Convergence Analysis for Principal Component Flows *

S. Yoshizawa
Department of Mathematics
Faculty of Science and Technology
Science University of Tokyo
Noda, Chiba 278, Japan

U. Helmke
Department of Mathematics
University of Würzburg
D–97074 Würzburg
Germany

K. Starkov
CITEDI–IPN
Otay Mesa
San Diego 92173
CA, USA

1 Introduction

In the theory of neural networks, dynamical systems methods are crucial for the analysis of learning algorithms and conversely, neural network learning algorithms provide interesting examples of dynamical systems that represent challenges for a rigorous mathematical analysis. Although some of these learning algorithms seem to work quite well in practise, their theoretical analysis is still lacking a rigorous convergence theory. So far, most results in the literature have focussed on local stability issues without addressing the problem of global convergence. In this paper we discuss a special class of learning algorithms for principal component analysis for which a complete phase portrait analysis can be developed.

The one-unit Oja flow is a cubic differential equation

$$\dot{x} = (I - xx^\top)Ax \tag{1.1}$$

on $\mathbb{R}^n$. It has the remarkable property that — for any nonzero initial condition $x(0)$ — its solutions converge for $t \to \infty$ to the eigenvectors of the positive definite covariance matrix A.

Oja and Karhunen's work has been the precursor for subsequent work on flows for principal component or principal subspace analysis. Oja [8] has considered the matrix generalization of (1.1)

$$\dot{X} = (I - XX^\top)AX, \quad X \in \mathbb{R}^{n\times k} \tag{1.2}$$

*Research partially supported by the German–Israeli Foundation for scientific research and development under grant GIF–I–526–034.06/97, and by the DFG project 436 RUS 113/275/0(R).

for principal subspace analysis and conjectured that solutions $X(t)$ of (1.2) converge for $t \to \infty$ to an orthonormal basis matrix X_∞ of an invariant subspace of A. This conjecture has been proven by Yan et al [15], where a phase portrait analysis of (1.2) is developed. In particular, an explicit description of the domain of attraction for the locally attracting equilibria has been given.

In this paper we focus on Xu's principal component flow

$$\dot{X} = (I - XDX^\top)AXD \tag{1.3}$$

and prove global convergence of (1.3) to the principal components of A. Moreover, the eigenvalues of the linearization of (1.3) around the quilibria are determined and thus a local stability analysis is given. Corresponding results for the flows proposed by Oja et al [9, 10] are obtained as well.

2 Phase Portrait of the Oja–Brockett flow

For $1 \leq k \leq n$ let $A \in \mathbb{R}^{n\times n}$ and $D \in \mathbb{R}^{k\times k}$ be positive definite matrices with $A = A^\top > 0$, $D = D^\top > 0$. Let $X \in \mathbb{R}^{n\times k}$. We consider the matrix differential equation on $\mathbb{R}^{n\times k}$

$$\dot{X} = AXD - XDX^\top AX. \tag{2.1}$$

We refer to (2.1) as the *Oja–Brockett system* as it is a natural generalization of both Oja's flow (1.2) for principal subspace analysis (for $D = I_k$) as well as Brockett's [3] flow

$$\dot{X} = AXDX^\top X - XDX^\top AX$$

on orthogonal matrices for symmetric matrix diagonalization (for $k = n$ and $X \in \mathbb{R}^{n\times n}$ a real orthogonal matrix). The unconstrained flow (2.1) has been first analyzed by Xu [14], but the full analysis given here appears to be new.

Our first main result on the Oja–Brockett flow (2.1) on $\mathbb{R}^{n\times k}$ is that it is actually a gradient flow. Consider the real analytic function $f : \mathbb{R}^{n\times k} \to \mathbb{R}$ defined by

$$f(X) := \frac{1}{4}\operatorname{tr}(AXDX^\top)^2 - \frac{1}{2}\operatorname{tr}(A^2XD^2X^\top) = \frac{1}{4}\|(A^{\frac{1}{2}}XDX^\top A^{\frac{1}{2}}\|^2 - \frac{1}{2}\|AXD\|^2 \tag{2.2}$$

where $\|X\|^2 = \operatorname{tr}(XX^\top)$ denotes the Frobenius norm.

Lemma 2.1 *$f : \mathbb{R}^{n\times k} \to \mathbb{R}$ is a lower bounded function with compact sublevel sets.*

In the sequel we endow $\mathbb{R}^{n\times k}$ with the Riemannian metric

$$\langle \Omega_1, \Omega_2\rangle_{A,D} := \operatorname{tr}(A\Omega_1 D\Omega_2^\top), \quad \Omega_i \in \mathbb{R}^{n\times k} \tag{2.3}$$

and corresponding norm square

$$\|\Omega\|^2_{A,D} := \operatorname{tr}(A\Omega D\Omega^\top).$$

Note that (2.3) defines a positive definite inner product on $\mathbb{R}^{n\times k}$, as A, D are assumed to be positive definite. If A, D were only positive semidefinite then (2.3) would define a sub-Riemannian metric. In order to avoid technical difficulties we therefore assume positive definiteness of A and D.

Recall, that for any Riemannian metric $\langle\, , \,\rangle$ on $\mathbb{R}^{n\times k}$ an associated gradient vector field $\operatorname{grad} f(X)$ is defined by the characterising property

$$\mathrm{df}(X)\Omega = \langle \operatorname{grad} f(X), \Omega\rangle \qquad \forall X, \Omega \in \mathbb{R}^{n\times k}.$$

We now compute the gradient of our cost function f with respect to the above Riemannian metric.

Proposition 2.2 *The gradient of the cost function* (2.2) *with respect to the Riemannian metric* (2.3) *is the real analytic vectorfield* $\operatorname{grad} f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$ *given as*

$$\operatorname{grad} f(X) = XDX^\top AX - AXD.$$

In particular, the Oja–Brockett flow (2.1) is just the negative gradient flow

$$\dot{X} = -\operatorname{grad} f(X)$$

of the cost function (2.2).

Choosing other Riemannian metrics on $\mathbb{R}^{n\times k}$, different gradients of the same function $f : \mathbb{R}^{n\times k} \to \mathbb{R}$ are obtained. Using the above proposition we can now prove our main convergence result for the Oja–Brockett flow.

Theorem 2.3 *(i) The solutions* $X(t) \in \mathbb{R}^{n\times k}$ *of the Oja–Brockett flow* (2.1) *exist for all* $t \geq 0$ *and every solution of* (2.1) *converges for* $t \to +\infty$ *to a single equilibrium point* $X_\infty = \lim_{t\to\infty} X(t)$.
(ii) The equilibrium points of (2.1) *are characterized by*

$$AX_\infty D = X_\infty D X_\infty^\top A X_\infty. \tag{2.4}$$

Moreover, every equilibrium point X_∞ *satisfies the additional matrix equation*

$$X_\infty^\top A X_\infty (X_\infty^\top X_\infty - I_k) = 0. \tag{2.5}$$

We now derive an explicit description of the equilibrium points of (2.1) in terms of the eigenspace decomposition of A, D. For simplicity we assume that D is in diagonal form, i.e.

$$D = \operatorname{diag}(\mu_1, \ldots, \mu_1, \ldots, \mu_s, \ldots, \mu_s) \tag{2.6}$$

with $\mu_1 > \ldots > \mu_s > 0$ and μ_i occuring with multiplicity k_i, $k_1 + \ldots + k_s = k$.

Theorem 2.4 *Let* $\Phi_0 \in \mathrm{SO}(n)$ *be a basis matrix of eigenvectors of* $A = A^\top > 0$ *and let* D *be a diagonal form* (2.6). *Then* $X_\infty \in \mathbb{R}^{n\times k}$ *is an equilibrium point of the Oja–Brockett flow* (2.1) *if and only if*

$$X_\infty = \Phi_0 \pi \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} PS \tag{2.7}$$

where $0 \leq r \leq k \leq n$, P, π *are* $k \times k$ *and* $n \times n$ *permutation matrices, respectively, and* $S \in \mathrm{O}(k)$ *with* $S = \mathrm{diag}(S_1, \ldots, S_s)$, $S_i \in \mathrm{O}(k_i)$. *Equivalently,*

$$X_\infty = \Phi_0 \pi \begin{bmatrix} \Gamma \\ 0 \end{bmatrix} S \tag{2.8}$$

where π *is an* $n \times n$ *permutation matrix,* $\Gamma = \mathrm{diag}(\epsilon_1, \ldots, \epsilon_k)$, $\epsilon_i \in \{0, 1\}$ *and* $S = \mathrm{diag}(S_1, \ldots, S_s)$, $S_i \in \mathrm{O}(k_i)$, $i = 1, \ldots, s$.

Remark 2.5 Formula (2.8) in the above theorem generalizes a result of Xu [14], who has derived the equivalent representation

$$X_\infty = \Psi_0 \pi \begin{bmatrix} S \\ 0 \end{bmatrix},$$

where $S = \mathrm{diag}(\delta_1, \ldots, \delta_k)$, $\delta_i \in \{0, 1, -1\}$, $i = 1, \ldots, k$, for the generic case, where $D = \mathrm{diag}(\mu_1, \ldots, \mu_k)$, $\mu_1 > \ldots > \mu_k > 0$. Our representation applies also in the nongeneric case, and thus in particular to Oja's flow where $D = I_k$.

3 The Weighted Subspace Flow

Another flow that achieves PCA is the *weighted subspace flow* introduced by Oja et al [9, 10]. This is given as the matrix differential equation on $\mathbb{R}^{n \times k}$

$$\dot{X} = AX - XX^\top AXD \tag{3.1}$$

where $A = A^\top > 0$ and $D = D^\top > 0$. (Actually Oja et al only consider the case where $D = \mathrm{diag}(\mu_1, \ldots, \mu_k)$, $\mu_1 > \ldots > \mu_k > 0$.) We show that (3.1) admits a similar convergence analysis as (2.1). We just state the relevant results; the proofs run completely similar to those for (2.1) and are therefore omitted.

Proposition 3.1 *The gradient of the function* $g : \mathbb{R}^{n \times k} \to \mathbb{R}$

$$g(X) := \frac{1}{4} \mathrm{tr}(X^\top AX)^2 - \frac{1}{2} \mathrm{tr}(A^2 X D^{-1} X^\top) \tag{3.2}$$

with respect to the Riemannian metric (2.3) *is the real analytic vector field* $\mathrm{grad}\, g : \mathbb{R}^{n \times k} \to \mathbb{R}^{n \times k}$ *given as*

$$\mathrm{grad}\, g(X) = XX^\top AXD - AX.$$

In particular, the weighted subspace flow (3.1) *is the negative gradient flow* $\dot{X} = -\,\mathrm{grad}\, g(X)$ *of* (2.2).

Theorem 3.2 *(i) The cost function* (3.2) *on* $\mathbb{R}^{n \times k}$ *has compact sublevel sets and is lower bounded. The solutions* $X(t) \in \mathbb{R}^{n \times k}$ *of the weighted subspace flow* (3.1) *exist for all* $t \geq 0$ *and every solution* $X(t)$ *converges for* $t \to +\infty$ *to a single equilibrium point* $X_\infty = \lim_{t \to \infty} X(t)$.
(ii) Equilibrium points of (3.1) *are characterized by*

$$AX_\infty = X_\infty X_\infty^\top AX_\infty D.$$

There is a close connection between the solutions of the Oja–Brockett flow and the weighted subspace flow. This becomes particularly transparent if D is assumed to be diagonal

$$D = \operatorname{diag}(\mu_1, \ldots, \mu_k), \qquad \mu_1 > \ldots > \mu_k > 0. \tag{3.3}$$

One interesting additional feature of the subspace flow in contrast to (2.1) is that it defines a rank preserving flow.

Proposition 3.3 *Let $A = A^\top > 0$, $D = D^\top > 0$. Then the weighted subspace flow* (3.1) *is rank preserving, i.e. for every solution $X(t)$ of* (3.1) *we have*

$$\operatorname{rank} X(t) = \operatorname{rank} X(0) \quad \forall t \geq 0.$$

4 Hessians and Asymptotic Stability

Finally we establish the local stability properties of the Oja–Brockett flow. This immediately implies corresponding results for the weighted subspace flow (3.1), which are however not explicitly stated here. Let

$$X_\infty = \Psi_0 \pi \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} PS \in \mathbb{R}^{n \times k}$$

be an equilibrium point of (2.1). We calculate the signature of the Hessian of the cost function $f : \mathbb{R}^{n \times k} \to \mathbb{R}$, (2.2), at X_∞ and characterize the asymptotically stable equilibria of (2.1).

Theorem 4.1 *Let $D = \operatorname{diag}(\mu_1, \ldots, \mu_k)$ with $\mu_1 > \ldots > \mu_k > 0$ and let $A = A^\top > 0$ have distinct eigenvalues $\lambda_1 > \ldots > \lambda_n > 0$.*
(i) Every critical point of $f : \mathbb{R}^{n \times k} \to \mathbb{R}$ is nondegenerate.
(ii) The signature of the Hessian at the equilibrium points $X_\infty = \Psi_0 \pi \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} PS$ is given by

$$\begin{aligned} \nu_+ &:= \dim \operatorname{Eig}_+(\operatorname{Hess} f)(X_\infty) &&= nk - (n-r)(k-r) - s, \\ \nu_- &:= \dim \operatorname{Eig}_-(\operatorname{Hess} f)(X_\infty) &&= (n-r)(k-r) + s \end{aligned}$$

where ν_+ and ν_- are the numbers of positive and negative eigenvalues of the Hessian, respectively, and

$$s := \operatorname{card}\Big\{(i,j); \quad 1 \leq i < j \leq r \ \text{ and } \ (\pi^\top(i) - \pi^\top(j))(p(i) - p(j)) < 0\Big\}$$

(iii) In particular, the asymptotically stable equilibria are precisely those of the form

$$X_\infty = \Psi_0 \begin{bmatrix} I_k \\ 0 \end{bmatrix} S$$

with $S = \operatorname{diag}(\epsilon_1, \ldots, \epsilon_k)$, $\epsilon_i \in \{-1, 1\}$.

References

[1] P. Baldi and K. Hornik. "Back-propagation and unsupervised learning in linear networks", *in Backpropagation: Theory, architectures and applications* (Y. Chauvin and D. E. Rumelhart, Eds.), Hillsdale, NJ: Erlbaum Associates, 1991.

[2] P. Baldi and K. Hornik. "Learning in linear neural networks: A survey", *IEEE Trans. Neural Networks* 6 (1995), 837–858.

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

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

[5] S. Lojasiewicz. "Sur les trajectoires du gradient d'une fonction analytique", *Seminari di Geometria*, Bologna (1983), 115–117.

[6] E. Oja. "A simplified neuron model as a principal component analyzer", *J. Math. Biol.* 15 (1982), 267–273.

[7] E. Oja and 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.

[8] E. Oja. "Neural networks, principal components, and subspaces", *Int. J. Neural Systems* 1 (1989), 61–68.

[9] E. Oja, H. Ogawa and J. Wangviwattana. "Principal component analysis by homogeneous neural networks, Part I: The weighted subspace criterion", *IEICE Trans. Inf. & Syst.* 3 (1992), 366–375.

[10] E. Oja, H. Ogawa and J. Wangviwattana. "Principal component analysis by homogeneous neural networks, Part II: Analysis and extensions of the learning algorithms", *IEICE Trans. Inf. & Syst.* 3 (1992), 376–382.

[11] T.D. Sanger. "Optimal unsupervised learning in a single-layer linear feedforward network", *Neural Networks* 2 (1989), 459–473.

[12] R. Williams. "Feature discovery through error-correcting learning", *Techn. Report* 8501, UCSD, Institute of Cognitive Science, (1985).

[13] J.L. Wyatt and I.M. Elfadel. "Time–domain solutions of Oja's equations", *Neural Computation* 7 (1995), 915–922.

[14] L. Xu. "Least mean square error recognition principle for self organizing neural nets", *Neural Networks* 6 (1993), 627–648.

[15] W.Y. Yan, U. Helmke and J.B. Moore. "Global analysis of Oja's flow for neural networks", *IEEE Trans. Neural Networks* 5 (1994), 674–683.