Continuous-time subspace flows related to the symmetric eigenproblem*

P.-A. Absil† R. Sepulchre‡ R. Mahony§

June 11, 2008

PREPRINT

Abstract

The classes of continuous-time flows on $\mathbb{R}^{n\times p}$ that induce the same flow on the set of p-dimensional subspaces of $\mathbb{R}^n$ are described. The power flow is briefly reviewed in this framework, and a subspace generalization of the Rayleigh quotient flow [Linear Algebra Appl. **368C**, 2003, pp. 343–357] is proposed and analyzed. This new flow displays a property akin to deflation in finite time.

Key words. Matrix flows, subspace flows, power flow, Grassmann Rayleigh quotient flow, principal component analysis, invariant subspace, finite-time deflation.

1 Introduction

Let $\mathbb{R}_*^{n\times p}$, with $p < n$, denote the set of all $n \times p$ matrices with full column rank, let $\mathrm{Gr}(p, n)$ denote the set of all p-dimensional linear subspaces of $\mathbb{R}^n$, and let $\mathsf{span}(Y) \in \mathrm{Gr}(p, n)$ denote the column space of $Y \in \mathbb{R}_*^{n\times p}$. It is well known (see, e.g., [Boo75, AMS08]) that the set $\mathrm{Gr}(p, n)$ admits one and only one "natural" differentiable structure, the one for which the span mapping is a submersion (i.e., span is differentiable and its differential map is a surjection at every point of $\mathbb{R}_*^{n\times p}$). Endowed with this differentiable structure, $\mathrm{Gr}(p, n)$ is called the *Grassmann manifold* of p-planes in $\mathbb{R}^n$.

Consider in $\mathbb{R}_*^{n\times p}$ the continuous-time dynamical system

$$\dot{Y} = F(Y), \tag{1}$$

*This paper presents research results of the Belgian Network DYSCO (Dynamical Systems, Control, and Optimization), funded by the Interuniversity Attraction Poles Programme, initiated by the Belgian State, Science Policy Office. The scientific responsibility rests with its authors. This work was supported in part by the United States National Science Foundation under Grant OCI0324944 and by the School of Computational Science of Florida State University through a postdoctoral fellowship.

†Département d'ingénierie mathématique, Université catholique de Louvain, Belgium (www.inma.ucl.ac.be/∼absil).

‡Department of Electrical Engineering and Computer Science, Université de Liège, Institut Montefiore (B28), Grande Traverse 10, B-4000 Liège, Belgium (www.montefiore.ulg.ac.be/systems).

§Australian National University, ACT, 0200, Australia (Robert.Mahony@anu.edu.au).

where $\dot{Y}(t) = \frac{\mathrm{d}}{\mathrm{d}t}Y(t)$ denotes the time derivative of Y at t, and where F is a vector field on $\mathbb{R}_*^{n\times p}$. Note that F can be viewed as a function from $\mathbb{R}_*^{n\times p}$ to $\mathbb{R}^{n\times p}$ since the tangent spaces of $\mathbb{R}_*^{n\times p}$ are copies of $\mathbb{R}^{n\times p}$. An *integral curve* of the vector field F—or a *solution trajectory* of (1)—passing through $Y_0 \in \mathbb{R}_*^{n\times p}$ at time t_0 is a curve $\gamma : J \to \mathbb{R}_*^{n\times p}$, defined on an interval $J \subseteq \mathbb{R}$ containing t_0, such that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\gamma(t) = F(\gamma(t)) \text{ for all } t \in J,\\
&\gamma(t_0) = Y_0.
\end{aligned}$$

From now on we assume that F is a continuously differentiable vector field defined on an open subset $\mathcal{D}$ of $\mathbb{R}_*^{n\times p}$. Given $Y_0 \in \mathbb{R}_*^{n\times p}$, we let T_{Y_0} denote the largest time such that there exists on $[0, T_{Y_0})$ an integral curve of F through Y_0 at $t = 0$. A classical existence and uniqueness result ensures that $T_{Y_0} > 0$ and that the integral curve is unique. We let $t \mapsto \varphi_t^F(Y_0)$ denote this unique integral curve of F through Y_0 at $t = 0$. The function $(Y_0, t) \mapsto \varphi_t^F(Y_0)$ on $\{(Y_0, t) \in \mathbb{R}_*^{n\times p} \times \mathbb{R} : t \in [0, T_{Y_0})\}$ into $\mathbb{R}_*^{n\times p}$ is called the *flow* of the vector field F.

In this paper, we are interested in those flows in $\mathbb{R}_*^{n\times p}$ that induce a flow on the Grassmann manifold $\mathrm{Gr}(p, n)$, in the following sense.

Definition 1.1 (subspace flows) *Let $\mathcal{D}$ be an open subset of $\mathbb{R}_*^{n\times p}$ and let F be a continuously differentiable vector field on $\mathcal{D}$. The flow φ^F of F is said to* induce *or to* realize a subspace flow *if, for all Y_a and Y_b in $\mathcal{D}$ such that $\mathsf{span}(Y_a) = \mathsf{span}(Y_b)$, it holds that $\mathsf{span}(\varphi_t^F(Y_a)) = \mathsf{span}(\varphi_t^F(Y_b))$ for all $t \in [0, \min\{T_{Y_a}, T_{Y_b}\})$. The* induced subspace flow Φ^F *is the flow on $\mathsf{span}(\mathcal{D}) := \{\mathsf{span}(Y) : Y \in \mathcal{D}\} \subseteq \mathrm{Gr}(p, n)$ that satisfies*

$$\Phi_{t+\tau}^F(\mathcal{Y}_0) = \mathsf{span}(\varphi_\tau^F(Y))$$

for all $\mathcal{Y}_0 \in \mathsf{span}(\mathcal{D})$, all $Y \in \mathsf{span}^{-1}(\Phi_t^F(\mathcal{Y}_0))$, and all $\tau \in [0, T_Y)$. The flow φ^F is termed a matrix realization *of the subspace flow Φ^F.*

If (1) induces a subspace flow, then its analysis can be decomposed into two parts: (i) the behavior of the subspace $\mathsf{span}(Y(t))$ evolving on the Grassmann manifold and (ii) the behavior of the representation $Y(t)$ of $\mathsf{span}(Y(t))$. Subspace flows can be useful, e.g., for computing the invariant subspace of a matrix A corresponding to p clustered eigenvalues. In this case, the p principal eigenvectors are ill-conditioned with respect to errors in A (see, e.g., the introduction of [Ste73]) while the principal subspace (i.e., the column space of the p principal eigenvectors) is well-conditioned. It is then expected that the flow on $\mathbb{R}^{n\times p}$ will have poor convergence properties, while the induced subspace flow will have good convergence properties.

The paper is organized as follows. The context of this work is discussed in Section 2. The general results on subspace flows and their realizations as matrix flows are presented in Section 3. We give necessary and sufficient conditions on F for (1) to induce a subspace flow, and we describe the class of functions F that induce the same subspace flow. Within this class of functions we show that there are degrees of freedom that may be used to specify properties of the matrix realization of the flow without altering the subspace flow itself. Then, in Section 4, we briefly review the case

$$F(Y) = AY \tag{2}$$

where A is assumed to be symmetric. The flow of F induces a subspace flow called the *power flow*, which can be interpreted as the gradient flow of a Rayleigh quotient on the Grassmann manifold [HM94]. In Section 5, we consider the function F implicitly defined by

$$AF(Y) - F(Y)(Y^TY)^{-1}Y^TAY = Y, \tag{3}$$

where A is a symmetric matrix and Y^T denotes the transpose of the matrix Y. The flow of F induces a subspace flow that we call the *Grassmann Rayleigh quotient flow (GRQF)*. The case $p = 1$ was analyzed in [MA03]. Using a particular matrix representation of the GRQF, we show that the Ritz values of A with respect to $\mathsf{span}(Y)$ increase with constant rate. This property is analogous to the monotonic increase of the Rayleigh quotient seen in the $p = 1$ case [Chu88, MA03]. We show that, when a Ritz value reaches an eigenvalue of the matrix A, the associated eigenvector lies in the span of the subspace flow. We briefly indicate how the flow may be deflated to generate a flow on a reduced representation of the matrix A. Conclusions are drawn in Section 6.

2 History and Context

The origins of (2) can be traced back to the power method, which is probably the simplest of all methods for solving for eigenvalues and eigenvectors. The basic idea of the power method is to apply A repeatedly to a well chosen starting vector, namely

$$x_{k+1} = \frac{Ax_k}{\|Ax_k\|}, x_0 \in S^{n-1},$$

where $\|\cdot\|$ denotes the Euclidean 2-norm and $S^{n-1} = \{x \in \mathbb{R}^n : x^Tx = 1\}$ denotes the unit sphere in $\mathbb{R}^n$. If $\lambda_1 > \lambda_2$, then the power method converges to $\pm v_1$ for almost all initial condition; see, e.g., [Wil65]. Note that the norm of the iterates of the power method is irrelevant: all the information is conveyed in the direction of x_k and the purpose of the normalization is merely to prevent overflow or underflow. This means that the power method can be regarded as a *subspace iteration* evolving on the set of one-dimensional subspaces of $\mathbb{R}^n$ [AM86, HM94].

When the dominant eigenvalues of A are clustered, or when a multi-dimensional dominant eigenspace is sought, it is advisable to iterate on multi-dimensional subspaces. A natural p-dimensional version of the power method is the *direct subspace iteration* [Par80] defined by $\mathcal{Y}_+ = A\mathcal{Y}$, where $\mathcal{Y}$ and $\mathcal{Y}_+$ are p-dimensional subspaces of $\mathbb{R}^n$ and $A\mathcal{Y} := \{Ay : y \in \mathcal{Y}\}$.

The direct subspace iteration is an example of *subspace iteration*, i.e., an iterative process evolving on the set of p-dimensional subspaces of $\mathbb{R}^n$. Several subspace iterations have been built on the framework described in Algorithm 2.1 below. Observe that the set of matrices that have the same column space as $Y \in \mathbb{R}^{n\times p}_*$ is

$$\mathsf{span}^{-1}(\mathsf{span}(Y)) = \{YM : M \in \mathrm{GL}_p\} =: Y\mathrm{GL}_p,$$

where GL_p denotes the set of all invertible $p \times p$ matrices.

Algorithm 2.1 (Subspace Iteration) *Requires: function $f : \mathbb{R}^{n\times p}_* \mapsto \mathbb{R}^{n\times p}_*$ such that*

$$\forall Y \in \mathbb{R}^{n\times p}_*, \forall M \in \mathrm{GL}_p : \exists N \in \mathrm{GL}_p : f(YM) = f(Y)N. \tag{4}$$

Input: $\mathcal{Y}_0 \in \mathrm{Gr}(p,n)$
Output: $\mathcal{Y}_{k_{max}}$
For $k = 0, 1, \ldots, k_{max} - 1$
  Pick $Y_k \in \mathbb{R}_*^{n\times p}$ *such that* $\mathcal{Y}_k = \mathsf{span}(Y_k)$
  $Y_+ = f(Y_k)$
  $\mathcal{Y}_{k+1} = \mathsf{span}(Y_+)$
End

Condition (4) guarantees that $\mathcal{Y}_{k+1}$ does not depend on the choice of Y_k in $\mathsf{span}^{-1}(\mathcal{Y}_k)$.

The correspondence between the iteration function f and the subspace iteration is not one-to-one. Two functions f and $\hat{f}$ define the same subspace iteration through Algorithm 2.1 if and only if

$$\forall Y \in \mathbb{R}_*^{n\times p} : \exists N \in \mathrm{GL}_p : \hat{f}(Y) = f(Y)N.$$

Several iterations for invariant subspace computation are instances of Algorithm 2.1. This includes the direct subspace iteration mentioned above, obtained with f as in (2); shifted inverse iterations [Ips97, AM86], where A is replaced by $(A-\mu I)^{-1}$; a Grassmannian version of the Rayleigh quotient iteration [Smi97, AMSV02] obtained with f given by (3) (see also the two-sided version presented in [AV07]); Grassmannian Newton methods [EAS98, LST98, LE02, AMS02]; and the Grassmannian implicit trust-region method proposed in [BAG07]. See also [ASVM04] for an overview.

Since the early 1980's, there has been considerable interest in studying continuous-time flows related to discrete-time iterations. The result that ignited interest in such flows was when iterates of the unshifted QR-algorithm (which is closely related to the power method; see [Wat82]) were shown to be unit time samples of a particular Lax-pair equation [Fla74, Sym82, DNT83, Nan85]. This work sparked extensive research on using dynamical systems to solve linear algebraic problems; see, e.g., [Bro89, WE88, Chu88, CD90, Chu92, CU92, HM94, Chu94, Deh95, CG02, Prz03, MA03, BI04, CDLP05, MHM05, GL06]. More generally, there is a vast literature on continuous-time algorithms, spanning several areas of computational science, including, but not limited to, linear programming [BL89a, BL89b, Bro91, Fay91b], continuous nonlinear optimization [Fay91a, LW00], discrete optimization [Hop84, HT85, Vid95, AS04], signal processing [AC98, Dou00, CG03], model reduction [HM94, YL99] and automatic control [HM94, MH98, GS01]. There is a computability theory for continuous-time algorithms, initiated by the work of Shannon on the general-purpose analog computer [Sha41]; see [Orp97, Cam04, MC05] and references therein. A theory of computational complexity for continuous-time algorithms has recently started developing [BHSF02]. Continuous-time flows are also used to study the asymptotic behavior of their discrete counterpart, referring to the theory of Ljung [Lju77] and Kushner and Clark [KC78]; see for example [OK85]. Finding adequate discretizations of flows endowed with computational properties that preserve or even improve those properties is also an important area of research; see [Hig99, FK05, KQL$^+$06, LKLT06] and references therein.

Equation (1) can be thought of as the continuous-time analogue of Algorithm 2.1 when $f = F$. In this sense, the power flow, given by (1) and (2), is a continuous analogue of the power method, defined by Algorithm 2.1 and (2). This is a strong analogy in the sense that the iterates of the power method are the unit-time samples of the power flow of $\log(A)$. Likewise, the Grassmann Rayleigh quotient flow, given by (1) and (3), can be viewed as a continuous-time analogue of the Grassmann Rayleigh quotient iteration, defined by Algorithm 2.1 and (3).

Here, the analogy is weaker since the interpolation property does not hold. Nevertheless the Grassmann Rayleigh quotient flow has interesting convergence properties, as shown in Section 5.

3 Subspace Flows and Their Matrix Realizations

In this section, we give necessary and sufficient conditions for (1) to induce a subspace flow and we describe the classes of functions that induce the same subspace flow. First we introduce some notation and definitions as well as two preparation lemmas. Given $Y \in \mathbb{R}_*^{n\times p}$, we let $\Pi_{Y_\perp} := I - Y(Y^TY)^{-1}Y^T$ denote the orthogonal projector onto the orthogonal complement of $\mathsf{span}(Y)$. The tangent space to $Y\mathrm{GL}_p$ at $Y \in \mathbb{R}_*^{n\times p}$ is called the *vertical space* [KN63]

$$\mathcal{V}_Y = \{YM : M \in \mathbb{R}^{p\times p}\},$$

and we choose the *horizontal space* as its complement

$$\mathcal{H}_Y = \{Z \in \mathbb{R}^{n\times p} : Y^TZ = 0\}.$$

Observe that $\mathcal{H}_Y$ is the range of $\Pi_{Y_\perp}$. We refer the reader to [AMS08] for details about manifolds, vector fields, and differential maps.

Lemma 3.1 *The Grassmann manifold* $\mathrm{Gr}(p,n)$ *is diffeomorphic to the set* $\mathcal{P}_{p,n}$ *of all rank-p orthogonal projectors in* $\mathbb{R}^n$ *(viewed as a submanifold of* $\mathbb{R}^{n\times n}$*), with diffeomorphism* $\mathsf{span}(Y) \in \mathrm{Gr}(p,n) \mapsto P_Y = Y(Y^TY)^{-1}Y^T \in \mathcal{P}_{p,n}$.

Proof. See, e.g., [HM94, Ch. 1]. □

Definition 3.2 (Ψ-related vector fields) *Let* F, G *be vector fields on manifolds* $\mathcal{M}$ *and* $\mathcal{N}$ *respectively and let* $\Psi : \mathcal{M} \to \mathcal{N}$ *be a smooth map. Then* F *and* G *are* Ψ-related *if the following diagram commutes:*

$$\begin{tikzcd} \mathcal{M} \arrow[r, "\Psi"] \arrow[d, "F"'] & \mathcal{N} \arrow[d, "G"] \\ T\mathcal{M} \arrow[r, "\mathrm{D}\Psi"'] & T\mathcal{N} \end{tikzcd}$$

where $\mathrm{D}\Psi$ *denotes the differential of* Ψ*. In other words,* $\mathrm{D}\Psi(x)[F] = G(\Psi(x))$ *for all* $x \in \mathcal{M}$.

Lemma 3.3 (integral curves of Ψ-related vector fields) *Let* F *and* G *be vector fields on* $\mathcal{M}$ *and* $\mathcal{N}$ *respectively and let* $\Psi : \mathcal{M} \to \mathcal{N}$ *be a smooth map. Then the vector fields* F *and* G *are* Ψ*-related if and only if, for every integral curve* γ *of* F, $\Psi \circ \gamma$ *is an integral curve of* G.

Proof. See, e.g., [AMR88]. □

We are now ready to give a characterization of vectors fields on $\mathbb{R}_*^{n\times p}$ that induce the same subspace flow.

Theorem 3.4 *Let F and $\hat{F}$ be continuously differentiable functions from an open domain $\mathcal{D} \subseteq \mathbb{R}_*^{n\times p}$ to $\mathbb{R}^{n\times p}$. Let $F(Y) = F^h(Y) + YU_Y$ denote the decomposition of $F(Y)$ into a* horizontal part *$F^h(Y) = \Pi_{Y_\perp} F(Y)$ and a* vertical part *$YU_Y = Y(Y^TY)^{-1}Y^TF(Y)$. Let $\hat{F}(Y) = \hat{F}^h(Y) + Y\hat{U}_Y$ denote the same decomposition for $\hat{F}(Y)$. For all $Y_0 \in \mathcal{D}$, let T_{Y_0}, resp. $\hat{T}_{Y_0}$, be the largest time such that for all $t \in [0, T_{Y_0})$, the differential system*

$$\dot{Y} = F(Y) \equiv F^h(Y) + YU_Y, \tag{5}$$

resp. $\dot{Y} = \hat{F}(Y) \equiv \hat{F}^h(Y) + Y\hat{U}_Y$, has one and only one solution $\varphi_t^F(Y_0)$, resp. $\varphi_t^{\hat{F}}(Y_0)$, in $\mathcal{D}$ with initial condition $\varphi_0^F(Y_0) = \varphi_0^{\hat{F}}(Y_0) = Y_0$. The flow of F induces a subspace flow (Definition 1.1) if and only if

$$F^h(YM) = F^h(Y)M \tag{6}$$

for all $Y \in \mathcal{D}$ and all $M \in \mathrm{GL}_p$ such that $YM \in \mathcal{D}$. Moreover, the subspace flows Φ^F and $\Phi^{\hat{F}}$ coincide if and only if

$$\hat{F}^h(Y) = F^h(Y) \tag{7}$$

for all $Y \in \mathcal{D}$.

Proof. We first prove the "only if" statements. Let $Y \in \mathcal{D}$ and $M \in \mathrm{GL}_p$ be such that YM belongs to $\mathcal{D}$. Assume that the flow of F induces a subspace flow. We show that (6) holds. Since the flow of F induces a subspace flow, it follows that, for all $t \in [0, T_Y)$, there exists $N(t) \in \mathrm{GL}_p$ such that

$$\varphi_t^F(YM) = \varphi_t^F(Y)N(t). \tag{8}$$

Equation (8) at $t = 0$ gives $YM = YN(0)$. Since Y has full rank, it follows that $N(0) = M$. Since the curves $t \mapsto \varphi_t^F(YM)$ and $t \mapsto \varphi_t^F(Y)$ are continuously differentiable, it follows that the curve $t \mapsto N(t)$ is continuously differentiable. Taking the time derivative of (8) at $t = 0$, we obtain $F(YM) = F(Y)N(0) + Y\dot{N}(0)$. Applying $\Pi_{Y_\perp}$ on the left of this equation gives the homogeneity property (6).

Assume further that φ^F and $\varphi^{\hat{F}}$ induce the same subspace flow. Then, for all $t \in [0, \min\{T_Y, \hat{T}_Y\})$, there exists $\tilde{N}(t) \in \mathrm{GL}_p$ such that

$$\varphi_t^{\hat{F}}(Y) = \varphi_t^F(Y)\tilde{N}(t), \tag{9}$$

with $N(0) = I$. Taking the time derivative of (9) and applying $\Pi_{Y_\perp}$ on the left yields (7).

We now prove the converse statements. Assume that $F^h(YM) = F^h(Y)M$ for all $Y \in \mathcal{D}$ and all $M \in \mathrm{GL}_p$ such that $YM \in \mathcal{D}$. Consider the mapping

$$\Psi : \mathbb{R}_*^{n\times p} \to \mathcal{P}_{p,n} : Y \mapsto Y(Y^TY)^{-1}Y^T.$$

Observe that the mapping Ψ identifies to the span mapping through the diffeomorphism defined in Lemma 3.1. In particular, $\Psi^{-1}(\Psi(Y)) = Y\mathrm{GL}_p$. Routine matrix manipulations lead to

$$
\begin{aligned}
\mathrm{D}\Psi(Y)[F(Y)] &= F(Y)(Y^TY)^{-1}Y^T \\
&\quad - Y(Y^TY)^{-1}((F(Y))^TY + Y^TF(Y))(Y^TY)^{-1}Y^T \\
&\quad + Y(Y^TY)^{-1}(F(Y))^T \\
&= F^h(Y)(Y^TY)^{-1}Y^T \\
&\quad - Y(Y^TY)^{-1}((F^h(Y))^TY + Y^TF^h(Y))(Y^TY)^{-1}Y^T \\
&\quad + Y(Y^TY)^{-1}(F^h(Y))^T \\
&\quad + YU_Y(Y^TY)^{-1}Y \\
&\quad - Y(Y^TY)^{-1}(U_Y^TY^TY + Y^TYU_Y)(Y^TY)^{-1}Y^T \\
&\quad + Y(Y^TY)^{-1}U_Y^TY^T \\
&= F^h(Y)(Y^TY)^{-1}Y^T + Y(Y^TY)^{-1}(F^h(Y))^T =: \tilde{G}(Y), \qquad (10)
\end{aligned}
$$

where we have used the identity $Y^TF^h(Y) = 0$. Note that $\tilde{G}$ is a function from $\mathcal{D} \subseteq \mathbb{R}_*^{n\times p}$ to $T\mathcal{P}_{p,n}$ that maps any $Y \in \mathcal{D}$ to $\tilde{G}(Y) \in T_{\Psi(Y)}\mathcal{P}_{p,n}$. It is readily checked that $\tilde{G}$ satisfies $\tilde{G}(YM) = \tilde{G}(Y)$ whenever Y and YM belong to $\mathcal{D}$. It follows that there exists a projected map $G : \Psi(\mathcal{D}) \to \mathcal{P}_{p,n}$ such that $G(\Psi(Y)) = \tilde{G}(Y)$ for all $Y \in \mathcal{D}$. Hence we have $\mathrm{D}\Psi(Y)[F(Y)] = G(\Psi(Y))$, i.e., F and G are Ψ-related (Definition 3.2).

$$
\begin{tikzcd}
 & \mathrm{Gr}(p,n) \arrow[d, leftrightarrow, "\text{diffeo}"] \\
\mathbb{R}_*^{n\times p} \arrow[ur, "\mathsf{span}"] \arrow[r, "\Psi"] \arrow[d, "F"'] \arrow[dr, "\tilde{G}"] & \mathcal{P}_{p,n} \arrow[d, "G"] \\
T\mathbb{R}_*^{n\times p} \arrow[r, "\mathrm{D}\Psi"'] & T\mathcal{P}_{p,n}
\end{tikzcd}
$$

Since F is continuously differentiable and Ψ is smooth, it follows that G is a continuously-differentiable vector field, which guarantees uniqueness of the integral curves of G. From Lemma 3.3, we conclude that $\Psi(\varphi_t^F(Y)) = \varphi_t^G(\Psi(Y))$ for all $t \in [0, T_Y)$. From Lemma 3.1, it follows that $\mathsf{span}(\varphi_t^F(Y)) = \varphi_t^G(\mathsf{span}(Y))$ for all $t \in [0, T_Y)$, i.e., the flow of F induces a subspace flow.

Moreover, if (7) holds, then, by (10), we have $\mathrm{D}\Psi(Y)[F(Y)] = \mathrm{D}\Psi(Y)[\hat{F}(Y)]$, and thus the flows of F and $\hat{F}$ induce the same subspace flow. $\square$

Note that if $\dot{Y} = F(Y) = F^h(Y) + YU_Y$, then $\frac{\mathrm{d}}{\mathrm{d}t}\det(Y^TY) = \mathrm{trace}(\mathrm{adj}(Y^TY)(\dot{Y}^TY + Y^T\dot{Y})) = 2\,\mathrm{trace}(\mathrm{adj}(Y^TY)Y^TYU_Y) = 2\,\mathrm{trace}(\det(Y^TY)U_Y) = 2\,\mathrm{trace}(U_Y)\det(Y^TY)$. Hence $\det(Y(t)^TY(t)) = \det(Y(0)^TY(0))\exp[\int_0^t 2\mathrm{trace}(U_{Y(\tau)})\,\mathrm{d}\tau]$. Thus, the requirement that $\varphi_t^F(Y_0)$ is full-rank for all $t \in [0, T_{Y_0})$ only depends on the vertical dynamics. In particular, if the integral $\int_0^T \mathrm{trace}(U_{Y(\tau)})\,\mathrm{d}\tau$ exists, then $Y(t)$ is full rank for all $t \in [0, T)$.

In the parlance of principal fiber bundle theory (see [KN63] or the more introductory [Was04]), the horizontality of $F^h(Y)$ and the homogeneity condition (6) ensure that $F^h(Y)$ is a *horizontal lift* of a tangent vector field on $\mathrm{Gr}(p,n) = \mathbb{R}_*^{n\times p}/\mathrm{GL}_p$; see [KN63, AMS04]. Theorem 3.4 states that (i) $\dot{Y} = F(Y)$ defines a subspace flow if and only if the horizontal component of F is a horizontal lift and (ii) $\dot{Y} = F(Y)$ and $\dot{Y} = \hat{F}(Y)$ define the same subspace flow if and only if they have the same horizontal component.

Consider the differential system

$$\dot{Y} = F(Y) + YR,\ Y(0) = Y_0, \tag{11}$$

where the horizontal part $F^h(Y)$ satisfies the homogeneity condition (6). It follows from Theorem 3.4 that the span of the solution $Y(t)$ of (11) does not depend on R. This flexibility can be exploited for analysis purposes (e.g., in order to simplify the equations governing the dynamics) or design purposes (e.g., choosing well-conditioned representations), as we will see in the following sections.

4 The Power Flow

As an illustration, we consider the well-known case where $F(Y) := AY$, i.e., the differential system (11) becomes

$$\dot{Y} = AY + YR, \tag{12}$$

where A is assumed to be symmetric, with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_n$ and associated orthonormal eigenvectors $v_1, \ldots, v_n$. From Theorem 3.4, one obtains that the horizontal component $F^h(Y) = \Pi_{Y_\perp} AY$ verifies the homogeneity condition. Therefore, (12) defines a subspace flow, called the *power flow*. This fact was already observed, e.g., in [Deh95, DMV99].

First consider the choice $R = 0$, i.e.

$$\dot{Y} = AY. \tag{13}$$

The solution of (13) is $Y(t) = e^{At}Y(0) = \sum_{i=1}^n e^{\lambda_i t} v_i v_i^T Y(0)$. Assume that $\lambda_p > \lambda_{p+1}$. Then for almost all starting point $Y(0)$, one has $\lim_{t\to\infty} \mathsf{span}(Y(t)) = \mathsf{span}(v_1|\ldots|v_p)$, i.e., $\mathsf{span}(Y(t))$ converges to the principal subspace of A. For this reason, one says that the power flow has Principal Subspace Analysis (PSA) properties. However, the matrix $Y(t)$ itself becomes arbitrarily close to rank-deficiency as time evolves: $Y(t) = e^{\lambda_1 t}((v_1|\ldots|v_1)\mathrm{diag}(v_1^T Y(0)) + \epsilon(t))$ with $\lim_{t\to\infty} \epsilon(t) = 0$. This is problematic because the span of an ill-conditioned matrix is very sensitive to perturbations in that matrix.

Fortunately, the vertical degrees of freedom contained in R can be taken advantage of in order to achieve well-conditioned representations (orthonormal representations in particular) without modifying the induced subspace flow. In fact, R can be chosen regardless of F in such a way that the solution of (11) always converges to the set of orthonormal matrices, as we now show. Take $R = (Y^TY)^{-1}(-Y^TF(Y) + S + \Omega)$ where $\Omega = -\Omega^T$ is skew-symmetric and $S = S^T$ is symmetric. Then (11) becomes

$$\dot{Y} = F^h(Y) + Y(Y^TY)^{-1}\Omega + Y(Y^TY)^{-1}S \tag{14}$$

where $F^h(Y) = \Pi_{Y_\perp} F(Y)$. Define $W = Y^TY - I$. The dynamics (14) yields $\dot{W} = 2S$. The choice $S = -cW$, $c > 0$, produces $\dot{W} = -2cW$, which implies that $W(t) = e^{-2ct}W(0)$, and then $Y(t)^TY(t)$ goes to I as t goes to infinity. This means that for the choice $S = -c(Y^TY - I)$, the solution trajectories of (14) converge to the set of orthonormal matrices for all initial conditions in $\mathbb{R}_*^{n\times p}$. In conclusion, for almost all initial conditions, the solution of

$$\dot{Y} = \Pi_{Y_\perp} AY + cY(Y^TY)^{-1}(I - Y^TY) + Y(Y^TY)^{-1}\Omega,\ \Omega^T = -\Omega,\ c > 0, \tag{15}$$

converges to the set of orthonormal bases of the dominant eigenspace of A.

5 The Grassmann-Rayleigh Quotient Flow

As an illustrative application of Theorem 3.4, we now analyze the dynamics (11) for F defined by (3), namely

$$\dot{Y} = Z_Y + YR \tag{16a}$$

$$AZ_Y - Z_Y(Y^TY)^{-1}Y^TAY = Y. \tag{16b}$$

The solution of (16b) possesses the homogeneity property $Z_{YM} = Z_Y M$, therefore by Theorem 3.4 the flow defined by (16) induces a unique subspace flow, regardless of R. We call this flow the Grassmann-Rayleigh Quotient Flow (GRQF).

We will take advantage of the vertical degrees of freedom contained in R (see Section 3) in order to facilitate the analysis of (16) by making (16b) simpler. If $R_A(Y) := (Y^TY)^{-1}Y^TAY$ is diagonal, then (16b) decouples into p independent linear systems where the unknowns are the p columns of Z_Y; see e.g. [LE02, AMSV02]. Therefore, we will attempt to keep $(Y^TY)^{-1}Y^TAY$ diagonal. Moreover, we will require Y^TY to remain constant. From (16), one obtains

$$\frac{\mathrm{d}}{\mathrm{d}t}(Y^TY) = Y^TZ_Y + Y^TYR + Z_Y^TY + R^TY^TY$$

and, assuming that $\frac{\mathrm{d}}{\mathrm{d}t}(Y^TY) = 0$ holds,

$$\frac{\mathrm{d}}{\mathrm{d}t}((Y^TY)^{-1}Y^TAY) = (Y^TY)^{-1}Y^TA(Z_Y + YR) + (Y^TY)^{-1}(Z_Y^T + R^TY^T)AY.$$

From (16b), one has $AZ_Y = Y + Z_Y(Y^TY)^{-1}(Y^TAY)$. Substitute this expression in the above equation to obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}((Y^TY)^{-1}Y^TAY) =& I + (Y^TY)^{-1}Y^TZ_Y(Y^TY)^{-1}Y^TAY + (Y^TY)^{-1}Y^TAYR \\ &+ I + (Y^TY)^{-1}(Y^TAY)(Y^TY)^{-1}Z_Y^TY + (Y^TY)^{-1}R^TY^TAY.\end{aligned}$$

It is now easy to see that the choice $R = -(Y^TY)^{-1}Z_Y^TY$ achieves $\frac{\mathrm{d}}{\mathrm{d}t}(Y^TY) = 0$ and $\frac{\mathrm{d}}{\mathrm{d}t}((Y^TY)^{-1}Y^TAY) = 2I$, which keeps $R_A(Y(t))$ diagonal. We have thus proved the following result.

Proposition 5.1 (structure-preserving representation of the GRQF) *The solution $Y(t)$ of (16) with $R = -(Y^TY)^{-1}Z_Y^TY$, i.e.,*

$$\dot{Y} = Z_Y - Y(Y^TY)^{-1}Z_Y^TY \tag{17a}$$

$$AZ_Y - Z_Y(Y^TY)^{-1}Y^TAY = Y. \tag{17b}$$

satisfies the following properties

1. $\frac{\mathrm{d}}{\mathrm{d}t}(Y^TY) = 0$.
2. $\frac{\mathrm{d}}{\mathrm{d}t}[(Y^TY)^{-1}Y^TAY] = 2I$.

Define
$$\mathcal{C} := \{Y \in \mathbb{R}^{n\times p} : Y^TY = I \text{ and } Y^TAY \text{ is diagonal}\}. \tag{18}$$
Given $Y \in \mathcal{C}$, the diagonal entries of Y^TAY are called the *Ritz values* of A with respect to $\mathsf{span}(Y)$, and the columns of Y are associated *Ritz vectors*.

The following result directly follows from Proposition 5.1.

Proposition 5.2 (Ritz representation of the GRQF) *Let $Y(0)$ be such that $Y(0)^TY(0) = I$ and $Y(0)^TAY(0) = \Sigma(0) = \operatorname{diag}(\sigma_1(0), \dots, \sigma_p(0))$, i.e., $Y(0)$ belongs to $\mathcal{C}$ defined in* (18)*. Then the solution $Y(t)$ of* (17) *satisfies $Y(t)^TY(t) = I$ and*
$$R_A(Y(t)) = Y(t)^TAY(t) = \operatorname{diag}(\sigma_1(0), \dots, \sigma_p(0)) + 2It =: \operatorname{diag}(\sigma_1(t), \dots, \sigma_p(t))$$
with $\sigma_i(t) = \sigma_i(0) + 2t$. In particular, (17) *leaves $\mathcal{C}$ invariant. On $\mathcal{C}$,* (17) *simplifies to*
$$\dot{Y} = Z_Y - YZ_Y^TY \tag{19a}$$
$$AZ_Y - Z_Y\operatorname{diag}(\sigma_1(t), \dots, \sigma_p(t)) = Y. \tag{19b}$$

Equation (19b) decouples into p equations
$$(A - \sigma_i I)Z_{:,i} = Y_{:,i}, \quad i = 1, \dots, p,$$
which shows that Z_Y is well defined by (19b) as long as
$$\sigma_i(t) = \sigma_i(0) + 2t \notin \operatorname{spec}(A), \ i = 1, \dots, p. \tag{20}$$
The times t at which a Ritz value belongs to the spectrum of A are called *critical times*. When $p = 1$, one obtains the situation described in [MA03].

We will now use the Ritz representation to demonstrate the deflating property of the GRQF.

Theorem 5.3 *Let A be an $n \times n$ symmetric matrix with distinct eigenvalues $\lambda_1, \dots, \lambda_n$ and associated eigenvectors $v_1, \dots, v_n$. Let $\mathcal{Y}_0 \in \mathrm{Gr}(p, n)$ be such that the Ritz values $\sigma_1(0), \dots, \sigma_p(0)$ of A with respect to $\mathcal{Y}_0$ and the eigenvalues of A are disjoint. Let $t^* = \min\{(\lambda_i - \sigma_j(0))/2 : \lambda_i - \sigma_j(0) \geq 0, \ i = 1\dots, n, \ j = 1, \dots, p\}$. Then over the time interval $[0, t^*)$ the Grassmann Rayleigh quotient flow, i.e., the subspace flow induced by* (16)*, admits one and only one solution $\mathcal{Y}(t) \in \mathrm{Gr}(p, n)$ with initial condition $\mathcal{Y}(0) = \mathcal{Y}_0$.*
Assume moreover that there is only one (i', j') such that $t^* = (\lambda_{i'} - \sigma_{j'}(0))/2$. Let $\sigma_j(t)$, $j = 1, \dots, p$, denote the Ritz values of A with respect to $\mathcal{Y}(t)$, and let $x_1(t), \dots, x_p(t) \in \mathbb{R}^n$ be associated Ritz vectors. Then, either $\lim_{t\to t^*} x_{j'}(t) = \pm v_{i'}$, or $\lim_{t\to t^*} \angle(x_{j'}(t), v_{i'}) = \pi/2$. The latter case is unstable under perturbations.

Proof. To simplify the development, without loss of generality, we express A in the coordinate system achieving $v_1 = e_1, \dots, v_n = e_n$, where $\{e_1, \dots, e_n\}$ is the canonical basis of $\mathbb{R}^n$. Thus $A = \operatorname{diag}(\lambda_1, \dots, \lambda_n)$. We choose Y_0 such that $\mathcal{Y}_0 = \mathsf{span}(Y_0)$, $Y_0^TY_0 = I$, $Y_0^TAY_0 = \Sigma(0) = \operatorname{diag}(\sigma_1(0), \dots, \sigma_p(0))$. That is, the columns of Y_0 are Ritz vectors of A with respect to $\mathcal{Y}_0$. The differential system (19) with initial condition $Y(0) = Y_0$ admits one and only one solution $Y(t)$ on the time interval $[0, t^*)$ where t^* (as defined in the statement) is the first critical time.

Moreover $Y(t)$ has full rank since it is orthonormal. Therefore, over the time interval $[0, t^*)$, the Grassmann Rayleigh quotient flow admits the unique solution $\mathcal{Y}(t) = \mathsf{span}(Y(t))$ with initial condition $\mathcal{Y}(0) = \mathcal{Y}_0$.

We now prove the second claim. It follows from Proposition 5.2 that $Y(t) = [x_1(t)|\dots|x_p(t)]$ where the x_i's match the definition given in the statement. Since there is no constraint on the way the λ's and the σ's are ordered, we will assume that $i' = j' = 1$. Denote by Y_{ij} the ij element of Y and by Y_j the jth column of Y. The element-by-element expression of (19) is

$$\dot{Y}_{ij} = (\lambda_i - \sigma_j)^{-1} Y_{ij} - \sum_{km} Y_{ik}(\lambda_m - \sigma_k)^{-1} Y_{mk} Y_{mj} \tag{21}$$

where $\sigma_l = Y_l^T A Y_l$. Equivalently,

$$\dot{Y}_{11} = \frac{1}{\lambda_1 - \sigma_1} Y_{11}(1 - Y_{11}^2) \qquad - \sum_{(k,m)\neq(1,1)} Y_{ik}(\lambda_m - \sigma_k)^{-1} Y_{mk} Y_{mj} \tag{22a}$$

$$\dot{Y}_{ij} = -\frac{1}{\lambda_1 - \sigma_1} Y_{i1} Y_{1j} Y_{11} \qquad + \frac{1}{\lambda_i - \sigma_j} Y_{ij} - \sum_{(k,m)\neq(1,1)} Y_{ik}(\lambda_m - \sigma_k)^{-1} Y_{mk} Y_{mj}, \tag{22b}$$

where the first terms the right-hand sides are the ones that reach infinity at the first critical time t^*. Define the new time variable

$$\tau = -\ln(\lambda_1 - \sigma_1(t)) = -\ln(2(t^* - t)),\ t \le t^*,$$

and denote D_τ by $'$. Then, with $\xi := \lambda_1 - \sigma_1$, (22) becomes

$$Y_{11}' = Y_{11}(1 - Y_{11}^2) + \xi K_{11}(Y) \tag{23a}$$

$$Y_{ij}' = -Y_{i1} Y_{1j} Y_{11} + \xi K_{ij}(Y) \tag{23b}$$

$$\xi' = -\xi \tag{23c}$$

where $K(Y)$ is bounded on $t \in [0, t^*]$ because Y is orthonormal and $\lambda_m - \sigma_k$ for $(k, m) \neq (1, 1)$ evolves linearly with t (see (20)) and does not vanish for $t \in [0, t^*]$ by hypothesis.

Recall that there are constraints on Y and ξ, namely $(Y, \xi) \in \mathcal{C}'$ with

$$\mathcal{C}' := \{(Y, \xi) : Y \in \mathcal{C} \text{ and } \xi = \lambda_1 - Y_1^T A Y_1\}. \tag{24}$$

For a moment, let us relax the constraints on $(Y(0), \xi(0))$ and view (23) as a system in $\mathbb{R}^{n\times p} \times \mathbb{R}$. The set of equilibrium points is $\mathcal{U} \cup \mathcal{S}$ with

$$\mathcal{S} := \{(Y, \xi) : Y \in \mathbb{R}^{n\times p}, Y_{11} = \pm 1, \xi = 0\}$$
$$\mathcal{U} := \{(Y, \xi) : Y \in \mathbb{R}^{n\times p}, Y_{11} = 0, \xi = 0\}.$$

The linearization of (23) at a point $(\bar{Y}, \bar{\xi}) \in \mathcal{S}$ yields

$$\begin{aligned} Y_{11}' &= -2Y_{11} + \xi \bar{K}_{11} \\ Y_{ij}' &= \dots \\ \xi' &= -\xi \end{aligned}$$

and the set $\mathcal{S}$ is stable. The linearization of (23) at a point $(\bar{Y}, \bar{\xi}) \in \mathcal{U}$ yields

$$\begin{aligned} Y'_{11} &= Y_{11} + \xi \bar{K}_{11} \\ Y'_{ij} &= -\bar{Y}_{i1}\bar{Y}_{1j}Y_{11} + \xi \bar{K}_{ij} \\ \xi' &= -\xi \end{aligned}$$

and the set $\mathcal{U}$ is unstable, with a certain stable manifold W^s.

Now we reintroduce the constraint $(Y, \xi) \in \mathcal{C}'$. Since K_{11} is bounded and $\xi = e^{-\tau}$, equation (23a) implies that either

(i) $Y_{11} \to 0$. This happens only if $(Y(0), \xi(0))$ belongs to the stable manifold W^s of the unstable set $\mathcal{U}$. Or

(ii) $Y_{11} \to \pm 1$. This means that x_1, the first column of Y, converges to the eigenvector $\pm e_1$ of A.

Since $(Y(0), \xi(0)) \in \mathcal{C}'$, it remains to show that $\mathcal{C}'$ is not included in W^s. To this end, we show that $\mathcal{C}'$ is transverse to W^s at $\mathcal{U}$. Consider a small perturbation μ on Y_{11} from an equilibrium point. Then

$$Y^T A Y = \begin{pmatrix} \sigma_1 + \mu^2 \lambda_1 & \mu \lambda_1 Y_{12} & \dots & \mu \lambda_1 Y_{1p} \\ \mu \lambda_1 Y_{12} & \sigma_2 & & \\ \vdots & & \ddots & \\ \mu \lambda_1 Y_{1p} & & & \sigma_p \end{pmatrix}.$$

This implies that σ_1, thus ξ, is perturbed to the second order only. On the other hand, the stable manifold of (23) verifies to the first order $\xi = -\frac{2}{K_{11}} Y_{11}$. This means that the case (ii) is generic. $\square$

At time t^*, one has $\sigma_{j'}(t^*_-) = \sigma_{j'}(0) + 2t^* = \lambda_{i'}$, hence the solution Z_Y of (16b) is not well defined. Moreover, the Ritz vector $x_{j'}(t^*_-) \in \mathcal{Y}(t^*_-)$ is collinear with the eigenvector $v_{i'}$ of A. Choose W such that $[v_{i'}, W]$ is orthogonal. The deflation of the GRQF is the flow associated with the matrices $\hat{A} = W^T A W$ and $\hat{Y}(0) = W^T Y(t^*_-)$. The GRQF associated with $\hat{A}$, $\hat{Y}$ is a flow on $\mathrm{Gr}(p-1, n-1)$ that has the same properties as (16).

Concerning the condition in Theorem 5.3 that the Ritz values $\sigma_1(0), \dots, \sigma_p(0)$ of A with respect to $\mathcal{Y}_0$ and the eigenvalues of A be disjoint, we point out that the condition is false if and only if $\mathcal{Y}_0 = \mathsf{span}(Y_0)$ where Y_0 (of full rank) satisfies the polynomial equation $\prod_{i=1}^n \det(Y_0^T A Y_0 - \lambda_i Y_0^T Y_0) = 0$. The condition is thus not difficult to satisfy.

6 Conclusion

We have given a characterization of the flows on the set $\mathbb{R}_*^{n \times p}$ of full-rank $n \times p$ matrices that induce a subspace flow. Using this result, we have studied two continuous-time flows on the Grassmann manifold: the power flow and the novel Grassmann Rayleigh quotient flow (GRQF). The realization of these flows by differential equations in $\mathbb{R}_*^{n \times p}$ leaves degrees of freedom that can be exploited in order to achieve specific computations (principal component extraction) or simplify the analysis. The dynamics in the matrix space can be controlled, without altering the subspace flow, in such a way that the matrix representation converges to the set of orthonormal matrices. Moreover, we have constructed a matrix representation of the GRQF that clearly reveals the monotonic increase of the Ritz values along its solution and the finite time convergence to a subspace containing an eigenvector. In this sense the GRQF achieves finite-time deflation.

Acknowledgements

This work was initiated while the first author was a Research Fellow with the Belgian National Fund for Scientific Research (Aspirant du F.N.R.S.) at the University of Liège. Part of this work was performed while the first author was a guest in the Mathematics Department of the University of Würzburg under a grant from the European Nonlinear Control Network. The hospitality of the members of the department is gratefully acknowledged. Special thanks to K. Gallivan, U. Helmke, A. Srivastava, and S. Yoshizawa for useful discussions related to this topic.

References

[AC98] Shun-ichi Amari and Andrzej Cichocki, *Adaptive blind signal processing—neural network approaches*, Proc. IEEE **86** (1998), no. 10, 2026–2048.

[AM86] Gregory Ammar and Clyde Martin, *The geometry of matrix eigenvalue methods*, Acta Appl. Math. **5** (1986), no. 3, 239–278.

[AMR88] R. Abraham, J. E. Marsden, and T. Ratiu, *Manifolds, tensor analysis, and applications*, second ed., Applied Mathematical Sciences, vol. 75, Springer-Verlag, New York, 1988.

[AMS02] P.-A. Absil, R. Mahony, and R. Sepulchre, *The continuous-time Rayleigh quotient flow on the Grassmann manifold*, Proceedings of the 15th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2002), University of Notre Dame, August 2002.

[AMS04] P.-A. Absil, R. Mahony, and R. Sepulchre, *Riemannian geometry of Grassmann manifolds with a view on algorithmic computation*, Acta Appl. Math. **80** (2004), no. 2, 199–220.

[AMS08] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*, Princeton University Press, Princeton, NJ, January 2008.

[AMSV02] P.-A. Absil, R. Mahony, R. Sepulchre, and P. Van Dooren, *A Grassmann-Rayleigh quotient iteration for computing invariant subspaces*, SIAM Rev. **44** (2002), no. 1, 57–73.

[AS04] P.-A. Absil and R. Sepulchre, *Continuous dynamical systems that realize discrete optimization on the hypercube*, Systems Control Lett. **52** (2004), no. 3-4, 297–304.

[ASVM04] P.-A. Absil, R. Sepulchre, P. Van Dooren, and R. Mahony, *Cubically convergent iterations for invariant subspace computation*, SIAM J. Matrix Anal. Appl. **26** (2004), no. 1, 70–96.

[AV07] P.-A. Absil and P. Van Dooren, *Two-sided Grassmann-Rayleigh quotient iteration*, Tech. Report UCL-INMA-2007.024, Department of Mathematical Engineering, Université catholique de Louvain, Louvain-la-Neuve, Belgium, April 2007.

[BAG07] C. G. Baker, P.-A. Absil, and K. A. Gallivan, *Implicit trust-region methods on riemannian manifolds*, Tech. Report FSU-SCS-2007-449, School of Computational Science, Florida State University, http://www.scs.fsu.edu/publications/, June 2007.

[BHSF02] Asa Ben-Hur, Hava T. Siegelmann, and Shmuel Fishman, *A theory of complexity for continuous time systems*, J. Complexity **18** (2002), no. 1, 51–86.

[BI04] Anthony M. Bloch and Arieh Iserles, *On the optimality of double-bracket flows*, Int. J. Math. Math. Sci. **2004** (2004), no. 61-64, 3301–3319.

[BL89a] D. A. Bayer and J. C. Lagarias, *The nonlinear geometry of linear programming. I. Affine and projective scaling trajectories*, Trans. Amer. Math. Soc. **314** (1989), no. 2, 499–526.

[BL89b] ———, *The nonlinear geometry of linear programming. II. Legendre transform coordinates and central trajectories*, Trans. Amer. Math. Soc. **314** (1989), no. 2, 527–581.

[Boo75] William M. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London, 1975, Pure and Applied Mathematics, No. 63.

[Bro89] R. W. Brockett, *Least squares matching problems*, Linear Algebra and Applications **122-124** (1989), 761–777.

[Bro91] R. W. Brockett, *Dynamical systems that sort lists, diagonalize matrices, and solve linear programming problems*, Linear Algebra Appl. **146** (1991), 79–91.

[Cam04] Manuel Lameiras Campagnolo, *Continuous-time computation with restricted integration capabilities*, Theoret. Comput. Sci. **317** (2004), no. 1-3, 147–165.

[CD90] Moody T. Chu and Kenneth R. Driessel, *The projected gradient method for least squares matrix approximations with spectral constraints*, SIAM J. Numer. Anal. **27** (1990), no. 4, 1050–1060.

[CDLP05] M. Chu, N. Del Buono, L. Lopez, and T. Politi, *On the low-rank approximation of data on the unit sphere*, SIAM J. Matrix Anal. Appl. **27** (2005), no. 1, 46–60.

[CG02] Moody T. Chu and Gene H. Golub, *Structured inverse eigenvalue problems*, Acta Numer. **11** (2002), 1–71.

[CG03] Andrzej Cichocki and Pando Georgiev, *Blind source separation algorithms with matrix constraints*, IEICE Trans. Fundam. **E86-A** (2003), no. 1, 1–9.

[Chu88] Moody T. Chu, *On the continuous realization of iterative processes*, SIAM Rev. **30** (1988), no. 3, 375–387.

[Chu92] ———, *Matrix differential equations: a continuous realization process for linear algebra problems*, Nonlinear Anal. **18** (1992), no. 12, 1125–1146.

[Chu94] ———, *A list of matrix flows with applications*, Hamiltonian and gradient flows, algorithms and control, Fields Inst. Commun., vol. 3, Amer. Math. Soc., Providence, RI, 1994, pp. 87–97.

[CU92] A. Cichocki and R. Unbehauen, *Neural networks for computing eigenvalues and eigenvectors*, Biological Cybernetics **68** (1992), no. 2, 155–164.

[Deh95] Jeroen Dehaene, *Continuous-time matrix algorithms, systolic algorithms and adaptive neural networks*, Ph.D. thesis, 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/.

[DMV99] Jeroen Dehaene, Marc Moonen, and Joos Vandewalle, *Analysis of a class of continuous-time algorithms for principal component analysis and subspace tracking*, IEEE Trans. Circuits Systems I Fund. Theory Appl. **46** (1999), no. 3, 364–372.

[DNT83] P. Deift, T. Nanda, and C. Tomei, *Ordinary differential equations and the symmetric eigenvalue problem*, SIAM J. Numer. Anal. **20** (1983), no. 1, 1–22.

[Dou00] Scott C. Douglas, *Self-stabilized gradient algorithms for blind source separation with orthogonality constraints*, IEEE Trans. Neural Networks **11** (2000), no. 6, 1490–1497.

[EAS98] Alan Edelman, Tomás A. Arias, and Steven T. Smith, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl. **20** (1998), no. 2, 303–353.

[Fay91a] L. Faybusovich, *Dynamical systems which solve optimization problems with linear constraints*, IMA J. Math. Control Inform. **8** (1991), no. 2, 135–149.

[Fay91b] Leonid Faybusovich, *Hamiltonian structure of dynamical systems which solve linear programming problems*, Phys. D **53** (1991), no. 2-4, 217–232.

[FK05] K. R. Fowler and C. T. Kelley, *Pseudo-transient continuation for nonsmooth nonlinear equations*, SIAM J. Numer. Anal. **43** (2005), no. 4, 1385–1406 (electronic).

[Fla74] H. Flashka, *The Toda latice, II. Existence of integrals*, Physical Review B **9** (1974), no. 4, 1924–1925.

[GL06] Gene H. Golub and Li-Zhi Liao, *Continuous methods for extreme and interior eigenvalue problems*, Linear Algebra Appl. **415** (2006), no. 1, 31–51.

[GS01] F. Grognard and R. Sepulchre, *Global stability of a continuous-time flow which computes time-optimal switchings*, Proceedings of the 16th IEEE Conference on Decision and Control, 2001, pp. 3826–3831.

[Hig99] Desmond J. Higham, *Trust region algorithms and timestep selection*, SIAM J. Numer. Anal. **37** (1999), no. 1, 194–210 (electronic).

[HM94] Uwe Helmke and John B. Moore, *Optimization and dynamical systems*, Communications and Control Engineering Series, Springer-Verlag London Ltd., London, 1994, With a foreword by R. Brockett.

[Hop84] J. J. Hopfield, *Neurons with graded response have collective computational cababilities like those of two-state neurons*, Proc. Natl. Acad. Sci. USA **81** (1984), 3088–3092.

[HT85] J. J. Hopfield and D. W. Tank, *"Neural" computation of decision optimization problems*, Biol. Cybernet. **52** (1985), 141–152.

[Ips97] Ilse C. F. Ipsen, *Computing an eigenvector with inverse iteration*, SIAM Rev. **39** (1997), no. 2, 254–291.

[KC78] Harold J. Kushner and Dean S. Clark, *Stochastic approximation methods for constrained and unconstrained systems*, Applied Mathematical Sciences, vol. 26, Springer-Verlag, New York, 1978.

[KN63] Shoshichi Kobayashi and Katsumi Nomizu, *Foundations of differential geometry*, Interscience Publishers, a division of John Wiley & Sons, New York-London, 1963, Volumes 1 and 2.

[KQL+06] C. T. Kelley, L. Qi, L-Z. Liao, J. P. Reese, and C. Winton, *Pseudo-transient continuation and optimization*, http://www4.ncsu.edu/∼ctk/pubs.html, 2006.

[LE02] Eva Lundström and Lars Eldén, *Adaptive eigenvalue computations using Newton's method on the Grassmann manifold*, SIAM J. Matrix Anal. Appl. **23** (2001/02), no. 3, 819–839.

[Lju77] Lennart Ljung, *Analysis of recursive stochastic algorithms*, IEEE Trans. Automatic Control **AC-22** (1977), no. 4, 551–575.

[LKLT06] X.-L. Luo, C. T. Kelley, L.-Z. Liao, and H.-W. Tam, *Combining trust region techniques and Rosenbrock methods for gradient systems*, http://www4.ncsu.edu/∼ctk/pubs.html, 2006.

[LST98] Ralf Lösche, Hubert Schwetlick, and Gisela Timmermann, *A modified block Newton iteration for approximating an invariant subspace of a symmetric matrix*, Linear Algebra Appl. **275/276** (1998), 381–400.

[LW00] Xue-Bin Liang and Jun Wang, *A recurrent neural network for nonlinear optimization with a continuously differentiable objective function and bound constraints*, IEEE Trans. Neural Networks **11** (2000), no. 6, 1251–1262.

[MA03] R. Mahony and P.-A. Absil, *The continuous-time Rayleigh quotient flow on the sphere*, Linear Algebra Appl. **368C** (2003), 343–357.

[MC05] Jerzy Mycka and José Félix Costa, *The computational power of continuous dynamic systems*, Machines, Computations, and Universality, Lecture Notes in Computer Science, vol. 3354, Springer-Verlag, 2005, pp. 164–175.

[MH98] R. E. Mahony and U. Helmke, *System assignment and pole placement for symmetric realisations*, J. Math. Systems Estim. Control **8** (1998), no. 3, 321–352.

[MHM05] Jonathan H. Manton, Uwe Helmke, and Iven M. Y. Mareels, *A dual purpose principal and minor component flow*, Systems Control Lett. **54** (2005), no. 8, 759–769.

[Nan85] T. Nanda, *Differential equations and the QR algorithm*, SIAM J. Numer. Anal. **22** (1985), no. 2, 310–321.

[OK85] Erkki Oja and Juha Karhunen, *On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix*, J. Math. Anal. Appl. **106** (1985), no. 1, 69–84.

[Orp97] Pekka Orponen, *A survey of continuous-time computation theory*, Advances in Algorithms, Languages, and Complexity (D.-Z. Du and K.-I. Ko, eds.), Kluwer, Dordrecht, 1997, pp. 209–224.

[Par80] Beresford N. Parlett, *The symmetric eigenvalue problem*, Prentice-Hall Inc., Englewood Cliffs, N.J., 1980, Prentice-Hall Series in Computational Mathematics.

[Prz03] Maria Przybylska, *Isospectral-like flows and eigenvalue problem*, Future Generation Computer Syst. **19** (2003), 1165–1175.

[Sha41] Claude E. Shannon, *Mathematical theory of the differential analyzer*, J. Math. Phys. Mass. Inst. Tech. **20** (1941), 337–354.

[Smi97] Paul Smit, *Numerical analysis of eigenvalue algorithms based on subspace iterations*, Ph.D. thesis, CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, 1997.

[Ste73] G. W. Stewart, *Error and perturbation bounds for subspaces associated with certain eigenvalue problems*, SIAM Review **15** (1973), no. 4, 727–764.

[Sym82] W. W. Symes, *The QR algorithm and scattering for the finite nonperiodic Toda lattice*, Physica D **4** (1982), no. 2, 275–280.

[Vid95] M. Vidyasagar, *Minimum-seeking properties of analog neural networks with multilinear objective functions*, IEEE Trans. Automat. Control **40** (1995), no. 8, 1359–1375.

[Was04] Robert H. Wasserman, *Tensors and manifolds*, second ed., Oxford University Press, Oxford, 2004, With applications to physics.

[Wat82] David S. Watkins, *Understanding the* QR *algorithm*, SIAM Rev. **24** (1982), no. 4, 427–440.

[WE88] D. S. Watkins and L. Elsner, *Self-similar flows*, Linear algebra and its applications **110** (1988), 213–242.

[Wil65] J. H. Wilkinson, *The algebraic eigenvalue problem*, Clarendon Press, Oxford, 1965.

[YL99] Wei-Yong Yan and James Lam, *An approximate approach to H^2 optimal model reduction*, IEEE Trans. Automat. Control **44** (1999), no. 7, 1341–1358.