Available online at www.sciencedirect.com
SCIENCE DIRECT®
Systems & Control Letters 54 (2005) 759–769
SYSTEMS & CONTROL LETTERS
www.elsevier.com/locate/sysconle

A dual purpose principal and minor component flow

Jonathan H. Manton[a,*], Uwe Helmke[b], Iven M.Y. Mareels[a]

[a]Department of Electrical and Electronic Engineering, ARC Special Research Centre for Ultra-Broadband Information Networks, The University of Melbourne, Melbourne, Vic. 3010, Australia
[b]Department of Mathematics, University of Würzburg, Am Hubland, D-97074 Würzburg, Germany

Received 26 April 2003; received in revised form 21 October 2004; accepted 13 November 2004
Available online 21 January 2005

Abstract

Principal component flows are flows converging to the eigenvectors associated with the largest eigenvalues of a given symmetric matrix. Similarly, minor component flows converge to the eigenvectors associated with the smallest eigenvalues. Traditional flows required the matrix to be positive definite, and moreover, finding well-behaved minor component flows appeared to be harder and unrelated to the principal component case. This paper derives a flow which can be used to extract either the principal or the minor components and which does not require the matrix to be positive definite. The flow is shown to be a generalisation of the Oja–Brockett flow.

Keywords: Subspace flow; Subspace tracking; Principal components; Minor subspace; Oja flow

1. Introduction

There is considerable interest in constructing and analysing families of ordinary differential equations which are parameterised by a symmetric matrix $C \in \mathbb{R}^{n\times n}$, which evolve on matrix space $\mathbb{R}^{n\times p}$, and which have the following property. Starting from a generic initial matrix $X_0 \in \mathbb{R}^{n\times p}$, the flow converges to a matrix X_∞ whose columns are particular eigenvectors of C. If these eigenvectors are those associated with the p largest eigenvalues of C then the flow is said to be a principal component flow. Similarly, the flow is a minor component flow if it converges to the eigenvectors associated with the p smallest eigenvalues of C. An example of a principal component flow is

$$\dot{X} = CXN - XNX^{\mathrm{T}}CX, \tag{1}$$

where $\dot{X}$ denotes the derivative of $X \in \mathbb{R}^{n\times p}$ with respect to time, superscript T denotes matrix transpose, and N is an arbitrary diagonal matrix with distinct positive eigenvalues. This flow was introduced and partially studied in [17,18,27]. A detailed analysis appears in [32], which termed (1) the Oja–Brockett flow

* Corresponding author. Tel.: +61 383 446 791; fax: +61 383 446 678.

E-mail addresses: jon@ee.mu.oz.au (J.H. Manton), helmke@mathematik.uni-wuerzburg.de (U. Helmke).

0167-6911/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.sysconle.2004.11.015

because it is a natural generalisation of both Oja's flow [13,14] for principal subspace analysis and Brockett's flow [2] on orthogonal matrices for symmetric matrix diagonalisation.

The contribution of this paper is two-fold. The novel flow

$$\dot{X} = -CXN + \mu X(N - X^{\mathrm{T}}X) \tag{2}$$

is introduced and analysed. For appropriate choices of the constants $\mu \in \mathbb{R}$ and $N \in \mathbb{R}^{p\times p}$, it is proved to be a minor component flow. Moreover, and somewhat surprisingly, it is shown to be a generalisation of the Oja–Brockett principal component flow (1). Specifically, provided μ is sufficiently large, a coordinate transformation converting the Oja–Brockett flow (1) into the proposed minor component flow (2) is exhibited. It is remarked that whereas most flows, including (1), require C to be positive definite and symmetric, the proposed flow (2) only requires C to be symmetric.

The main interest in principal and minor component flows arises from being able to derive from them discrete time stochastic averaging algorithms for tracking the principal or minor components of a time-varying data matrix [1,9,11]. Subspace tracking algorithms are widely used in signal processing and control applications [5,8,12,20,21,25,29,31] such as direction or frequency estimation in antenna arrays [22,24,30], data compression via the Karhunen–Loéve transform [19] and multiuser detection in wireless communications [26].

In practice then, it is important to consider flows $\dot{X} = f(X, C)$, where the function $f(X, C)$ can be evaluated on a computer quickly. In particular, although a minor component flow is readily obtained from a principal component flow by replacing C with C^{-1}, it is not desirable to do so. Another trick sometimes resulting in a minor component flow is to reverse the sign of a principal component flow and incorporate a projection operator to prevent the flow from diverging. However, the computation of the projection operator requires a matrix inverse and hence is not desirable either.

With the proviso that $f(X, C)$ in the flow $\dot{X} = f(X, C)$ be an element-wise polynomial function, so that matrix inverses cannot appear, history shows that minor component flows evolving on matrix space are harder to construct than principal component flows. Moreover, there is no historical evidence that principal and minor component flows are somehow related. This makes the connection between the proposed minor component flow (2) and the existing principal component flow (1) interesting for two reasons; the very existence of a connection is itself interesting, and it is interesting that the minor component flow is more general than the principal component flow rather than the other way round.

Related work is now summarised. The starting point for most of the current work in principal component analysis and subspace tracking has been Oja's system from neural network theory [13–15]. Oja's principal subspace flow is

$$\dot{X} = (I - XX^{\mathrm{T}})CX, \tag{3}$$

where I denotes the identity matrix. A principal subspace flow differs from a principal component flow in that its stable equilibrium points X_∞ do not determine the principal eigenvectors individually, but instead, the space spanned by the columns of X_∞ is the same as the space spanned by the principal eigenvectors of C. In fact, although (3) was proved in [13,16] to be a principal subspace flow if $p=1$ (recall p is the number of columns of X), the conjecture that it is a principal subspace flow for $p>1$ was not proved till much later in [28].

An interesting feature of the Oja flow (3) is that it has been shown very recently to be a gradient flow [32] with respect to a suitable Riemannian metric on $\mathbb{R}^{n\times p}$. This is despite confusing remarks made earlier in the literature claiming that it cannot be a gradient flow because the linearisation is not a symmetric matrix. This claim is only valid for gradient flows with respect to the Euclidean metric.

Principal component flows were first studied in [23,17,18,27]. However, pointwise convergence to the equilibria points was not established in these papers. Using an early result by Lojasiewicz [10] on real analytic gradient flows, the pointwise convergence of the Oja–Brockett flow (1) was established in [32]. It is also mentioned that although sufficient conditions for initial matrices in the Oja flow (3) to converge to a principal subspace are given in [4,28], a complete characterisation of the stable and unstable manifolds is currently lacking for flows (1) and (3).

The remainder of this paper is organised as follows. Section 2 motivates the introduction of a novel cost function whose critical points are related to the eigenstructure of a matrix C. Under mild assumptions, all the critical points are characterised. The proposed flow (2) is a gradient flow for this cost function. Section 3 formally states the definition of a minor component flow and proves that (2) does indeed meet the requirements of this definition. Informal arguments and several numerical examples in Section 4 provide a feel for the convergence rate of the flow. Section 5 shows that replacing C with $-C$ converts (2) into a principal component flow. It also establishes a connection between the proposed flow (2) and the Oja–Brockett flow (1). Section 6 concludes the paper.

2. A minor component cost function

This section introduces and analyses a cost function necessary for the understanding of the proposed minor component flow (2).

2.1. Motivation

Subject to the orthogonality constraint $X^{\mathrm{T}}X = I$, it is a standard result [6] that the generalised Rayleigh quotient $\mathrm{tr}\{CXNX^{\mathrm{T}}\}$ takes its smallest value when the columns of the tall matrix X are the eigenvectors (arranged in a suitable order) corresponding to the smallest eigenvalues of the symmetric matrix C, hereafter referred to as the minor components of C. Here, N is an arbitrary diagonal matrix with distinct positive elements and $\mathrm{tr}\{\cdot\}$ is the trace operator. With this in mind, define the penalised cost function

$$\tilde{f}(X) = \mathrm{tr}\{CXNX^{\mathrm{T}}\} + \mu\|I - X^{\mathrm{T}}X\|^2, \tag{4}$$

where I is the identity matrix and $\|\cdot\|$ the Frobenius norm. From above, it follows that in the limit $\mu \to \infty$, the minimum of $\tilde{f}(X)$ occurs when X contains the minor components of C. This paper is based on the novel observation that even for finite but sufficiently large μ, the minimum of $\tilde{f}(X)$ still occurs when X contains the minor components of C.

It is remarked that (4) is a generalisation of the cost function studied in [3], the latter corresponding to the special case of X being a vector. Algorithms for subspace tracking using this special case of the cost function were developed in [8,12].

2.2. The cost function and assumptions

In Appendix A, it is shown that the cost function (4) can have critical points unrelated to the eigenstructure of C. To avoid this, the penalty term is modified thus

$$f(X) = \tfrac{1}{2}\,\mathrm{tr}\{CXNX^{\mathrm{T}}\} + \tfrac{1}{4}\mu\|N - X^{\mathrm{T}}X\|^2. \tag{5}$$

The following assumptions are made throughout.

- A1. The scalar $\mu \in \mathbb{R}$ is strictly positive.
- A2. The matrix $C \in \mathbb{R}^{n\times n}$ is symmetric.
- A3. The matrix $N \in \mathbb{R}^{p\times p}$ is diagonal with distinct positive eigenvalues.

When discussing local minima, it is convenient to make the following additional assumptions.

- A4. The matrix C has distinct eigenvalues.
- A5. The scalar μ does not equal any eigenvalue of C.

2.3. Critical points

The directional derivative of (5) in the direction $\xi \in \mathbb{R}^{n\times p}$ is readily calculated to be

$$\mathscr{D}f(X)\xi = \mathrm{tr}\{\xi^{\mathrm{T}}CXN + \mu\xi^{\mathrm{T}}(-XN + XX^{\mathrm{T}}X)\}. \tag{6}$$

Therefore, the critical points of $f(X)$ are the points X satisfying

$$CXN = \mu X(N - X^{\mathrm{T}}X). \tag{7}$$

The solutions of (7) are stated explicitly in Proposition 3, the proof of which requires the following two lemmas.

Lemma 1. *If matrices A and N commute and N is diagonal with distinct eigenvalues then A is diagonal.*

Proof. If $N = \mathrm{diag}\{\lambda_1, \ldots, \lambda_p\}$ then $AN - NA = 0$ implies the ijth element of A satisfies $(\lambda_i - \lambda_j)A_{ij} = 0$. □

Lemma 2. *Assume* A1–A3 *hold. A necessary condition for X to satisfy* (7) *is that it can be written in the form $X = QD$ where D is diagonal and Q is isometric* ($Q^{\mathrm{T}}Q = I$).

Proof. Pre-multiplication of (7) by X^T shows that $X^T(\mu I - C)XN = \mu(X^TX)^2$. Since the right-hand side is symmetric, $X^T(\mu I - C)X$ commutes with N and is diagonal by Lemma 1. Thus $(X^TX)^2$, and hence X^TX, is diagonal. (Recall $\mu \neq 0$ by A1.) This implies X is of the form $X = QD$ with $Q^TQ = I$ and D diagonal. □

Proposition 3 (*Critical points*). *Assume* A1–A3 *hold. Let N_{ii} denote the ith diagonal element of N. A necessary and sufficient condition for $X = [x_1, \ldots, x_p]$ to be a critical point of* (5) *is that, for all i, either*

(1) *x_i is the null vector, or*
(2) *x_i is an eigenvector of C with corresponding eigenvalue λ_i and $\|x_i\|^2 = N_{ii} - (\lambda_i/\mu)N_{ii}$.*

Proof. From Lemma 2, write $X = QD$ where D is diagonal and Q is isometric. Substituting into (7) yields $CQDN = \mu QD(N - D^2)$. Let q_i denote the ith column of Q. Then $Cq_i D_{ii} N_{ii} = \mu q_i D_{ii}(N_{ii} - D_{ii}^2)$. Thus, either $D_{ii} = 0$, in which case x_i is the null vector, or $Cq_i = \mu N_{ii}^{-1}(N_{ii} - D_{ii}^2)q_i$, implying q_i, and hence x_i, is an eigenvector of C with associated eigenvalue $\lambda_i = \mu N_{ii}^{-1}(N_{ii} - D_{ii}^2)$. The proposition now follows by noting $\|x_i\|^2 = D_{ii}^2$. □

Proposition 3 shows that the only effect decreasing μ has on the location of the critical points is to cause the columns of each critical point to shrink towards and ultimately equal the null vector.

2.4. Local stability analysis of critical points

An immediate consequence of Proposition 3 is that if X is a critical point of (5) then there exists an orthogonal matrix Q and a diagonal matrix D such that $X = Q[D\ 0]^T$ and the columns of Q are the eigenvectors of C, that is, $\Lambda = Q^TCQ$ is diagonal. Moreover, each real valued diagonal element D_{ii} can take at most three values; either $D_{ii} = 0$ or $D_{ii}^2 = N_{ii} - \mu^{-1}\Lambda_{ii}N_{ii}$. This representation is used throughout this section.

The following lemma expresses the Hessian of (5) about any critical point in block diagonal form, with each block either 1×1 or 2×2. This enables the type of critical point, such as a local minimum or saddle point, to be determined by inspection.

Lemma 4 (*Hessian*). *Assume* A1–A3 *hold. Let X be a critical point of the cost function $f(X)$ defined in* (5). *Let the orthogonal matrix Q and diagonal matrix D be such that $\Lambda = Q^TCQ$ is diagonal and $X = Q[D\ 0]^T$; such Q and D always exist. Then, the second directional derivative $g(\xi) = \mathscr{D}^2 f(X)\xi$ of $f(X)$ in the direction ξ at the point X is given by the quadratic form*

$$g(Q\xi) = \sum_{i=1}^{p} \alpha_i \xi_{ii}^2 + \sum_{i=p+1}^{n} \sum_{j=1}^{p} \beta_{ij}\xi_{ij}^2 + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} [\xi_{ij}\ \ \xi_{ji}]\Gamma_{(ij)}[\xi_{ij}\ \ \xi_{ji}]^T, \tag{8}$$

where the scalars $\alpha_i, \beta_{ij} \in \mathbb{R}$ and matrices $\Gamma_{(ij)} \in \mathbb{R}^{2\times 2}$ are

$$\alpha_i = N_{ii}\Lambda_{ii} - \mu(N_{ii} - 3D_{ii}^2), \tag{9}$$

$$\beta_{ij} = N_{jj}\Lambda_{ii} - \mu(N_{jj} - D_{jj}^2), \tag{10}$$

$$\Gamma_{(ij)} = \begin{bmatrix} N_{jj}\Lambda_{ii} - \mu(N_{jj} - D_{jj}^2 - D_{ii}^2) & \mu D_{ii}D_{jj} \\ \mu D_{ii}D_{jj} & N_{ii}\Lambda_{jj} - \mu(N_{ii} - D_{ii}^2 - D_{jj}^2) \end{bmatrix}. \tag{11}$$

Proof. Differentiating (6) yields

$$\mathscr{D}^2 f(X)\xi = \mathrm{tr}\{\xi^T C\xi N + \mu\,\xi^T(-\xi N + \xi X^T X + X\xi^T X + XX^T\xi)\}. \tag{12}$$

For a given critical point X, define $g(\xi)$ to be this quadratic form. By Proposition 3, there exists an orthogonal matrix Q and a diagonal matrix D such that $\Lambda = Q^TCQ$ is diagonal and $X = Q[D\ 0]^T$. Thus

$$g(Q\xi) = \mathrm{tr}\{\xi^T\Lambda\xi N - \mu\xi^T\xi(N - D^2) + \mu\xi^T[D\ 0]^T\xi^T[D\ 0]^T + \mu\xi^T[D\ 0]^T[D\ 0]\xi\} \tag{13}$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{p}(N_{jj}\Lambda_{ii} - \mu(N_{jj} - D_{jj}^2))\xi_{ij}^2 + \mu\sum_{i=1}^{p}\sum_{j=1}^{p}(D_{ii}^2\xi_{ij}^2 + D_{ii}D_{jj}\xi_{ij}\xi_{ji}) \tag{14}$$

$$= \sum_{i=1}^{p} \alpha_i \xi_{ii}^2 + \sum_{i=p+1}^{n} \sum_{j=1}^{p} \beta_{ij} \xi_{ij}^2 + \sum_{i=1}^{p-1} \sum_{j=i+1}^{p} [\xi_{ij}\ \xi_{ji}] \Gamma_{(ij)} [\xi_{ij}\ \xi_{ji}]^{\mathrm{T}}, \quad (15)$$

where α_i, β_{ij} and $\Gamma_{(ij)}$ are as given in the lemma. □

Of most interest are the local minima of $f(X)$, defined to be the critical points for which the quadratic form $g(\xi)$ in Lemma 4 is positive definite. Referring to (8), the local minima are the critical points for which $\alpha_i > 0$, $\beta_{ij} > 0$ and $\Gamma_{(ij)} > 0$.

The following proposition proves that under A1–A5, $f(X)$ has a unique minimum, up to the sign of each column, given by the minor components of C arranged in an order governed by N, unless μ is too small, in which case some of the columns will be zero instead.

Proposition 5 (*Local minima*). *Assume* A1–A5 *hold. Define* $f(X)$ *as in* (5). *Let* σ *be the permutation of* $\{1,\ldots,p\}$ *such that* $N_{\sigma(1)\sigma(1)} > \cdots > N_{\sigma(p)\sigma(p)}$. *Let* $\lambda_1 < \cdots < \lambda_n$ *be the eigenvalues of* C *in ascending order and let* $v_1,\ldots,v_n$ *be the corresponding unit norm eigenvectors*. *Then* $X = [x_1,\ldots,x_p]$ *is a local minimum of* $f(X)$ *if and only if*

$$x_{\sigma(i)} = \pm\gamma_i v_i, \quad \gamma_i = \begin{cases} \sqrt{N_{\sigma(i)\sigma(i)}(1-\mu^{-1}\lambda_i)}, & \lambda_i < \mu, \\ 0 & \textit{otherwise} \end{cases} \quad (16)$$

for $i = 1,\ldots,p$.

Proof. Let Π denote the unique permutation matrix arranging the diagonal elements of $\Pi^{\mathrm{T}} N \Pi$ in descending order. The ith diagonal element of $\Pi^{\mathrm{T}} N \Pi$ is thus $N_{\sigma(i)\sigma(i)}$. Since $\Pi^{\mathrm{T}}\Pi = I$, substitution shows that the cost function (5) satisfies the symmetry $f(X;N) = f(X\Pi;\Pi^{\mathrm{T}} N \Pi)$. By observing that the ith column of $X\Pi$ is $x_{\sigma(i)}$, it becomes clear from (16) that it suffices to prove the proposition for the special case when the diagonal elements of N are already in descending order.

Henceforth, assume $i < j$ implies $N_{ii} > N_{jj}$. (Thus $\sigma(i) = i$.) Let X be an arbitrary critical point of $f(X)$. Let Q and D be as in Lemma 4 so that $\Lambda = Q^{\mathrm{T}} C Q$ is diagonal and $X = Q[D\ 0]^{\mathrm{T}}$. From Proposition 3, for each i there are at most two possibilities for D_{ii}^2; either $D_{ii}^2 = 0$ or, provided $\mu > \Lambda_{ii}$, $D_{ii}^2 = N_{ii}(1-\mu^{-1}\Lambda_{ii})$. With this in mind, consider α_i in (9) for $i = 1,\ldots,p$. If $D_{ii} = 0$ then direct substitution shows that $\alpha_i > 0$ if and only if $\Lambda_{ii} > \mu$. Similarly, if $D_{ii}^2 = N_{ii}(1-\mu^{-1}\Lambda_{ii})$ then $\alpha_i > 0$ if and only if $\Lambda_{ii} < \mu$. Next, consider β_{ij} in (10) for $i = p+1,\ldots,n$ and $j = 1,\ldots,p$. If $D_{jj} = 0$ then $\beta_{ij} > 0$ if and only if $\Lambda_{ii} > \mu$. If $D_{jj}^2 = N_{jj}(1-\mu^{-1}\Lambda_{jj})$ then $\beta_{ij} > 0$ if and only if $\Lambda_{ii} > \Lambda_{jj}$. Lastly, consider $\Gamma_{(ij)}$ in (11) for $1 \leqslant i < j \leqslant p$. There are four cases. If $D_{ii} = D_{jj} = 0$ then $\Gamma_{(ij)} > 0$ if and only if both $\Lambda_{ii} > \mu$ and $\Lambda_{jj} > \mu$. If $D_{ii} = 0$ but $D_{jj}^2 = N_{jj}(1-\mu^{-1}\Lambda_{jj})$ then $\Gamma_{(ij)} > 0$ if and only if both $\Lambda_{ii} > \Lambda_{jj}$ and $(N_{ii}-N_{jj})(\Lambda_{jj}-\mu) > 0$. However, the last condition is always false: Because $i < j$ implies $N_{ii} > N_{jj}$ while $D_{jj}^2 > 0$ implies $\mu > \Lambda_{jj}$, $(N_{ii}-N_{jj})(\Lambda_{jj}-\mu)$ is always negative. If $D_{ii}^2 = N_{ii}(1-\mu^{-1}\Lambda_{ii})$ but $D_{jj} = 0$ then $\Gamma_{(ij)} > 0$ if and only if both $\mu > \Lambda_{ii}$ and $\Lambda_{jj} > \Lambda_{ii}$. The case $D_{ii}^2 = N_{ii}(1-\mu^{-1}\Lambda_{ii})$ and $D_{jj}^2 = N_{jj}(1-\mu^{-1}\Lambda_{jj})$ is slightly more involved. Substitution into (11) yields

$$\Gamma_{(ij)} = \begin{bmatrix} N_{jj}(\Lambda_{ii}-\Lambda_{jj})+\mu D_{ii}^2 & \mu D_{ii} D_{jj} \\ \mu D_{ii} D_{jj} & N_{ii}(\Lambda_{jj}-\Lambda_{ii})+\mu D_{jj}^2 \end{bmatrix}. \quad (17)$$

A 2×2 matrix is positive definite if and only if its trace and determinant are both positive. The trace and determinant of $\Gamma_{(ij)}$ are

$$\operatorname{tr}\{\Gamma_{(ij)}\} = (N_{ii}-N_{jj})(\Lambda_{jj}-\Lambda_{ii}) + \mu D_{ii}^2 + \mu D_{jj}^2, \quad (18)$$

$$|\Gamma_{(ij)}| = (\Lambda_{jj}-\Lambda_{ii})(N_{ii}N_{jj}\Lambda_{ii} - N_{ii}N_{jj}\Lambda_{jj} + \mu N_{ii} D_{ii}^2 - \mu N_{jj} D_{jj}^2) \quad (19)$$

$$= (N_{ii}-N_{jj})(\Lambda_{jj}-\Lambda_{ii})(N_{ii}(\mu-\Lambda_{ii}) + N_{jj}(\mu-\Lambda_{jj})). \quad (20)$$

The term $N_{ii}(\mu-\Lambda_{ii}) + N_{jj}(\mu-\Lambda_{jj})$ is always positive because $D_{ii}^2 > 0$ and $D_{jj}^2 > 0$ imply $\mu > \Lambda_{ii}$ and $\mu > \Lambda_{jj}$. Similarly, $N_{ii}-N_{jj} > 0$ because $i < j$. Therefore $|\Gamma_{(ij)}| > 0$ if and only if $\Lambda_{jj} > \Lambda_{ii}$. Furthermore, $\Lambda_{jj} > \Lambda_{ii}$ implies $\operatorname{tr}\{\Gamma_{(ij)}\} > 0$. Thus, if $D_{ii}^2 = N_{ii}(1-\mu^{-1}\Lambda_{ii})$ and $D_{jj}^2 = N_{jj}(1-\mu^{-1}\Lambda_{jj})$ then $\Gamma_{(ij)} > 0$ if and only if $\Lambda_{jj} > \Lambda_{ii}$.

To prove one direction, assume X is a local minimum and define Q, D and Λ as above. (Observe that the ith column of Q is an eigenvector of C with associated eigenvalue Λ_{ii}.) First, assume no diagonal element of D is zero, so that $D_{ii}^2 = N_{ii}(1-\mu^{-1}\Lambda_{ii}) > 0$. This necessitates $\Lambda_{ii} < \mu$ for $i = 1, \ldots, p$. The condition $\Gamma_{(ij)} > 0$ then forces $\Lambda_{ii} < \Lambda_{jj}$ for $1 \leqslant i < j \leqslant p$. Finally, the condition $\beta_{ij} > 0$ forces $\Lambda_{ii} > \Lambda_{pp}$ for $i = p+1, \ldots, n$. That is, X must be given by (16) in this case. (Recall $\sigma(i) = i$ by assumption.) Now, assume instead that a diagonal element of D is zero. Let k be the smallest integer such that $D_{kk} = 0$. Then, since $D_{ii}^2 = N_{ii}(1-\mu^{-1}\Lambda_{ii}) > 0$ for $i = 1, \ldots, k-1$, $\Lambda_{ii} < \mu$ for $i = 1, \ldots, k-1$. Since $D_{kk} = 0$, the condition $\Gamma_{(ij)} > 0$ implies $D_{jj} = 0$ for $j = k+1, \ldots, p$. It also implies $\Lambda_{ii} > \mu$ for $i = k, \ldots, p$. Finally, the condition $\beta_{ij} > 0$ implies $\Lambda_{ii} > \mu$ for $i = p+1, \ldots, n$. Again, X must be given by (16), completing the proof in one direction.

To prove the other direction, define X as in (16). (Recall $\sigma(i) = i$ by assumption.) It follows from Proposition 3 that X is a critical point. Let $Q = [v_1, \ldots, v_n]$ and $D = \operatorname{diag}\{\gamma_1, \ldots, \gamma_p\}$, so that $\Lambda = Q^{\mathrm{T}} C Q = \operatorname{diag}\{\lambda_1, \ldots, \lambda_n\}$ is diagonal and $X = Q[D\ 0]^{\mathrm{T}}$. It follows that $\alpha_i > 0$ because either $\lambda_i > \mu$ in which case $D_{ii} = \gamma_i = 0$ and thus $\alpha_i > 0$, or $\lambda_i < \mu$ in which case $D_{ii}^2 = \gamma_i^2 = N_{ii}(1-\mu^{-1}\Lambda_{ii})$ and thus $\alpha_i > 0$. (Note that A6 excludes the case $\lambda_i = \mu$.) Similarly, it can be shown that $\beta_{ij} > 0$ and $\Gamma_{(ij)} > 0$, proving that (16) is a local minimum. □

Propositions 3 and 5 show that to find the p minor components of C it is necessary and sufficient to choose $\mu > \lambda_p$, since then the essentially unique local minimum of (5) corresponds to the minor components of C. Interestingly, if μ is also chosen to be less than λ_{p+1} then all the critical points associated with the principal eigenvectors, which are nevertheless unstable, are eliminated.

The eigenvalues of the Hessian about a critical point influence the asymptotic convergence rate of an optimisation algorithm to that critical point and are thus of interest.

Proposition 6 (*Eigenvalues of Hessian*). *Assume* A1–A3 *hold. Let* $X = [x_1, \ldots, x_p] \in \mathbb{R}^{n\times p}$ *be a critical point of* $f(X)$ *defined in* (5). *Assume*[1] *that no column of* X *is the null vector. Then each* x_i *is an eigenvector of* C; *let* λ_i *denote its associated eigenvalue. The np eigenvalues of the Hessian of* $f(X)$ *about the critical point* X *are*

$$\begin{aligned}&\{2N_{ii}(\mu-\lambda_i), i = 1, \ldots, p\}\\ &\quad \cup \{N_{ii}(\lambda_j - \lambda_i), i = 1, \ldots, p, j = p+1, \ldots, n\}\\ &\quad \cup \{(N_{ii} - N_{jj})(\lambda_j - \lambda_i), 1 \leqslant i < j \leqslant p\}\\ &\quad \cup \{N_{ii}(\mu - \lambda_i) + N_{jj}(\mu - \lambda_j), 1 \leqslant i < j \leqslant p\}.\end{aligned} \tag{21}$$

Proof. That each x_i is an eigenvector follows from Proposition 3. Lemma 4 implies that the eigenvalues of the Hessian are simply α_i, β_{ij} and the eigenvalues of $\Gamma_{(ij)}$ (with $\Lambda_{ii} = \lambda_i$). From Proposition 3, if $D_{ii} \neq 0$ then it must satisfy $D_{ii}^2 = N_{ii}(1-\mu^{-1}\lambda_i)$. Substituting D_{ii}^2 into α_i and β_{ij} yields $\alpha_i = 2N_{ii}(\mu - \lambda_i)$ and $\beta_{ij} = N_{jj}(\lambda_i - \lambda_j)$. (Note that i and j have been interchanged in the term corresponding to β_{ij} in (21).) Substitution of D_{ii} into $\Gamma_{(ij)}$ and taking its trace and determinant yields (18) and (20), where $\Lambda_{ii} = \lambda_i$ and $\Lambda_{jj} = \lambda_j$. Since $\mu D_{ii}^2 + \mu D_{jj}^2 = N_{ii}(\mu - \lambda_i) + N_{jj}(\mu - \lambda_j)$, it follows from (18) and (20) that the eigenvalues of $\Gamma_{(ij)}$ are $(N_{ii} - N_{jj})(\lambda_j - \lambda_i)$ and $N_{ii}(\mu - \lambda_i) + N_{jj}(\mu - \lambda_j)$. This completes the proof. □

2.5. Compact sub-level sets

The following lemma establishes that (5) is lower bounded. Thus, its essentially unique local minimum (16) is also its global minimum. The lemma also proves the technical condition that (5) has compact sub-level sets.

Lemma 7. *Assume* A1 *holds. The cost function* (5) *is lower bounded and, for any constant* $c \in \mathbb{R}$, *its sub-level set* $\{X : f(X) \leqslant c\}$ *is compact.*

Proof. Function (5) can be written as

$$\begin{aligned}f(X) = \tfrac{1}{4}\operatorname{tr}\{&\mu N^2 + 2X^{\mathrm{T}}CXN - 2\mu X^{\mathrm{T}}XN\\ &+ \mu X^{\mathrm{T}}XX^{\mathrm{T}}X\}\end{aligned} \tag{22}$$

[1] It is straightforward to modify the proof so as to find the eigenvalues of the Hessian if one or more columns of X are zero. Indeed, if either $D_{ii} = 0$ or $D_{jj} = 0$ then $\Gamma_{(ij)}$ reduces to a diagonal matrix.

from which it is clear that $f(X)$ is a fourth degree polynomial in the elements of X. For large X, the dominant term is $(\mu/4)\,\mathrm{tr}\{X^{\mathrm{T}}XX^{\mathrm{T}}X\}$ which is lower bounded by zero and has compact sub-level sets. The lemma now follows. □

3. The minor component flow

Proposition 5 established that the essentially unique local minimum of the cost function (5) corresponds to the minor components of C. It is therefore natural to consider the corresponding gradient flow, which the following lemma shows is precisely (2). For an introduction to gradient flows, see [7].

Lemma 8 (*Gradient flow*). *Define $f(X)$ as in* (5). *Then* (2) *is the negative gradient flow*

$$\dot{X} = -\mathrm{grad}\, f(X), \tag{23}$$

where the gradient is with respect to the Euclidean inner product $\langle A, B\rangle = \mathrm{tr}\{B^{\mathrm{T}}A\}$ on matrix space.

Proof. It follows from (6) that

$$\mathscr{D}f(X)\xi = \langle CXN - \mu X(N - X^{\mathrm{T}}X), \xi\rangle. \tag{24}$$

Thus $\mathrm{grad}\, f(X) = CXN - \mu X(N - X^{\mathrm{T}}X)$, proving the lemma. □

The main result of this paper is that (2) is a minor component flow. In order to state this rigorously, the definition of a minor component flow is first given.

Definition 9 (*Minor component flow*). The flow $\dot{X} = f(X, C)$ on $\mathbb{R}^{n\times p}$ is a minor component flow for matrices C belonging to the class $\mathscr{C} \subset \mathbb{R}^{n\times n}$ if the following hold:

(1) For any initial condition X_0 and any $C \in \mathscr{C}$, a solution $X(t)$ of the flow $\dot{X} = f(X, C)$ satisfying $X(0) = X_0$ exists and is unique for all $t \geqslant 0$.
(2) The limit $X_\infty = \lim_{t\to\infty} X(t)$ always exists.
(3) If X_0 is a generic initial condition, then the p columns of X_∞ are orthogonal to each other and each one is an eigenvector associated with one of the p smallest eigenvalues of C, counting multiplicities.

Theorem 10 (*Minor component flow*). *Let n and p be arbitrary integers with $n \geqslant p \geqslant 1$. For any real valued $\mu > 0$, define $\mathscr{C}_\mu$ to be the set of all symmetric matrices in $\mathbb{R}^{n\times n}$ whose eigenvalues are distinct, are not equal to μ, and at least p of them are less than μ. Then, for any diagonal matrix $N \in \mathbb{R}^{p\times p}$ with distinct positive eigenvalues, the flow* (2) *is a minor component flow for $C \in \mathscr{C}_\mu$.*

The proof of Theorem 10 relies on the following standard result.

Lemma 11. *Consider the negative gradient flow $\dot{X} = -\mathrm{grad}\, f(X)$ and let X_0 be an arbitrary initial condition. If $f(X)$ is a smooth, lower bounded function with compact sub-level sets, then there exists a unique trajectory $X(t)$ defined and bounded for all $t \geqslant 0$ and satisfying $X(0) = X_0$. Moreover, if the critical points of $f(X)$ are isolated, and if the Hessian of $f(X)$ at every critical point is non-singular, then $X_\infty = \lim_{t\to\infty} X(t)$ exists and is a critical point of $f(X)$, and if X_0 is chosen generically then X_∞ is a local minimum of $f(X)$.*

Proof of Theorem 10. Lemma 8 established that (2) is the negative gradient flow of the cost function $f(X)$ defined in (5). Lemma 7 proved $f(X)$ is lower bounded and has compact sub-level sets. Being a fourth degree polynomial in the elements of X, $f(X)$ is smooth. If $C \in \mathscr{C}_\mu$, $\mu > 0$ and N is diagonal with distinct positive eigenvalues, then A1–A5 hold. Thus, Proposition 3 implies that the critical points of $f(X)$ are isolated. Inspection of (21) reveals that the eigenvalues of the Hessian about any critical point are all non-zero because A3–A5 hold. Proposition 5 shows that the only local minima are those corresponding to the minor components of C. Therefore, Lemma 11 implies that all three requirements in Definition 9 are satisfied. □

4. Numerical examples and convergence rates

The results of this section are informal, the intention being merely to give a feel for how μ, N and C influence the convergence rate of the proposed minor component flow (2).

It is assumed for this section only that

- $C = \mathrm{diag}\{\lambda_1, \ldots, \lambda_n\}$, where $\lambda_1 < \cdots < \lambda_n$,

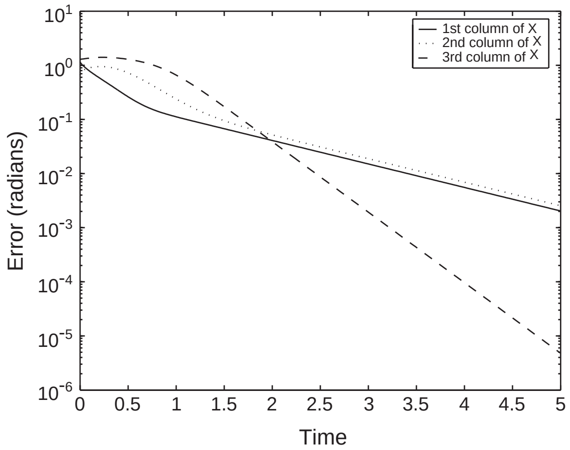

Fig. 1. Graph of the evolution of the errors $\theta_1(t), \ldots, \theta_3(t)$, defined in (25), of the flow (2) with $\mu=6$, $C=\mathrm{diag}\{\ \}1, 2, 5, 8, 11$ and $N=\mathrm{diag}\{\ \}3, 2, 1$. Initial point was chosen randomly.

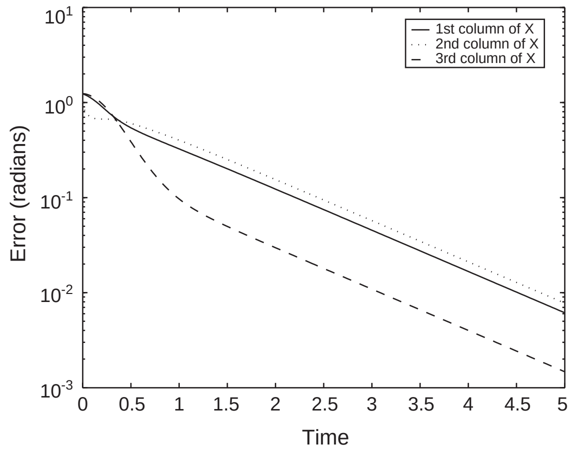

Fig. 2. Graph of the evolution of the errors $\theta_1(t), \ldots, \theta_3(t)$, defined in (25), of the flow (2) with $\mu=6$, $C=\mathrm{diag}\{\ \}1, 2, 3, 8, 11$ and $N=\mathrm{diag}\{\ \}3, 2, 1$. Initial point was chosen randomly.

- $N=\mathrm{diag}\{N_{11}, \ldots, N_{pp}\}$, where $N_{11} > \cdots > N_{pp}$,
- $\lambda_p < \mu$.

This causes no loss of generality because replacing X with $UX\Pi$ leaves the form of (2) unchanged if U is an orthogonal matrix and Π a permutation matrix.

Under these assumptions, Theorem 10 (see also Proposition 5) shows that the essentially unique stable equilibrium point of (2) is $X=[I\ \ 0]^{\mathrm{T}}$. Let v_i denote the ith column of this stable equilibrium point, so that if $X(t)$ is the value of X in the flow (2) at time t then it can be assumed that, by appropriate choice of the initial condition $X(0)$, the ith column $x_i(t)$ of $X(t)$ converges to v_i. The distance $x_i(t)$ is from its limit v_i can be measured by the angle between them, given by

$$\theta_i(t) = \arccos \frac{x_i(t) \cdot v_i}{\|x_i(t)\|\, \|v_i\|}, \tag{25}$$

where $\cdot$ denotes vector dot product.

The eigenvalues of the linearised flow about the stable equilibrium point are precisely the eigenvalues of the Hessian about the equilibrium point, and are given in (21). Since the Hessian is positive definite about the stable equilibrium point, $\theta_i(t)$ converges to zero asymptotically like $\mathrm{e}^{-\lambda t}$ where λ is related to the eigenvalues of the Hessian and depends on i in general. Figs. 1–3 bear testament to this; the log of the error decreases linearly with time.

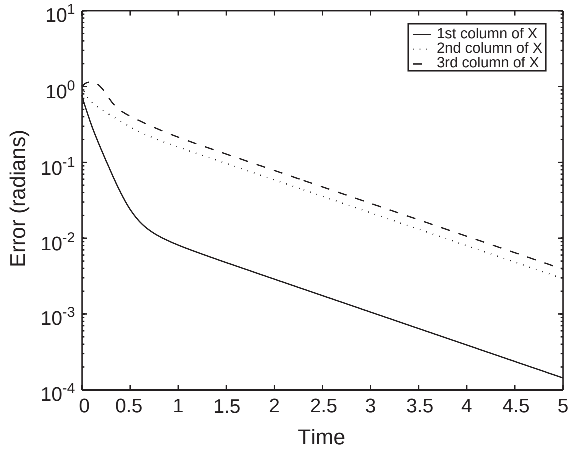

Fig. 3. Same as in Fig. 2 except a different initial point was chosen randomly.

Define the largest error to be $\max_i \theta_i(t)$. Its rate of decrease is dominated by the smallest eigenvalue in the set (21). Clearly, since $N_{11} > \cdots > N_{pp}$ and $\lambda_1 < \cdots < \lambda_p$, the smallest eigenvalue in (21) must belong to the set

$$\begin{aligned}&\{2N_{pp}(\mu-\lambda_p), N_{pp}(\lambda_{p+1}-\lambda_p),\\ &\quad (N_{11}-N_{22})(\lambda_2-\lambda_1), \ldots,\\ &\quad (N_{p-1,p-1}-N_{pp})(\lambda_p-\lambda_{p-1})\}.\end{aligned} \tag{26}$$

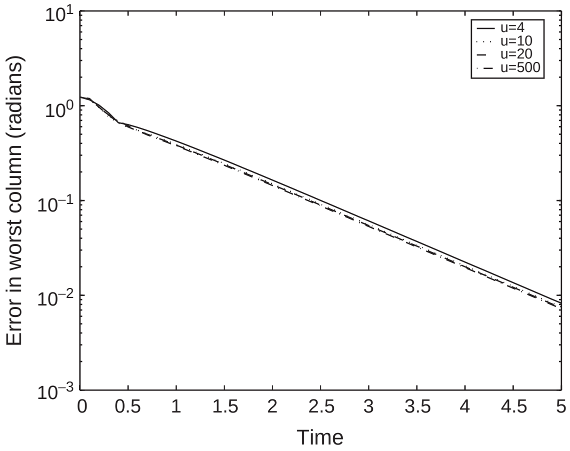

Fig. 4. Graph demonstrating the relative insensitivity of the flow to changes in the parameter μ. Here, $C = \mathrm{diag}\{\}1, 2, 3, 8, 11$ and $N = \mathrm{diag}\{\}3, 2, 1$.

Usually, such as if $\mu \geqslant (\lambda_p + \lambda_{p+1})/2$, the minimum of (26) does not depend on μ. Thus, it can be expected that the largest error decreases at a rate relatively insensitive to μ. Fig. 4 supports this argument. When choosing μ though, recall from Proposition 3 that one mild advantage of choosing μ to lie between λ_p and λ_{p+1} is that the number of unstable critical points is reduced.

Doubling N causes all the eigenvalues in (21) to be doubled. Therefore, the asymptotic convergence rate increases by a factor of two (that is, the rate $\mathrm{e}^{-\lambda t}$ becomes $\mathrm{e}^{-2\lambda t}$).

Observe from (26) that if λ_i is close to λ_{i-1} then $N_{i-1,i-1} - N_{ii}$ must be large if the error is to converge to zero at a reasonable rate. If the eigenvalue distribution of C is unknown, as is usually the case, then a sensible choice for N is one such that the differences $N_{i-1,i-1} - N_{ii}$ are equal for all i; for instance, $N = \mathrm{diag}\{p, p-1, \ldots, 1\}$ is suitable.

Lastly, the possibility of replacing C with $C + \lambda I$ to improve the convergence rate is ruled out. If C is replaced by $C + \lambda I$ and μ is replaced by $\mu + \lambda$ then eigenvalues (21), and in particular (26), are unchanged. Therefore, the asymptotic convergence rate is insensitive to shifts in C.

5. Connection with principal component flows

This section shows that the Oja–Brockett flow (1) is a special case of the minor component flow (2). It also derives a novel principal component flow based on (2).

For convenience, the Oja–Brockett flow is restated here but with different variable names,

$$\dot{Z} = AZN - ZNZ^{\mathrm{T}}AZ, \quad Z \in \mathbb{R}^{n \times p}, \tag{27}$$

where $A \in \mathbb{R}^{n \times n}$ and $N \in \mathbb{R}^{p \times p}$ are positive definite symmetric matrices with distinct eigenvalues. The columns of Z converge to the eigenvectors associated with the p largest eigenvalues of A in an order determined by N; see [32] for a proof.

Theorem 12. *Under the linear coordinate transformation*

$$X = \lambda^{-1/2} A^{1/2} Z N^{1/2}, \tag{28}$$

which is only defined if both A and N are positive definite symmetric matrices, the Oja–Brockett flow (27) *becomes*

$$\dot{X} = (A - \lambda I)XN + \lambda X(N - X^{\mathrm{T}}X), \tag{29}$$

which is the minor component flow (2) *with* $C = \lambda I - A$ *and* $\mu = \lambda$.

Proof. Define X as in (28). Then

$$\dot{X} = \lambda^{-1/2} A^{1/2} \dot{Z} N^{1/2} \tag{30}$$

$$= AXN - \lambda XX^{\mathrm{T}}X \tag{31}$$

$$= (A - \lambda I)XN + \lambda X(N - X^{\mathrm{T}}X), \tag{32}$$

where (31) is obtained by substituting (27) into (30). □

Since the Oja–Brockett flow (27) requires A and N to be positive definite and symmetric, transformation (28) from the Oja–Brockett flow to the minor component flow (2) is always valid. However, as is now shown, the reverse transformation from (2) to (27) is not always possible, meaning that the minor component flow (2) is a strict generalisation of the Oja–Brockett flow. The reverse of $C = \lambda I - A$ and $\mu = \lambda$ is $A = \mu I - C$ and $\lambda = \mu$. Since A must be positive definite for (28) to be defined (and for (27) to be stable), the reverse transformation is only valid if μ is larger than the largest eigenvalue of C.

Since (2) is a generalisation of (27), it is natural to consider using (2) with $C = \lambda I - A$ to find the principal components of A. Since there is no requirement for C to be positive definite in (2), the choice of λ is relatively unimportant. Indeed, the informal analysis in Section 4 suggests λ does not affect the asymptotic convergence rate.

6. Conclusion

The novel minor component flow (2) was derived and analysed. It was shown to be a generalisation of the Oja–Brockett principal component flow (1). The derivation of (2) was based on the observation that the penalty term in the cost function (4) has a benign effect on the critical points. Section 2 performed a local stability analysis about all critical points of a mild modification of this penalised cost function, showing that the only local minima are those corresponding to the minor components of C. Section 3 derived the flow (2) as a gradient flow minimising this modified cost function. Moreover, pointwise convergence of this flow to the minor components of C was proved. The effect of changing μ, N and C on the convergence rate of the flow was investigated in Section 4. Section 5 elucidated the connection between (2) and the Oja–Brockett flow. It also stated that replacing C by $-C$ in (2) results in a satisfactory principal component flow.

Acknowledgements

The authors gratefully acknowledge several helpful conversations with Dr Maziar Nikpour.

Appendix A. The orthonormal penalty function

The reason the cost function (4), whose penalty term forces X to be orthonormal, is not used is due to the possibility of critical points unrelated to the eigenvectors of C, as is now shown.

From Section 2, the critical points of (4) with $\mu = 1$ are the solutions of $CXN = X(I - X^{\mathrm{T}}X)$. Thus $X^{\mathrm{T}}CXN = X^{\mathrm{T}}X(I - X^{\mathrm{T}}X)$ and $CXNX^{\mathrm{T}} = (I - XX^{\mathrm{T}})XX^{\mathrm{T}}$ both hold. Assume C and N are both diagonal with distinct elements. It follows that both $X^{\mathrm{T}}X(I-X^{\mathrm{T}}X)$ and $(I-XX^{\mathrm{T}})XX^{\mathrm{T}}$ are diagonal. Unfortunately, this is not enough to conclude that $X^{\mathrm{T}}X$ is diagonal, as required if the columns of X are to correspond to eigenvectors of C. For example, if X is any two-by-two matrix with singular values $\sqrt{s}$ and $\sqrt{1-s}$, then $X^{\mathrm{T}}X(I - X^{\mathrm{T}}X) = (I - XX^{\mathrm{T}})XX^{\mathrm{T}} = s(1-s)I$.

The above observation can be used to construct examples of undesired critical points as follows. Let U and V be arbitrary orthogonal two-by-two matrices and set $S = \mathrm{diag}\{\sqrt{s}, \sqrt{1-s}\}$ for some constant s. Substitute $X = USV^{\mathrm{T}}$ into $CXN = X(I - X^{\mathrm{T}}X)$ to yield $S(U^{\mathrm{T}}CU)S = s(1-s)(V^{\mathrm{T}}N^{-1}V)$. Even though C and N are assumed to be diagonal, note that $U^{\mathrm{T}}CU$ and $V^{\mathrm{T}}N^{-1}V$ are arbitrary positive definite matrices. In particular, if V and N are chosen at random, with N positive definite to satisfy A3, a solution of $S(U^{\mathrm{T}}CU)S = s(1-s)(V^{\mathrm{T}}N^{-1}V)$ is found simply by taking the singular value decomposition of $S^{-1}(V^{\mathrm{T}}N^{-1}V)S^{-1}$ and reading off suitable values for U and C. In this way, diagonal matrices C and N can be found such that there exists a critical point X whose columns are not related to the eigenvectors of C.

References

[1] A. Benveniste, M. Métivier, P. Priouret, Adaptive Algorithms and Stochastic Approximations, Springer, Berlin, 1990.
[2] R.W. Brockett, Dynamical systems that sort lists, diagonalise matrices, and solve linear programming problems, Linear Algebra Appl. 146 (1991) 79–91.
[3] Y. Chauvin, Principal component analysis by gradient descent on a constrained linear Hebbian cell, in: Proceedings of the Joint International Conference on Neural Networks, San Diego, CA, 1989, pp. 373–380.
[4] T. Chen, Y. Hua, W.-Y. Yan, Global convergence of Oja's subspace algorithm for principal component extraction, IEEE Trans. Neural Networks 9 (1) (1998) 58–67.
[5] P. Comon, G.H. Golub, Tracking a few extreme singular values and vectors in signal processing, Proc. IEEE 78 (8) (1990) 1327–1343.
[6] U. Helmke, J.B. Moore, Optimization and Dynamical Systems, Springer, Berlin, 1994.
[7] M.W. Hirsch, S. Smale, Differential Equations, Dynamical Systems, and Linear Algebra, Academic Press, New York, 1974.
[8] Z. Kang, C. Chatterjee, V.P. Roychowdhury, An adaptive quasi-Newton algorithm for eigensubspace estimation, IEEE Trans. Signal Process. 48 (12) (2000) 3328–3333.
[9] H.J. Kushner, G.G. Yin, Stochastic Approximation Algorithms and Applications, Springer, Berlin, 1997.
[10] S. Lojasiewicz, Sur les trajectoires du gradient d'une fonction analytique, Seminari di Geometria, 1983, pp. 115–117.

[11] I.Y. Mareels, J.W. Polderman, Adaptive Systems: An Introduction, Birkhäuser, Basel, 1996.
[12] G. Mathew, V.U. Reddy, S. Dasgupta, Adaptive estimation of eigensubspace, IEEE Trans. Signal Process. 43 (2) (1995) 401–411.
[13] E. Oja, A simplified neuron model as a principal component analyzer, J. Math. Biol. 15 (1982) 267–273.
[14] E. Oja, Neural networks, principal components, and subspaces, Int. J. Neural Systems 1 (1989) 61–68.
[15] E. Oja, Principal components, minor components, and linear neural networks, Neural Networks 5 (1992) 927–935.
[16] E. Oja, J. Karhunen, On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix, J. Math. Anal. Appl. 106 (1985) 69–84.
[17] E. Oja, H. Ogawa, J. Wangviwattana, Principal component analysis by homogeneous neural networks, Part I: the weighted subspace criterion, IEICE Trans. Inform. Systems 3 (1992) 366–375.
[18] E. Oja, H. Ogawa, J. Wangviwattana, Principal component analysis by homogeneous neural networks, Part II: analysis and extensions of the learning algorithms, IEICE Trans. Inform. Systems 3 (1992) 376–382.
[19] W.K. Pratt, Digital Image Processing, Wiley, New York, 1978.
[20] D.J. Rabideau, Fast, rank adaptive subspace tracking and applications, IEEE Trans. Signal Process. 44 (9) (1996) 2229–2244.
[21] E.C. Real, D.W. Tufts, J.W. Cooley, Two algorithms for fast approximate subspace tracking, IEEE Trans. Signal Process. 47 (7) (1999) 1936–1945.
[22] V.U. Reddy, B. Egardt, T. Kailath, Least squares type algorithm for adaptive implementation of Pisarenko's harmonic retrieval method, IEEE Trans. Acoust. Speech Signal Process. 30 (3) (1982) 399–403.
[23] T.D. Sanger, Optimal unsupervised learning in a single-layer linear feedforward network, Neural Networks 2 (1989) 459–473.
[24] R. Schreiber, Implementation of adaptive array algorithms, IEEE Trans. Acoust. Speech Signal Process. 34 (5) (1986) 1038–1045.
[25] P. Strobach, Square-root QR inverse iteration for tracking the minor subspace, IEEE Trans. Signal Process. 48 (11) (2000) 2994–2999.
[26] X. Wang, H.V. Poor, Blind multiuser detection: a subspace approach, IEEE Trans. Inform. Theory 44 (2) (1998) 677–689.
[27] L. Xu, Least mean square error recognition principle for self organizing neural nets, Neural Networks 6 (1993) 627–648.
[28] W.-Y. Yan, U. Helmke, J.B. Moore, Global analysis of Oja's flow for neural networks, IEEE Trans. Neural Networks 5 (5) (1994) 674–683.
[29] B. Yang, Projection approximation subspace tracking, IEEE Trans. Signal Process. 43 (1) (1995) 95–107.
[30] J.-F. Yang, M. Kaveh, Adaptive eigensubspace algorithms for direction or frequency estimation and tracking, IEEE Trans. Acoust. Speech Signal Process. 36 (2) (1988) 241–251.
[31] X. Yang, T.K. Sarkar, E. Arvas, A survey of conjugate gradient algorithms for solution of extreme eigen-problems of a symmetric matrix, IEEE Trans. Acoust. Speech Signal Process. 37 (10) (1989) 1550–1556.
[32] S. Yoshizawa, U. Helmke, K. Starkov, Convergence analysis for principal component flows, Int. J. Appl. Math. Comput. Sci. 11 (2001) 223–236.