ_i is smaller than 1 then it is locally increasing along the integral curve of X_2. Therefore, the singular values along an integral curve can never converge to 0. Hence, the integral curves are never moving into the set of singular $n\times p$ matrices. Note that the singular values on an orbit of ψ_2 are constant. Thus, by the properties of v described above, all integral curves on the orbit space of the ψ_2 action are defined on the interval $(0,\infty)$. Therefore, as $\dot v = -4v$, v converges exponentially to 0 on all integral curves. Hence, the orthogonal shape as the single equilibrium is asymptotically stable with the whole manifold $\mathbb{R}_*^{n\times p}$ as region of attraction.

As the intersection of the vertical distributions $V_1\cap V_2$ is not trivial, we have third choice for the construction of our new vector field. We use it to enforce PCA properties of the flow, as the flow on the Grassmannian gives only PSA properties. The distribution $V_1\cap V_2$ is the vertical distribution of a third action on $\mathbb{R}_*^{n\times p}$, the action $\psi_3(U,Y) = Y(Y^TY)^{-1/2}U(Y^TY)^{1/2}$ of $O(p)$. Hence, we have to smoothly choose vector fields on the orbits of this action to give us convergence of the integral curves to matrices of eigenvectors of A. To achieve this we use the vector field

$$X_3(Y) = Y(Y^TY)^{-1}[Y^TAY, N], \tag{6}$$

where $[A, B] = AB - BA$ and $N = \mathrm{diag}(p, \ldots, 1)$. The choice is justified by the following lemma.

Lemma 4.1: Let $A \in R^{n\times n}$ be symmetric. Let $V \in \mathbb{R}_*^{n\times p}$ be a orthonormal matrix which spans a p-dimensional eigenspace of A. Then the set $VO(p)$ is invariant under X_3 and, writing $Y = VQ$, we obtain the vector field

$$\tilde{X}_3(Q) = Q[Q^T V^T AVQ, N] \tag{7}$$

on $O(p)$. This is the gradient vector field of the Morse-Bott function $-\mathrm{trace}(Q^T V^T AVQN)$ and for the equilibria U the matrix VU consists only of eigenvectors of A. The eigenvectors in the stable equilibria are sorted by decreasing eigenvalues.

Proof: This is a direct consequence of the behavior of the double bracket flow, see e.g. [20, 2.1]. Note that for the stability [20, 2.1] consider only A with distinct eigenvalues. In the general case an approximation argument gives that the equilibria with sorted eigenvectors as above are stable. The lack of stability for equilibria with other orders can be seen be applying a rotation to two subsequent eigenvectors with increasing eigenvalues. ■

Using Construction 3.6 we get

$$X(Y) = X_1(Y) + X_2(Y) + X_3(Y) \tag{PCA}$$

as a vector field with partial symmetries ψ_1, ψ_2. In summary, we have obtained the matrix differential equation

$$\dot{Y} = (I - Y(Y^TY)^{-1}Y^T)AY + Y(Y^TY)^{-1}(I_p - Y^TY) + Y(Y^TY)^{-1}[Y^TAY, N]. \tag{8}$$

The following convergence results were obtained in [1].

Theorem 4.2: Let A be an $n \times n$ symmetric matrix (not necessarily positive definite) with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$ and associated orthonormal eigenvectors $v_1, \ldots, v_n$. Let $V = [v_1|\ldots|v_p]$. Let $N = \mathrm{diag}(p, \ldots, 1)$. Consider the dynamics (8) with full-rank $n \times p$ initial condition $Y(0) = Y_0 \in \mathbb{R}_*^{n\times p}$. Then
(i) $Y(t)^T Y(t) \to I_p$ as $t \to +\infty$.
(ii) The flow (8) induces a subspace flow, i.e., $\mathrm{span}(Y(t))$ only depends on $\mathrm{span}(Y(0))$.
(iii) There exists an eigenspace $\mathcal{S}$ of A such that $\mathrm{span}(Y(t)) \to \mathcal{S}$ as $t \to +\infty$.
(iv) The eigenspace $\mathcal{S}$ is asymptotically stable for the induced subspace flow if and only if it is the unique p-dimensional rightmost eigenspace of A (the *p-dimensional rightmost eigenspace* is the invariant subspace of A associated with the rightmost eigenvalues on the real line). $\mathcal{S}$ is unstable if it is not a rightmost eigenspace of A.
Now assume that $\lambda_1 > \ldots > \lambda_{p+1}$.
(v) The set $VO_p = \{VQ : Q^TQ = I_p\}$ is invariant with respect to (8). If the initial condition Y_0 is in VO_p, then $Y(t)$ converges to an orthonormal matrix $\tilde{V}$ whose columns are eigenvectors of A. The equilibrium point $\tilde{V}$ is stable conditionally to VO_p if and only if $\tilde{V} = V\mathrm{diag}(\pm 1, \ldots, \pm 1)$.
(vi) If $\mathrm{span}(Y(t))$ converges to the rightmost eigenspace (stable case), then $Y(t)$ converges to VP where P is a signed permutation matrix. Only the matrices $V\mathrm{diag}(\pm 1, \ldots, \pm 1)$ are stable equilibrium points of (8).

V. Conclusion and Future Work

The main purpose of this paper has been to present the partial-symmetry construction procedure (Construction 3.6) in the most general Lie group setting, and to illustrate it on a PCA flow. We will present elsewhere new tools, based on normal hyperbolicity, for analyzing the convergence of flows thus constructed. In particular, it possible to relax the assumption in Theorem 4.2 that the first p eigenvalues are simple.

References

[1] P.-A. Absil, "Continuous-time flows on quotient spaces for principal component analysis," in *Proceedings of the 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS2004)*, 2004.

[2] ——, "Continuous-time systems that solve computational problems," 2007, international Journal of Unconventional Computing, to appear.

[3] P.-A. Absil, R. Mahony, and R. Sepulchre, "Riemannian geometry of Grassmann manifolds with a view on algorithmic computation," *Acta Appl. Math.*, vol. 80, no. 2, pp. 199–220, January 2004.

[4] ——, *Optimization Algorithms on Matrix Manifolds.* Princeton, NJ: Princeton University Press, 2007, in preparation.

[5] P.-A. Absil and R. Sepulchre, "Continuous dynamical systems that realize discrete optimization on the hypercube," *Systems Control Lett.*, vol. 52, no. 3-4, pp. 297–304, 2004. [Online]. Available: http://authors.elsevier.com/sd/article/S0167691104000155

[6] S.-i. Amari and A. Cichocki, "Adaptive blind signal processing—neural network approaches," *Proc. IEEE*, vol. 86, no. 10, pp. 2026–2048, 1998.

[7] D. A. Bayer and J. C. Lagarias, "The nonlinear geometry of linear programming. I. Affine and projective scaling trajectories," *Trans. Amer. Math. Soc.*, vol. 314, no. 2, pp. 499–526, 1989.

[8] ——, "The nonlinear geometry of linear programming. II. Legendre transform coordinates and central trajectories," *Trans. Amer. Math. Soc.*, vol. 314, no. 2, pp. 527–581, 1989.

[9] R. W. Brockett, "Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems," *Linear Algebra Appl.*, vol. 146, pp. 79–91, 1991.

[10] M. T. Chu, "A list of matrix flows with applications," in *Hamiltonian and gradient flows, algorithms and control*, ser. Fields Inst. Commun. Providence, RI: Amer. Math. Soc., 1994, vol. 3, pp. 87–97.

[11] A. Cichocki and P. Georgiev, "Blind source separation algorithms with matrix constraints," *IEICE Trans. Fundamentals*, vol. E86-A, no. 1, pp. 1–9, 2003.

[12] J. Dehaene, "Continuous-time matrix algorithms, systolic algorithms and adaptive neural networks," Ph.D. dissertation, Katholieke Universiteit Leuven, Faculteit Toegepaste Wetenschappen, Departement elektrotechniek-ESAT, Kard. Mercierlaan 94, 3001 Leuven, Belgium, 1995, ftp://ftp.esat.kuleuven.ac.be/pub/SISTA/dehaene/phd/.

[13] P. Deift, T. Nanda, and C. Tomei, "Ordinary differential equations and the symmetric eigenvalue problem," *SIAM J. Numer. Anal.*, vol. 20, no. 1, pp. 1–22, 1983.

[14] S. C. Douglas, "Self-stabilized gradient algorithms for blind source separation with orthogonality constraints," *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1490–1497, 2000.

[15] J. J. Duistermaat and J. A. Kolk, *Lie groups*. Berlin: Springer-Verlag, 2000.

[16] L. Faybusovich, "Dynamical systems which solve optimization problems with linear constraints," *IMA J. Math. Control Inform.*, vol. 8, no. 2, pp. 135–149, 1991.

[17] ——, "Hamiltonian structure of dynamical systems which solve linear programming problems," *Physica D*, vol. 53, pp. 217–232, 1991.

[18] M. J. Field, "Equivariant dynamical systems," *Trans. Amer. Math. Soc.*, vol. 259, no. 1, pp. 185–205, 1980.

[19] F. Grognard and R. Sepulchre, "Global stability of a continuous-time flow which computes time-optimal switchings," in *Proceedings of the 16th IEEE Conference on Decision and Control*, 2001, pp. 3826–3831.

[20] U. Helmke and J. B. Moore, *Optimization and Dynamical Systems*. Springer, 1994.

[21] J. J. Hopfield, “Neurons with graded response have collective computational cababilities like those of two-state neurons,” *Proceedings of the National Academy of Sciences USA*, vol. 81, pp. 3088–3092, 1984.

[22] J. J. Hopfield and D. W. Tank, “‘Neural’ computation of decision optimization problems,” *Biological Cybernetics*, vol. 52, pp. 141–152, 1985.

[23] H. J. Kushner and D. S. Clark, *Stochastic approximation methods for constrained and unconstrained systems*, ser. Applied Mathematical Sciences. Springer-Verlag, 1978, vol. 26.

[24] X.-B. Liang and J. Wang, “A recurrent neural network for nonlinear optimization with a continuously differentiable objective function and bound constraints,” *IEEE Trans. Neural Networks*, vol. 11, no. 6, pp. 1251–1262, 2000.

[25] L. Ljung, “Analysis of recursive stochastic algorithms,” *IEEE Trans. Automat. Control*, vol. 22, no. 4, pp. 551–575, 1977.

[26] S. Łojasiewicz, “Sur les trajectoires du gradient d’une fonction analytique,” in *Seminari di Geometria 1982-1983*, Università di Bologna, Istituto di Geometria, Dipartimento di Matematica, 1984, pp. 115–117.

[27] R. Mahony and P.-A. Absil, “The continuous-time Rayleigh quotient flow on the sphere,” *Linear Algebra Appl.*, vol. 368C, pp. 343–357, 2003.

[28] R. E. Mahony and U. Helmke, “System assignment and pole placement for symmetric realisations,” *J. Math. Systems Estim. Control*, vol. 8, no. 3, pp. 321–352, 1998.

[29] R. E. Mahony, U. Helmke, and J. B. Moore, “Gradient algorithms for principal component analysis,” *J. Austral. Math. Soc. Ser. B*, vol. 37, pp. 430–450, 1996.

[30] J. H. Manton, U. Helmke, and I. M. Y. Mareels, “A dual purpose principal and minor component flow,” *Systems Control Lett.*, vol. 54, no. 8, pp. 759–769, 2005.

[31] T. Nanda, “Differential equations and the QR algorithm,” *SIAM J. Numer. Anal.*, vol. 22, no. 2, pp. 310–321, 1985.

[32] H. Nijmeijer and A. J. van der Schaft, “Partial symmetries for nonlinear systems,” *Math. Systems Theory*, vol. 18, no. 1, pp. 79–96, 1985.

[33] E. Oja, “Neural networks, principal components, and subspaces,” *Int. J. Neural Syst.*, vol. 1, pp. 61–68, 1989.

[34] E. Oja and J. Karhunen, “On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix,” *J. Math. Anal. Appl.*, vol. 106, no. 1, pp. 69–84, 1985.

[35] M. Przybylska, “Isospectral-like flows and eigenvalue problem,” *Future Generation Computer Systems*, vol. 19, pp. 1165–1175, 2003.

[36] W. Respondek, “On decomposition of nonlinear control systems,” *Systems Control Lett.*, vol. 1, no. 5, pp. 301–308, 1981/82.

[37] W. W. Symes, “The QR algorithm and scattering for the finite nonperiodic Toda lattice,” *Physica D*, vol. 4, no. 2, pp. 275–280, 1982.

[38] M. Vidyasagar, “Minimum-seeking properties of analog neural networks with multilinear objective functions,” *IEEE Trans. Automat. Control*, vol. 40, no. 8, pp. 1359–1375, 1995.

[39] W.-Y. Yan and J. Lam, “An approximate approach to H^2 optimal model reduction,” *IEEE Trans. Automat. Control*, vol. 44, no. 7, pp. 1341–1358, 1999.

[40] S. Yoshizawa, U. Helmke, and K. Starkov, “Convergence analysis for principal component flows,” *Int. J. Appl. Math. Comput. Sci.*, vol. 11, no. 1, pp. 223–236, 2001.