Proceedings of the 17th International Symposium on Mathematical Theory of Networks and Systems, Kyoto, Japan, July 24-28, 2006

A generalisation of the Oja subspace flow

Christfried Webers and Jonathan H. Manton

Abstract— **Recently, a novel flow for computing the eigenvectors associated with the smallest eigenvalues of a symmetric but not necessarily positive definite matrix was introduced. This meant that the eigenvectors associated with the smallest eigenvalues could be found simply by reversing the sign of the matrix. The current paper derives a cost function and the corresponding negative gradient flow which converges to the same subspace as is spanned by the minor components. The flow is related to Oja's major subspace flow which converges for positive definite matrices. With a suitable coordinate transformation, the Oja flow can be converted into the corresponding novel minor subspace flow. But the minor subspace flow does also converge for cases where no transformation back into an Oja flow exists. Thus, the novel flow is a generalisation of the Oja subspace flow.**

Keywords— **Algebraic and differential geometry, Eigenvalues and eigenfunctions, Hessian matrices, Iterative methods, Matrix decomposition, Signal subspaces, Stability analysis**

I. INTRODUCTION

In many areas of information processing, it is necessary to extract the minor (major) components or the corresponding minor (major) subspaces from a high dimensional input data stream. The p minor (major) components of a matrix C are the eigenvectors associated with the p smallest (largest) eigenvalues of C. The minor (major) subspace of dimension p of a matrix C is the subspace spanned by the p minor (major) components of C.

Recently, a continuous time gradient flow for an indefinite real symmetric matrix $C \in \mathbb{R}^{n\times n}$ was introduced in [1]. The flow minimises the cost function $f_N : \mathbb{R}^{n\times p} \to \mathbb{R}$,

$$f_N(X) = \frac{1}{2}\operatorname{tr}\{X^T C X N\} + \frac{1}{4}\mu\|N - X^T X\|_F^2 \qquad (1)$$

where $\operatorname{tr}\{\cdot\}$ denotes the trace of a matrix, $\|\cdot\|_F$ denotes the Frobenius norm of a matrix, $X \in \mathbb{R}^{n\times p}$ with $n \geq p$, $N \in \mathbb{R}^{p\times p}$ is a diagonal matrix with distinct positive eigenvalues and $\mu \in \mathbb{R}$ is a strictly positive constant different from the eigenvalues of C. It was shown that for μ large enough, the gradient flow of this cost function

$$\dot{X} = -CX + \mu X(N - X^T X) \qquad (2)$$

converges to the minor components of C.

It was also shown in [1], that the introduction of the diagonal matrix N in the penalty term of (1) is essential for the flow to converge to the minor components. Replacing N in the penalty term with the identity matrix, there exist minima X of the cost function (1) for which the columns of X are not the minor components of C.

Another flow not converging to the eigenvectors but only to a set of orthonormal vectors spanning the major subspace is Oja's principal subspace flow [2], [3]. It is defined as

$$\dot{Z} = (I - ZZ^T)AZ \qquad (3)$$

where $Z \in \mathbb{R}^{n\times p}$, $n \geq p$, I denotes the identity matrix and $A \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix. The convergence of the Oja flow was proven in [4] which also developed a phase portrait analysis for the flow. If $A < 0$, the Oja flow computes the minor subspace but diverges.

The main result of this paper is the analysis of the novel cost function

$$f(X) = \frac{1}{2}\operatorname{tr}\{X^T C X\} + \mu\frac{1}{4}\|I - X^T X\|_F^2 \qquad (4)$$

which is derived from (1) by replacing N with the identity matrix. The minimisation of this novel cost function leads to a matrix X with columns spanning the minor subspace of C.

Associated with the cost function $f(X)$ is the negative gradient flow $\dot{X} = -\operatorname{grad} f(X)$ given by

$$\dot{X} = -CX + \mu X(I - X^T X) \qquad (5)$$

where the gradient is taken with respect to the Euclidian inner product $\langle A, B\rangle = \operatorname{tr}\{B^T A\}$ on matrix space. The flow (5) generically converges to a minimum of the cost function $f(X)$.

It will be further shown that the novel flow is a generalisation of the Oja flow to indefinite matrices. With a suitable coordinate transformation, the Oja flow can be transformed into the flow (5). But there exist matrices C in (5) for which the inverse transformation is not defined. Hence, (5) is a strict generalisation of the Oja flow.

Finally, in a numerical experiment the convergence speed of the novel flow (5) will be compared to the flow (2).

II. COST FUNCTION

A. Properties of the cost function

The cost function (4) can be obtained from (1) by replacing N with the identity matrix. This change introduces an important symmetry into the cost function $f(X)$ expressed in the following lemma. This symmetry means that the analysis given in [1] does not carry over.

Lemma 1: Given an arbitrary orthogonal matrix $V \in \mathbb{R}^{p\times p}$, the following equation holds

$$f(X) = f(XV) \qquad (6)$$

C. Webers and J. H. Manton are with Department of Information Engineering, Research School of Information Sciences and Engineering, The Australian National University, Canberra ACT 0200, Australia {christfried.webers,jonathan.manton}@ieee.org

One consequence of this symmetry property of the cost function $f(X)$ is that in general minima of the cost function are not isolated points but manifolds. [1] This can be easily seen as for any minimum X and arbitrary orthogonal matrix V, all points of the connected set XV are also minima for which the cost function has the same value. This complicates the convergence analysis of the flow. Additionally to the convergence to the critical manifold, it will also be conjectured, that the flow converges to one point on the manifold.

The next lemma establishes the existence of a lower bound and bounded sublevel sets for the cost function.

Lemma 2: The cost function $f(X)$ of (4) is lower bounded and, for any constant $c \in \mathbb{R}$ and μ being strictly positive its sublevel set $\{X : f(X) \leq c\}$ is bounded.

Proof: The cost function $f(X)$ can be written as

$$f(X) = \frac{1}{2}\operatorname{tr}\{X^T CX\} + \mu\frac{1}{4}\operatorname{tr}\{I - 2X^T X + X^T XX^T X\} \tag{7}$$

and for large X the dominant contribution comes from the fourth order $X^T XX^T X$ term which is positive semidefinite and its trace has compact sublevel sets. □

The following section analyses the form of critical points for the cost function $f(X)$.

B. Critical points

The following assumptions are made throughout.

A1. The scalar μ is strictly positive.

A2. The matrix $C \in \mathbb{R}^{n\times n}$ is symmetric and has distinct eigenvalues.

A3. The scalar μ does not equal any eigenvalue of C.

The critical points of the cost function $f(X)$ are the points for which the directional derivative of $f(X)$ vanishes in all directions. The directional derivative in the direction $\xi \in \mathbb{R}^{n\times p}$ is

$$\mathcal{D}f(X)\xi = \operatorname{tr}\{\xi^T(CX + \mu XX^T X - \mu X)\} \tag{8}$$

For a point X to be a critical point of $f(X)$ it must satisfy the following equation

$$CX = \mu X(I - X^T X) \tag{9}$$

The following theorem establishes the basis for further analysis.

Proposition 1 (Critical Points): A necessary and sufficient condition for $X = \begin{bmatrix} x_1 & x_2 & \cdots & x_p \end{bmatrix}$ to be a critical point of (4) is that X has the form $X = QV$ for arbitrary orthogonal matrices $V \in \mathbb{R}^{p\times p}$ and for each of the columns q_r, $r = 1, \ldots, p$ of Q there are two choices.

(1) q_r is either the null vector, or

(2) q_r is an eigenvector of C with corresponding eigenvalue λ_r and $\|q_r\|^2 = 1 - \lambda_r/\mu$.

All nonzero vectors q_r are pairwise orthogonal to each other.

[1] We conjecture the minima are manifolds.

Proof: The singular value decomposition of any point $X \in \mathbb{R}^{n\times p}$ is

$$X = U\begin{bmatrix}\Sigma\\0\end{bmatrix}V^T \tag{10}$$

where $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{p\times p}$ are orthonormal matrices and the matrix $\Sigma \in \mathbb{R}^{p\times p}$ is diagonal with singular values $\sigma_r \geq 0$. Note, that with this definition no particular ordering of the singular values σ_r is assumed.

Inserting (10) into (9) one gets

$$CU\begin{bmatrix}\Sigma\\0\end{bmatrix}V^T = \mu U\begin{bmatrix}\Sigma\\0\end{bmatrix}V^T(I_p - V\Sigma U^T U\Sigma^T V^T) \tag{11}$$

This can be simplified to

$$CU\begin{bmatrix}\Sigma\\0\end{bmatrix} = \mu U\begin{bmatrix}\Sigma\\0\end{bmatrix}(I_p - \Sigma^2) \tag{12}$$

with $\Sigma^2 = \operatorname{diag}\{\sigma_1^2, \ldots, \sigma_p^2\} \in \mathbb{R}^{p\times p}$. Writing (12) for each of the first p columns of U leads to p equations of the form

$$Cu_r\sigma_r = \mu u_r\sigma_r(1 - \sigma_r^2) \tag{13}$$

With the definition $q_r = u_r\sigma_r$ this can be written as

$$Cq_r = \mu q_r(1 - \sigma_r^2) \tag{14}$$

Now, either $\sigma_r = 0$ in which case $q_r = 0$. Or $\sigma_r \neq 0$ in which case q_r is an eigenvector of C with corresponding eigenvalue $\lambda_r = \mu(1 - \sigma_r^2)$. The squared norm of the eigenvector q_r is $\|q_r\|^2 = \sigma_r^2 = 1 - \lambda_r/\mu$. The pairwise orthogonality of nonzero vectors q_r follows from the definition $q_r = u_r\sigma_r$ and the fact that the u_r are columns of an orthogonal matrix U.

□

III. Local stability of critical points

It follows from Proposition 1 that if X is a critical point of (4) then there exist orthogonal matrices $U \in \mathbb{R}^{n\times n}$, $V \in \mathbb{R}^{p\times p}$ and a diagonal matrix $\Sigma \in \mathbb{R}^{p\times p}$ such that $X = U\begin{bmatrix}\Sigma\\0\end{bmatrix}V^T$. Furthermore, each of the columns u_r, $r = p + 1, \ldots, n$ of U and all u_r, $r = 1, \ldots, p$ for which $\sigma_r = 0$ can be freely chosen as they do not contribute anything to the value of X. One special choice simplifying further analysis will be to fill the columns with unit norm eigenvectors of C not already associated with nonzero σ_r, $r = 1, \ldots, p$. This is always possible as C was assumed to have distinct eigenvalues. With this choice for U, $\Lambda = U^T CU$ will be a diagonal matrix cointaining the eigenvalues of C in an arbitrary order. The singular values in a singular value decomposition are all nonnegative, and hence, each of the values σ_r can take at most two different values; either $\sigma_r = 0$ or $\sigma_r = \sqrt{1 - \lambda_r/\mu}$.

The stability of the critical points can now be determined by directly inspecting the eigenvalues of the Hessian at the critical points.

Lemma 3: (Hessian of $f(X)$) Let X be a critical point of the cost function $f(X)$ defined in (9). Let the unitary matrix U and real diagonal matrix Σ be such that $\Lambda =$

	r	s	$\sigma_r = 0$	$\sigma_r = \sqrt{1-\lambda_r/\mu}$
α_r	$1 \ldots p$		$\lambda_r - \mu$	$2(\mu - \lambda_r)$
β_{sr}	$1 \ldots p$	$p+1 \ldots n$	$\lambda_s - \mu$	$\lambda_s - \lambda_r$

TABLE I

α_r AND β_{sr} AT A CRITICAL POINT X

	$\sigma_r = 0$		$\sigma_r > 0$	
	$\sigma_s = 0$	$\sigma_s > 0$	$\sigma_s = 0$	$\sigma_s > 0$
e_1	$\lambda_r - \mu$	0	0	0
e_2	$\lambda_s - \mu$	$\lambda_r - \lambda_s$	$\lambda_s - \lambda_r$	$2\mu - \lambda_r - \lambda_s$

TABLE II

EIGENVALUES $E = \operatorname{diag}\{e_1, e_2\}$ OF $\Gamma_{(rs)}V = VE$ FOR $r = 1 \ldots p-1$ AND $s = r+1 \ldots p$ AT A CRITICAL POINT X

U^TCU and $X = U \begin{bmatrix}\Sigma \\ 0\end{bmatrix} V^T$; such U, V and Σ always exists. Then, the second directional derivative $g(\xi) = \mathcal{D}^2 f(X)\xi$ in the direction ξ at the point X is given by the quadratic form

$$g(U\xi V^T) = \sum_{r=1}^{p} \alpha_r \xi_{rr}^2 + \sum_{s=p+1}^{n} \sum_{r=1}^{p} \beta_{sr} \xi_{sr}^2 + \sum_{r=1}^{p-1} \sum_{s=r+1}^{p} [\xi_{rs}\ \xi_{sr}]\, \Gamma_{(rs)}\, [\xi_{rs}\ \xi_{sr}]^T \tag{15}$$

where the scalars $\alpha_r, \beta_{sr} \in \mathbb{R}$ are

$$\alpha_r = \lambda_r - \mu + 3\mu\sigma_r^2 \tag{16}$$

$$\beta_{sr} = \lambda_s - \mu + \mu\sigma_r^2 \tag{17}$$

and the matrix $\Gamma_{(rs)} \in \mathbb{R}^{2\times 2}$ is

$$\Gamma_{(rs)} = \begin{bmatrix} \lambda_r - \mu + \mu(\sigma_r^2 + \sigma_s^2) & \mu\sigma_r\sigma_s \\ \mu\sigma_r\sigma_s & \lambda_s - \mu + \mu(\sigma_r^2 + \sigma_s^2) \end{bmatrix} \tag{18}$$

Proof: The second directional derivative of $f(X)$ is the term quadratic in h when expanding the cost function $f(X + h\xi)$

$$\mathcal{D}^2 f(X)\xi = \operatorname{tr}\{\xi^T(C\xi - \mu\xi(I - X^TX) + \mu X\xi^TX + \mu XX^T\xi)\}. \tag{19}$$

For a given critical point X, define $g(\xi)$ to be this quadratic form. By Proposition 1, there exists orthogonal matrices U and V and a diagonal matrix Σ such that $U^TCU = \Lambda$ or $C = U\Lambda U^T$ and $X = U \begin{bmatrix}\Sigma \\ 0\end{bmatrix} V^T$. Thus

$$\begin{aligned} g(U\xi V^T) &= \operatorname{tr}\{\xi^T\Lambda\xi - \mu\xi^T\xi(I - \Sigma^2) + \mu\xi^T \begin{bmatrix}\Sigma \\ 0\end{bmatrix} [\,\Sigma\ 0\,]\,\xi + \mu\xi^T \begin{bmatrix}\Sigma \\ 0\end{bmatrix} \xi^T \begin{bmatrix}\Sigma \\ 0\end{bmatrix}\} \\ &= \sum_{s=1}^{n}\sum_{r=1}^{p} (\lambda_s - \mu(1 - \sigma_r^2))\xi_{rs}^2 + \mu \sum_{r=1}^{p}\sum_{s=1}^{p} (\sigma_r^2\xi_{rs}^2 + \sigma_r\sigma_s\xi_{rs}\xi_{sr}) \\ &= \sum_{r=1}^{p} \alpha_r \xi_{rr}^2 + \sum_{s=p+1}^{n}\sum_{r=1}^{p} \beta_{sr}\xi_{sr}^2 + \sum_{r=1}^{p-1}\sum_{s=r+1}^{p} [\xi_{rs}\ \xi_{sr}]\, \Gamma_{(rs)}\, [\xi_{rs}\ \xi_{sr}]^T \end{aligned} \tag{20}$$

where the α_r, β_{sr} and $\Gamma_{(rs)}$ are as given in the lemma. □

The minima of the cost function $f(X)$ are the critical points for which the eigenvalues of the Hessian are non-negative. Each of the singular values σ_r, $r = 1,\ldots, p$ can be either $\sigma_r = 0$ or, provided $\mu > \lambda_r$, $\sigma_r^2 = 1 - \lambda_r/\mu$. Tables I and II show these eigenvalues for all possible choices of σ_r. The eigenvalues of the Hessian are used to establish the form of X at a local minimum.

Proposition 2 (Local minima): Let $\lambda_1 < \ldots < \lambda_n$ be the eigenvalues of C in ascending order and let $u_1, \ldots, u_n$ be the corresponding eigenvectors. Then $X = \begin{bmatrix} q_1 & q_2 & \cdots & q_p \end{bmatrix} V^T$ is a local minimum of $f(X)$ for an arbitrary orthogonal matrix $V \in \mathbb{R}^{p\times p}$ if and only if $q_r = \pm\gamma_r u_r$ for all $r = 1,\ldots, p$ with $\gamma_r = \sqrt{1 - \lambda_r/\mu}$ for $\lambda_r < \mu$ and $\gamma_r = 0$ otherwise.

Proof: First, assume all $\sigma_r > 0$. For the eigenvalues of α_r to be positive, this necessitates $\mu > \lambda_r$ for $r = 1,\ldots, p$. Now, for β_{sr} to be positive $\lambda_s > \lambda_r$ for all $r = 1,\ldots, p$ and $s = p+1,\ldots, n$. Finally, one of the eigenvalues of $\Gamma_{(rs)}$ is zero indicating that the minima are not isolated points. The other eigenvalue for $\Gamma_{(rs)}$ is always positive because the above condition on α_r implies $\mu > \lambda_r$, for $r = 1,\ldots, p$ and therefore $2\mu - \lambda_r - \lambda_s > 0$ for $r, s = 1,\ldots, p$.

Now assume that one singular value is zero. Let k be the smallest integer such that $\sigma_k = 0$. Then $\lambda_k > \mu$ in order to satisfy $\alpha_k > 0$. From $\beta_{sk} > 0$ follows $\lambda_s > \mu$, $s = p+1,\ldots, n$. Now from $\Gamma_{(ks)} \geq 0$, $s = k+1,\ldots, p$ there are two cases. Either $\sigma_s = 0$ and therefore $\lambda_k > \mu$ and $\lambda_s > \mu$ or $\sigma_s \neq 0$. In the latter case, one eigenvalue is zero and from the other $\lambda_k > \lambda_s$ is concluded. This means, all eigenvalues associated with $\sigma_r \neq 0$ have to be smaller than the eigenvalues of eigenvectors associated with $\sigma_r = 0$. This means, the eigenvectors q_r for which the corresponding eigenvalues $\lambda_r < \mu$ are the eigenvectors associated with the smallest eigenvalues of C. Now observe, that the eigenvectors q_r can be sorted in ascending order by multiplying from right with a permutation matrix Π. As Π is an orthogonal matrix and orthogonal matrices form a group. Thus, ΠV^T is another orthogonal matrix. □

IV. MINOR SUBSPACE FLOW

It was shown that the critical points of the cost function $f(X)$ are not isolated but sets. In the following, a conjecture about the convergence of the flow to the minor subspace and an overview of the proof is given. First, a definition for a minor subspace flow is proposed.

Definition 1 (Minor subspace flow): The flow $\dot{X} = f(X, C)$ on $\mathbb{R}^{n\times p}$ is a minor subspace flow for matrices

C belonging to the class $\mathcal{C} \subset \mathbb{R}^{n\times n}$ if the following hold:

(1) For any initial condition X_0 and any $C \in \mathcal{C}$, a solution $X(t)$ of the flow $\dot{X} = f(X, C)$ satisfying $X(0) = X_0$ exists and is unique for all $t \geq 0$.
(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_∞ span the same subspace as the subspace spanned by the eigenvectors associated with the p smallest eigenvalues of C, counting multiplicities.

With these definition, the convergence for the minor subspace flow (5) can now be conjectured.

Conjecture 1 (Minor subspace flow): For arbitrary integers $1 \leq p \leq n$, let $\mu > 0$ and define $\mathcal{C}_\mu$ 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, the flow (5) is a minor subspace flow for $C \in \mathcal{C}_\mu$.

In order to proof the conjecture, one has to show that $f(X)$ is a Morse-Bott function which implies the following.

Definition 2: [5] Let M be a smooth manifold and let $\Phi : M \to \mathbb{R}$ be a smooth function. Let $C(\Phi) \subset M$ denote the set of all critical points of Φ. Then Φ is called a *Morse-Bott* function provided the following three conditions are satisfied.

1) $\Phi : M \to \mathbb{R}$ has compact sublevel sets.
2) $C(\Phi) = \bigcup_{j=1}^{k} N_j$ with N_j disjoint, closed and connected submanifolds of M such that Φ is constant on N_j, $j = 1, \ldots, k$.
3) $\ker(H_\Phi(x)) = T_x N_j, \forall x \in N_j$, $j = 1, \ldots, k$.

The following section will relate this minor subspace flow to Oja's major subspace flow.

V. Relation to the Oja major subspace flow

This section will show that the novel flow is a generalisation of the Oja flow.

Theorem 1: Consider the Oja flow (3). As A is positive definite, the linear coordinate transformation

$$X = \lambda^{-1/2} A^{1/2} Z \tag{21}$$

is defined for all $Z, X \in \mathbb{R}^{n\times p}$ if $\lambda > 0$. With this transformation the Oja flow (3) becomes

$$\dot{X} = (A - \lambda I)X + \lambda X(I - X^T X) \tag{22}$$

which is the minor subspace flow (5) with $C = \lambda I - A$ and $\mu = \lambda$.

Proof: Differentiate (21) and insert (3) for $\dot{Z}$ to get

$$\dot{X} = \lambda^{-1/2} A^{1/2} \dot{Z} \tag{23}$$
$$= AX - \lambda X X^T X \tag{24}$$
$$= (A - \lambda I)X + \lambda X(I - X^T X). \tag{25}$$

□

Hence, the transformation from the Oja flow (3) to the flow (5) is always valid.

Now consider the inverse transformation from the flow (5) to the Oja flow (3) given by

$$Z = \lambda^{1/2} A^{-1/2} X \tag{26}$$

If μ is chosen larger than then pth smallest eigenvalue of C but smaller than the largest eigenvalue of C, the minor subspace flow converges. The matrix A defined via the inverse transformation of (21)

$$A = \mu I - C \qquad \mu = \lambda \tag{27}$$

is not positive definite. Therefore, $A^{-1/2}$ in (26) is not defined and the corresponding Oja flow does not exist. In this respect, the subspace flow (5) is a generalisation of the Oja subspace flow (3).

However, the novel flow does not converge to orthogonal columns as the Oja flow does. This can be seen by applying the linear transformation (21) to the limit point Z_∞ observing that $Z_\infty^T Z_\infty = I$ is valid.

$$I = Z_\infty^T Z_\infty = \mu X_\infty^T A^{-1} X_\infty \tag{28}$$

If orthogonal vectors spanning the minor subspace are needed, an extra step orthogonalising the colums of X_∞ will be necessary.

VI. Relation between novel minor and major subspace flow

The minor subspace flow (5) can be directly turned into a major subspace flow by reversing the sign of the matrix C. This can be seen in the following way. First note, that the matrices C and $-C$ share the same set of unit norm eigenvectors u_r, $r = 1, \ldots, p$ as

$$CU = U\Lambda \tag{29}$$
$$(-C)U = U(-\Lambda) \tag{30}$$

where Λ is the diagonal matrix of eigenvalues and $U = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix}$. Note, that the unit norm eigenvectors are uniquely defined (apart from a sign) because the matrix C is assumed to have distinct eigenvalues. Now, the p smallest eigenvalues of Λ correspond to the p largest eigenvalues of $-\Lambda$ and vice versa.

Hence, the flow

$$\dot{X} = CX + \mu X(I - X^T X) \tag{31}$$

is a major subspace flow for the matrix C. It converges, if and only if $\mu > 0$ and μ is greater than at least p of the eigenvalues of $-C$.

VII. Numerical examples

A. Angle between subspaces

In order to quantify the convergence of the numerical examples, a measure for the similarity of two subspaces of $\mathbb{R}^n$ spanned by the columns of two matrices $A \in \mathbb{R}^{n\times r}$ and $B \in \mathbb{R}^{n\times s}$ is needed. One possible measure is based on the concept of principal angles recursively defined between subspaces ([6], [7]). Refinements introduced in [8] ensure that also small angles are calculated accurately. [2] If both matrices A and B have the same number of columns, the angle can be computed in the following way:

[2]Note that the Matlab command *subspace* prior to release 13 does not provide the accurate answers for small angles.

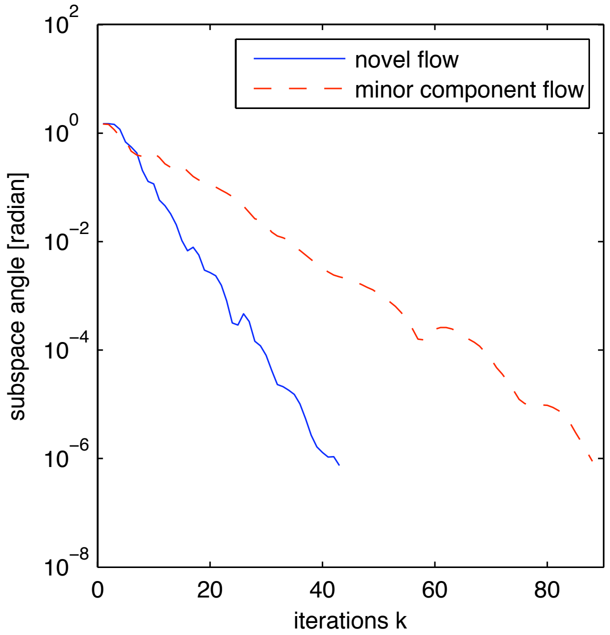

Fig. 1. Comparison of convergence rate between the novel subspace flow (5) and the minor component flow (2) ($n = 20$, $p = 6$)

1. Find orthonormal bases Q_A and Q_B for the subspaces spanned by the columns of A and B employing the QR factorisation of the matrices.
2. Calculate the angle $\theta = \arcsin(\|Q_A^T - (Q_A^T Q_B)Q_B^T\|)$

In all subsequent experiments, the angle will be measured between the subspace generated from the discretisation of the flow and the subspace generated by the minor (major) components calculated with the Matlab command *eig*.

B. Iteration setup

The goal is to find an approximation to the solution of the flow for a given cost function $f(x)$ given an initial value $X_0 \in \mathbb{R}^{n\times p}$ by discretising the corresponding flow. Given an approximation X_k at step k, the next iteration is calculated as

$$X_{k+1} = X_k + \alpha^* D_k \tag{32}$$

where D_k is the search direction and α^* is the positive step size minimising $h(\alpha) = f(X + \alpha D_k)$. As the cost functions used here are of order four, minimisation of $h(\alpha)$ can be done by explicitely solving a third order scalar equation in α resulting in one or three real solutions for the optimal step size α^*. For the search direction, the steepest descent (SD) with

$$D_k = -G_k \tag{33}$$

or the Polak-Ribiere conjugate gradient (PRCG) with

$$D_k = -G_k + \frac{< G_k - G_{k-1}, G_k >}{< G_{k-1}, G_{k-1} >} \tag{34}$$

is employed [9]. Here, G_k is the gradient of the cost function for the iterate X_k.

C. Comparison between minor subspace and minor component flow

The minor subspace flow provides a faster convergence rate then the minor component flow as shown in Figure 1. This is an advantage for applications, where the minor components are not needed because a set of vectors spanning the minor subspace is sufficient. This increase in convergence speed can be explained with the minor subspace flow having less constraints to obey than the minor component flow where the order of the eigenvectors is specified by the order of the corresponding eigenvalues of the diagonal matrix N.

VIII. Conclusion

This paper introduced a novel minor subspace flow which is defined for indefinite symmetric matrices. A close relation to Oja's major subspace flow exists because all Oja flows can be converted to a corresponding novel flow. Moreover, the novel flow converges also for matrices which can not be related to an Oja flow. A numerical experiment illustrated the convegence of the novel flow.

References

[1] J. H. Manton, U. Helmke, and I. M. Y. Mareels, "A dual purpose principal and minor component flow," *Systems and Control Letters*, vol. 54, pp. 759–769, 2005.

[2] E. Oja, "A simplified neuron model as a principal component analyzer," *Journal of Mathematical Biology*, vol. 15, no. 3, pp. 267–273, November 1982.

[3] ——, "Neural networks, principal components, and subspaces," *International Journal of Neural Systems*, vol. 1, pp. 61–68, 1989.

[4] W. Y. Yan, U. Helmke, and J. B. Moore, "Global analysis of oja's flow for neural networks," *IEEE Transactions on Neural Networks*, vol. 5, no. 5, pp. 674–683, 1994.

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

[6] A. Bjorck and G. H. Golub, "Numerical methods for computing angles between linear subspaces," *Mathematics of Computation*, vol. 27, no. 123, pp. 579–594, 1973.

[7] G. H. Golub and C. F. van Loan, *Matrix Computations*. The Johns Hopkins University Press, 1996.

[8] P.-A. Wedin, *On angles between subspaces of a finite dimensional inner product space*, ser. Lecture Notes in Mathematics. Springer, 1983, vol. 973, pp. 263–285.

[9] M. Nikpour, J. H. Manton, M. Iven M.Y, and V. Adamyan, "Algorithms for extreme eigenvalue problems," in *16. International Symposium on Mathematical Theory of Networks and Systems (MTNS04)*, 2004.