

ASYMPTOTIC PERFORMANCE ANALYSIS OF PCA ALGORITHMS BASED ON THE WEIGHTED SUBSPACE CRITERION

Jean Pierre Delmas, Victor Gabillon

TELECOM & Management SudParis, 91011 Evry, France
Département CITI, CNRS UMR 5157

ABSTRACT

This paper studies the asymptotic distribution of the eigenvectors estimated by some PCA algorithms based on the weighted subspace criterion. This enables us to analyse how the choice of the weighting matrix affects the algorithm's performance, an issue previously overlooked.

Index Terms— Principal component analysis, weighted subspace criterion, asymptotic performance analysis.

1. INTRODUCTION

Research in subspace and component-based techniques was initiated in the Statistics literature in the middle of the last century through the problem of linear feature extraction solved by the Karhunen-Loève Transform. However the interest of the signal processing community in adaptive subspace and component-based schemes remains strong as it is evident from the numerous articles and reports published in this area each year. In contrast, only a few contributions have been dedicated to the theoretical performance analysis of these adaptive algorithms.

In particular, stochastic approximation gradient-like algorithms devoted to principal and minor subspace analysis (PSA and MSA) or principal and minor component analysis (PCA and MCA) have been mainly studied in stationary environments from two points of view. The first is their convergence in a decreasing step-size situation, using the stability theory of an associated ordinary differential equation (ODE) (see e.g., [1], [2] and the references therein). In a constant step-size situation, it has been shown [3] that the sequence of estimates can be approximated by the associated ODE, in the sense of weak convergence of random processes as the step size tends to zero. The second concerns the asymptotic Gaussian case when both the step size tends to zero and the number of samples tend to ∞ and has been derived for only a few algorithms (see e.g., [4], [5], [6]).

The purpose of this paper is to use the approach developed in [5], to derive the asymptotic distribution of the estimate issued from a family of adaptive PCA and MCA algorithms derived from the weighted criterion introduced in [7]. One of these gradient-descent like algorithms has just been

the subject of a new fast orthogonalization step [8]. From the estimation accuracy point of view, *the choice of the weighting matrix crucially affects the algorithm's performance* as been shown by simulations in [8]. Hence, it would be of interest to analyse the effect of this weighting matrix on the algorithm's performance and derive, if possible, its optimal value in some sense. This point will be an application of our analysis.

This paper is organized as follows. In Section 2, we give an overview of the family of adaptive PCA and MCA algorithms derived from the weighted criterion introduced in [7]. Then, in Section 3, after presenting a brief review of a general Gaussian approximation result, closed-form expressions of the covariance in the limiting distributions of the eigenvector estimators in a constant step size environment are given by solving Lyapunov equations. Finally, we present briefly in Section 4 some simulation results to illustrate the optimal choice of the weighting matrix.

2. REVIEW OF THE ALGORITHMS UNDER STUDY

For a given $n \times n$ covariance matrix $\mathbf{R}_x = E(\mathbf{x}\mathbf{x}^T)$ of a Gaussian distributed, zero mean real random vector $\mathbf{x}$, denote by $\lambda_1 \geq \dots \geq \lambda_n$ the eigenvalues of $\mathbf{R}_x$ and by $\mathbf{v}_1, \dots, \mathbf{v}_n$ corresponding normalized eigenvectors. Consider the problem of adaptively estimating $r < n$ normalized eigenvectors corresponding to the r largest [or smallest] distinct eigenvalues ($\lambda_1, \dots, \lambda_r$) with $\lambda_r > \lambda_{r+1}$ [resp. $\lambda_{n-r+1}, \dots, \lambda_n$ with $\lambda_{n-r} > \lambda_{n-r+1}$] of $\mathbf{R}_x$.

The PCA and MCA stochastic approximation algorithms that are studied in this paper, are derived from a small change in the subspace criterion [9] that reduces its ambiguous solution to a unique set of eigenvectors. If

$$J(\mathbf{W}) \stackrel{\text{def}}{=} \text{Tr}(\mathbf{\Omega} \mathbf{W}^T \mathbf{R}_x \mathbf{W}) = \sum_{i=1}^r \omega_i \mathbf{w}_i^T \mathbf{R}_x \mathbf{w}_i \quad (1)$$

where $\mathbf{W} \stackrel{\text{def}}{=} [\mathbf{w}_1, \dots, \mathbf{w}_r] \in \mathcal{R}^{n \times r}$ is an orthogonal matrix and $\mathbf{\Omega} \stackrel{\text{def}}{=} \text{Diag}(\omega_1, \dots, \omega_r)$ with $\omega_1 > \dots > \omega_r > 0$, we have the following lemma introduced for PCA in [7], proved in [11], then recently in [8] with a unifying proof for PCA and MCA.

Lemma 1 *The maximization or minimization of $J(\mathbf{W})$ (1) has the unique solution $\{\pm \mathbf{v}_1, \dots, \pm \mathbf{v}_r\}$ or $\{\pm \mathbf{v}_{n-r+1}, \dots, \pm \mathbf{v}_n\}$ respectively. All the stationary points of $J(\mathbf{W})$ are saddle points, except the global maximum and global minimum solutions specified above.*

The maximization or minimization of the *weighted subspace criterion* $J(\mathbf{W})$ subject to the constraint $\mathbf{W}^T \mathbf{W} = \mathbf{I}_r$ can be adaptively solved by constrained gradient-ascent/descent techniques using the instantaneous estimate $\mathbf{x}_k \mathbf{x}_k^T$ of $\mathbf{R}_x$.

Since the gradient function of $J(\cdot)$ under orthonormality constraint (see e.g., [2, Ch.7.2]) is equal to $\mathbf{R}_x \mathbf{W} \Omega - \mathbf{W} \Omega \mathbf{W}^T \mathbf{R}_x \mathbf{W}$, the PCA and MCA stochastic approximation algorithms, that we denote here by *weighted Oja algorithm 1* (WOja₁), follow clearly

$$\mathbf{W}_{k+1} = \mathbf{W}_k \pm \mu_k (\mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k \Omega - \mathbf{W}_k \Omega \mathbf{W}_k^T \mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k). \quad (2)$$

Another approach consists in using an unconstrained gradient ascent/descent algorithm followed by an orthonormalization step. Using the gradient function $\mathbf{R}_x \mathbf{W} \Omega$ of $J(\cdot)$ without constraint, this gives

$$\mathbf{W}_{k+1} = \{\mathbf{W}_k \pm \mu_k \mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k \Omega\} \mathbf{G}_{k+1}, \quad (3)$$

in which $\mathbf{G}_{k+1}$ is a matrix depending on $\mathbf{W}'_{k+1} \stackrel{\text{def}}{=} \mathbf{W}_k \pm \mu_k \mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k \Omega$, which approximately orthonormalizes the columns of $\mathbf{W}'_{k+1}$. Thus, $\mathbf{W}_k$ has approximately orthonormal columns for all k . Depending on the form of matrix $\mathbf{G}_{k+1}$, variants of (3) are obtained. If $\mathbf{G}_{k+1}$ is the symmetric square root inverse of $\mathbf{W}'_{k+1}{}^T \mathbf{W}'_{k+1}$, we obtain [10, ch.4] after a power expansion in μ_k where all but the zero and first order terms are omitted, the following stochastic approximation algorithm denoted here by *weighted Oja algorithm 2* (WOja₂)

$$\begin{aligned} \mathbf{W}_{k+1} = & \mathbf{W}_k \pm \mu_k [\mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k \Omega \\ & - \frac{1}{2} \mathbf{W}_k \Omega \mathbf{W}_k^T \mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k - \frac{1}{2} \mathbf{W}_k \mathbf{W}_k^T \mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k \Omega]. \end{aligned} \quad (4)$$

Instead of deriving a stochastic approximation algorithm from a specific orthonormalization matrix $\mathbf{G}_{k+1}$, an analogy with Oja's algorithm [1]

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu_k [\mathbf{I}_n - \mathbf{W}_k \mathbf{W}_k^T] \mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k$$

has been used in [7] to derive the following algorithm denoted here by *weighted Oja algorithm 3* (WOja₃)

$$\mathbf{W}_{k+1} = \mathbf{W}_k \pm \mu_k [\mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k - \mathbf{W}_k \Omega \mathbf{W}_k^T \mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k \Omega^{-1}].$$

Using $\mathbf{y}_k \stackrel{\text{def}}{=} \mathbf{W}_k^T \mathbf{x}_k$, $\mathbf{y}_{\Omega, k} \stackrel{\text{def}}{=} \Omega \mathbf{y}_k$ and $\mathbf{z}_k \stackrel{\text{def}}{=} \mathbf{W}_k \mathbf{y}_{\Omega, k}$, we note that WOja₁, WOja₂ and WOja₃ algorithms require only $O(nr)$ operations at each update.

To the best of our knowledge, no complete theoretical performance analysis of these algorithms has been carried out until now. However, the eigenvectors $[\pm \mathbf{v}_1, \dots, \pm \mathbf{v}_r]$, have

been proved to be locally asymptotically stable points for the associated ODE corresponding to the WOja₁ [11] and WOja₃ [12] PCA algorithms. As for the MCA algorithms, they have been considered as unstable by simulations [8]. Techniques to stabilize these MCA algorithms have been proposed in [8] consisting to an additional step of exact orthonormalization. The denoted MCA-OFRANS algorithms has been proposed where the matrix $\mathbf{G}_{k+1}$ of (3) is the exact symmetric square root inverse of $\mathbf{W}'_{k+1}{}^T \mathbf{W}'_{k+1}$. Applied to $\mathbf{W}'_{k+1} \stackrel{\text{def}}{=} \mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k \Omega - \mathbf{W}_k \Omega \mathbf{W}_k^T \mathbf{x}_k \mathbf{x}_k^T \mathbf{W}_k$ of (2), the denoted MCA-OOja algorithm has been derived. However these two MCA algorithms remain sensitive to numerical rounding errors that are eliminated by using an equivalent Householder transform implementation of this last step of exact orthonormalization.

3. ASYMPTOTICAL PERFORMANCE ANALYSIS

3.1. A short review of a general Gaussian approximation result

In this section, we evaluate the asymptotic distributions of eigenvector estimators given by the previous algorithms. For this purpose, we shall use the following result [13, Th.2, p.108]. Consider a constant step size recursive stochastic algorithm

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k + \mu f(\boldsymbol{\theta}_k, \mathbf{x}_k) \quad (5)$$

with $\mathbf{x}_k = g(\xi_k)$, where ξ_k is a Markov chain independent of $\boldsymbol{\theta}_k$. Suppose that the parameter vector $\boldsymbol{\theta}_k$ converges almost surely to the unique asymptotically stable point $\boldsymbol{\theta}_*$ in the corresponding decreasing step size algorithm. Consider the continuous Lyapunov equation

$$\mathbf{D} \mathbf{C}_\theta + \mathbf{C}_\theta \mathbf{D}^T + \mathbf{G} = \mathbf{O} \quad (6)$$

where $\mathbf{D}$ and $\mathbf{G}$ are, respectively, the derivative of the mean field and the covariance of the field of algorithm (5)

$$\mathbf{D} \stackrel{\text{def}}{=} \mathbb{E} \left[\frac{\partial f}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}, \mathbf{x}_k) \right]_{\boldsymbol{\theta}=\boldsymbol{\theta}_*},$$

$$\mathbf{G} \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} \text{Cov}[f(\boldsymbol{\theta}_*, \mathbf{x}_k), f(\boldsymbol{\theta}_*, \mathbf{x}_0)].$$

If all the eigenvalues of the derivative $\mathbf{D}$ of the mean field have strictly negative real parts, then, in a stationary situation, when $\mu \rightarrow 0$ and $k \rightarrow \infty$, we have the convergence in distribution

$$\mu^{-1/2}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_*) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{C}_\theta), \quad (7)$$

where $\mathbf{C}_\theta$ is the unique symmetric solution of the Lyapunov equation (6).

3.2. Asymptotic distributions of eigenvector estimators

To characterize the derivative of the mean field and the covariance of the field of the WOja₁, WOja₂ and WOja₃ algorithms, we use the Vec operator which turns the $n \times r$ matrix

$\mathbf{W}$ into the $nr \times 1$ vector parameter $\text{vec}(\mathbf{W})$. Thus these three PCA/MCA algorithms with constant step size are written in the form

$$\text{vec}(\mathbf{W}_{k+1}) = \text{vec}(\mathbf{W}_k) \pm \mu f(\text{vec}(\mathbf{W}_k, \mathbf{x}_k \mathbf{x}_k^T)) \quad (8)$$

3.2.1. Local characterization of the field

Derivative of the field: Using a first order expansion of the field $f(.,.)$ with respect to its first argument, we obtain after some algebraic manipulations the following derivatives associated with the three PCA algorithms

$$\begin{aligned} \mathbf{D}_1 &= \mathbf{\Omega} \otimes \mathbf{R}_x - \mathbf{\Lambda} \mathbf{\Omega} \otimes \mathbf{I}_n - \mathbf{I}_r \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T) \\ &\quad - [\mathbf{\Lambda} \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Omega}] \mathbf{K}_{n,r} \\ \mathbf{D}_2 &= \mathbf{\Omega} \otimes \mathbf{R}_x - \mathbf{\Lambda} \mathbf{\Omega} \otimes \mathbf{I}_n \\ &\quad - \frac{1}{2} [\mathbf{I}_r \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T) + \mathbf{\Omega} \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{W}^T)] \\ &\quad - \frac{1}{2} [(\mathbf{\Lambda} \mathbf{W}^T) \otimes \mathbf{W} \mathbf{\Omega} + (\mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T \otimes \mathbf{W})] \mathbf{K}_{n,r} \\ \mathbf{D}_3 &= \mathbf{I}_r \otimes \mathbf{R}_x - \mathbf{\Lambda} \otimes \mathbf{I}_n - \mathbf{\Omega}^{-1} \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T) \\ &\quad - [\mathbf{\Lambda} \mathbf{\Omega}^{-1} \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Omega}] \mathbf{K}_{n,r}, \end{aligned}$$

where $\mathbf{W} \stackrel{\text{def}}{=} [\pm \mathbf{v}_1, \dots, \pm \mathbf{v}_r]$, $\mathbf{\Lambda} \stackrel{\text{def}}{=} \text{Diag}(\lambda_1, \dots, \lambda_r)$ and $\mathbf{K}_{l,m}$ denotes the vec-permutation matrix that transforms $\text{vec}(\mathbf{A}^T)$ to $\text{vec}(\mathbf{A})$ for any $l \times m$ matrices.

From these expressions, and from those associated with the MCA algorithms, the following theorem is proved in [15]

Theorem 1 *The eigenvalues of the derivative $\mathbf{D}$ of the mean field of the WOja₁ PCA algorithm are strictly negative real and those of the WOja₂ and WOja₃ PCA algorithms have strictly negative real parts. In contrast, there exist eigenvalues with positive real parts for all the three MCA algorithms.*

Covariance of the field: Noting that field (8) is linear in $\mathbf{x}_k \mathbf{x}_k^T$ and using the covariance of $\text{vec}(\mathbf{x}_k \mathbf{x}_k^T)$ given by [14, p. 57] $\text{Cov}(\text{vec}(\mathbf{x} \mathbf{x}^T)) = \mathbf{R}_x \otimes \mathbf{R}_x + (\mathbf{R}_x \otimes \mathbf{R}_x) \mathbf{K}_{n,n}$, for independent Gaussian distributed data $\mathbf{x}_k$, allow us to derive, after simple algebraic manipulations, the following expressions of the covariance of the field

$$\begin{aligned} \mathbf{G}_1 &= \mathbf{\Lambda} \mathbf{\Omega}^2 \otimes \mathbf{R}_x \\ &\quad - 2 \mathbf{\Lambda} \mathbf{\Omega} \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T) + \mathbf{\Lambda} \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{\Omega}^2 \mathbf{W}^T) \\ &\quad + [2 \mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Lambda} \mathbf{\Omega} - \mathbf{\Lambda} \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Lambda} \mathbf{\Omega}^2 \\ &\quad \quad - \mathbf{\Lambda} \mathbf{\Omega}^2 \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Lambda}] \mathbf{K}_{n,r} \\ \mathbf{G}_2 &= \mathbf{\Lambda} \mathbf{\Omega}^2 \otimes \mathbf{R}_x - \frac{1}{2} \mathbf{\Lambda} \mathbf{\Omega} \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T) \\ &\quad - \frac{3}{4} \mathbf{\Lambda} \mathbf{\Omega}^2 \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{W}^T) + \frac{1}{4} \mathbf{\Lambda} \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{\Omega}^2 \mathbf{W}^T) \\ &\quad + [\frac{1}{2} \mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Lambda} \mathbf{\Omega} - \frac{1}{4} \mathbf{\Lambda} \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Lambda} \mathbf{\Omega}^2 \\ &\quad \quad - \frac{1}{4} \mathbf{\Lambda} \mathbf{\Omega}^2 \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Lambda}] \mathbf{K}_{n,r} \\ \mathbf{G}_3 &= \mathbf{\Lambda} \otimes \mathbf{R}_x \\ &\quad - 2 \mathbf{\Lambda} \mathbf{\Omega}^{-1} \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T) + \mathbf{\Lambda} \mathbf{\Omega}^{-2} \otimes (\mathbf{W} \mathbf{\Lambda} \mathbf{\Omega}^2 \mathbf{W}^T) \end{aligned}$$

$$+ [2 \mathbf{\Lambda} \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Lambda} - \mathbf{\Lambda} \mathbf{\Omega} \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Lambda} \mathbf{\Omega}^{-1} - \mathbf{\Lambda} \mathbf{\Omega}^{-1} \mathbf{W}^T \otimes \mathbf{W} \mathbf{\Lambda} \mathbf{\Omega}] \mathbf{K}_{n,r}.$$

3.2.2. Solution of the Lyapunov equation

For independent observations $\mathbf{x}_k$ and for the investigated algorithms, which can be written in a form similar to (5) for which the derivative of the mean field have strictly negative real parts (Theorem 1), the hypotheses of the model of Benveniste *et al* ([13, Th.2, p.108]) are fulfilled. However, the underlying assumption for the results by Benveniste *et al* is that the solution of the corresponding stochastic approximation type algorithms with decreasing step size, almost surely converges to the unique asymptotically stable solution of the associated ODE. Since the normalized eigenvectors are defined up to a sign, the global attractor $\mathbf{W}$ is not unique. Still, the practical use of the Benveniste results in such situation is usually justified by using formally a general approximation result ([13, Th.1, p.107]). Furthermore, the almost sure convergence of the associated decreasing step size algorithms are not strictly fulfilled for these WOja PCA algorithms. This a.s. convergence would need a boundedness condition, whose satisfaction is a challenging problem as discussed in [5]

If we allow ourselves the Benveniste results in our situation, the Lyapunov continuous equations can be solved exactly. Using the approach of [5] that consists in representing the matrices $\mathbf{D}$ and $\mathbf{G}$ in the orthonormal basis constituted by the nr columns of the r block diagonal matrix $\text{Diag}(\mathbf{V}, \dots, \mathbf{V})$ with $\mathbf{V} \stackrel{\text{def}}{=} [\mathbf{v}_1, \dots, \mathbf{v}_n]$, the following closed-form expressions of the covariance matrix $\mathbf{C}_W$ solution of the Lyapunov equation (6) are derived [15].

Theorem 2 *The covariance matrices $\mathbf{C}_W$ of the asymptotic distribution that appears in (7) read*

$$\begin{aligned} \mathbf{C}_W &= \sum_{\substack{1 \leq i \leq r \\ 1 \leq k \neq i \leq n}} b_{k,i} (\mathbf{e}_i^r \mathbf{e}_i^{rT} \otimes \mathbf{v}_k \mathbf{v}_k^T) \\ &\quad + \sum_{1 \leq i \neq j \leq r} c_{i,j} (\mathbf{e}_i^r \mathbf{e}_j^{rT} \otimes \mathbf{v}_j \mathbf{v}_i^T) \end{aligned} \quad (9)$$

with for the WOja₁, WOja₂ and WOja₃ PCA algorithms respectively

$$\begin{aligned} b_{k,i}^{\text{WOja}_1} &= \frac{\lambda_i \lambda_k (\omega_i - \omega_k)}{2(\lambda_i - \lambda_k)} \mathbb{1}_{1 \leq k \neq i \leq r} + \frac{\lambda_i \lambda_k \omega_i}{2(\lambda_i - \lambda_k)} \mathbb{1}_{r < k} \\ c_{i,j}^{\text{WOja}_1} &= -\frac{\lambda_i \lambda_j (\omega_i - \omega_j)}{2(\lambda_i - \lambda_k)} \mathbb{1}_{1 \leq i \neq j \leq r} \\ b_{k,i}^{\text{WOja}_2} &= \frac{\lambda_i \lambda_k (\omega_i - \omega_k)}{4(\lambda_i - \lambda_k)} \mathbb{1}_{1 \leq k \neq i \leq r} + \frac{\lambda_i \lambda_k \omega_i}{2(\lambda_i - \lambda_k)} \mathbb{1}_{r < k} \\ c_{i,j}^{\text{WOja}_2} &= -\frac{\lambda_i \lambda_j (\omega_i - \omega_j)}{4(\lambda_i - \lambda_k)} \mathbb{1}_{1 \leq i \neq j \leq r} \\ b_{k,i}^{\text{WOja}_3} &= f_{k,i} \mathbb{1}_{1 \leq i \neq k \leq r} + \frac{\lambda_i \lambda_k}{2(\lambda_i - \lambda_k)} \mathbb{1}_{r < k} \\ c_{i,j}^{\text{WOja}_3} &= g_{i,j} \mathbb{1}_{1 \leq i \neq j \leq r} \end{aligned}$$

where $\mathbf{e}_i^T$ and $\mathbf{1}_A$ denote the i th unit vector in $\mathcal{R}^r$ and the indicator function of the condition A and where

$$f_{k,i} = \frac{\lambda_i \lambda_k \frac{\omega_k}{\omega_i} \left(\sqrt{\frac{\omega_k}{\omega_i}} - \sqrt{\frac{\omega_i}{\omega_k}} \right) \left((\lambda_i - \lambda_k) \sqrt{\frac{\omega_k}{\omega_i}} - \lambda_i \sqrt{\frac{\omega_i}{\omega_k}} (1 + \frac{\omega_i}{\omega_k}) \right)}{2(\lambda_i - \lambda_k) \left(\frac{\lambda_i \omega_i}{\omega_k} + \frac{\lambda_k \omega_k}{\omega_i} \right)}$$

$$g_{i,j} = \frac{\lambda_i \lambda_j \left(\sqrt{\frac{\omega_i}{\omega_j}} - \sqrt{\frac{\omega_j}{\omega_i}} \right) \left(\lambda_i (2\sqrt{\frac{\omega_i}{\omega_j}} - \sqrt{\frac{\omega_j}{\omega_i}}) + \lambda_j (2\sqrt{\frac{\omega_j}{\omega_i}} - \sqrt{\frac{\omega_i}{\omega_j}}) \right)}{2(\lambda_j - \lambda_i) \left(\frac{\lambda_j \omega_j}{\omega_i} + \frac{\lambda_i \omega_i}{\omega_j} \right)}$$

4. ILLUSTRATIVE EXAMPLE

Simple measures of performance of the studied PCA algorithms can be derived from Theorem 2 if we suppose that both the first and second moments of the limiting distribution of $\mu^{-1/2}(\mathbf{W}_k - \mathbf{W})$ are equal to the corresponding asymptotic moments. Among them we have in the steady state, $\|\mathbf{E}(\mathbf{W}_k) - \mathbf{W}\|_{\text{Fro}}^2 = o(\mu)$, $\text{Cov}(\text{vec}(\mathbf{W}_k)) = \mu \mathbf{C}_W + o(\mu)$ which implies

$$\mathbf{E}\|\mathbf{W}_k - \mathbf{W}\|_{\text{Fro}}^2 = \mu \text{Tr}(\mathbf{C}_W) + o(\mu) \approx \mu \sum_{1 \leq i \leq r} b_{k,i},$$

$$\mathbf{E}\|\mathbf{w}_{k,i} - \mathbf{v}_i\|_{\text{Fro}}^2 = \mu \text{Tr}([\mathbf{C}_W]_{i,i}) + o(\mu) \approx \mu \sum_{1 \leq k \neq i \leq n} b_{k,i}$$

and [15] $\mathbf{E}\|\mathbf{W}_k^T \mathbf{W}_k - \mathbf{I}_r\|_{\text{Fro}}^2 = o(\mu)$.

We note that from the expressions of $b_{k,i}$ given by Theorem 2, the MSE of each estimated eigenvector $\mathbf{w}_{k,1}, \dots, \mathbf{w}_{k,r}$ can be arbitrarily fixed by an appropriate choice of $(\mu, \omega_1, \dots, \omega_r)$.

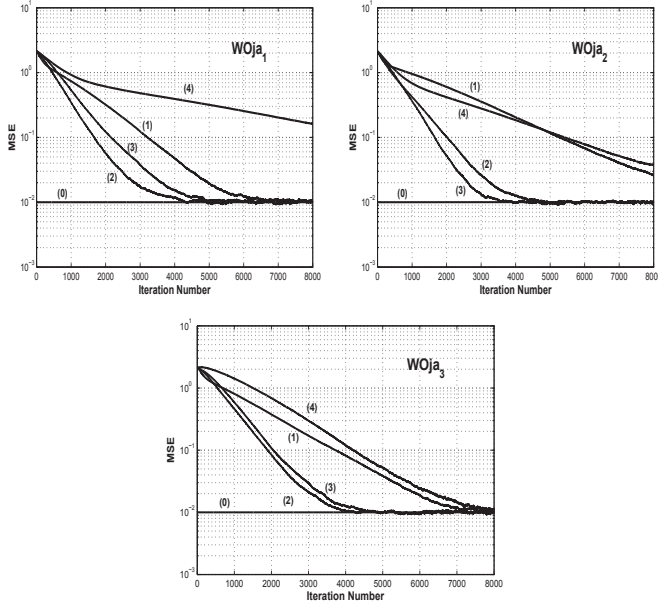


Fig.1 Learning curves of the MSE $\mathbf{E}\|\mathbf{W}_k - \mathbf{W}\|_{\text{Fro}}^2$ averaging 1000 independent runs for the 3 PCA WOja algorithms compared to $\mu \text{Tr}(\mathbf{C}_W)$ (0) for different values 0.9 (1), 0.6 (2), 0.4 (3), 0.1 (4) of ω_1/ω_2 for $n = 3, p = 2$ and $(\lambda_1, \lambda_2, \lambda_3) = (1, 0.6, 0.2)$.

Concerning the minimization of the misadjustment $\mathbf{E}\|\mathbf{W}_k - \mathbf{W}\|_{\text{Fro}}^2$, we note that it vanishes for $\omega_1 = \dots = \omega_r$ for which the speed of convergence tends to ∞ . Consequently a misadjustment/convergence speed tradeoff crops up for the choice of the "optimal" weighting matrix. To solve this dilemma, a solution consists in fixing the theoretical value $\mu \text{Tr}(\mathbf{C}_W)$ and searching for the value of $(\omega_1, \dots, \omega_r)$ that maximizes

the speed of convergence. Noting that the set of parameters $(\mu, \omega_1, \dots, \omega_r) = (\mu/c, c\omega_1, \dots, c\omega_r)$ [resp. $(\mu, \omega_1, \dots, \omega_r) = (\mu, c\omega_1, \dots, c\omega_r)$] (for arbitrary value of $c > 0$) yields the same WOja₁ and WOja₂ algorithms [resp. WOja₃ algorithm], this speed of convergence depends only on the ratios $\omega_2/\omega_1, \dots, \omega_r/\omega_{r-1}$. To illustrate this choice, Fig.1 displays the similar behavior of the three studied PCA WOja algorithms in a particular scenario. A thorough analysis in relation with $(\lambda_1, \dots, \lambda_r)$ will be presented in [15].

5. REFERENCES

- [1] E. Oja, "Principal components, minor components and linear neural networks," *Neural Networks*, vol. 5, pp. 927-935, 1992.
- [2] J. Dehaene, *Continuous-time matrix algorithms, systolic algorithms and adaptive neural networks*, Ph.D. dissertation, Katholieke Univ. Leuven, Belgium, Oct. 1995.
- [3] C.M. Kuan, K. Hornik, "Convergence of learning algorithms with constant learning rates," *IEEE Trans. Neural Networks*, vol. 2, no. 5, pp. 484-489, Sept. 1991.
- [4] B. Yang, F. Gersemsky, "Asymptotic distribution of recursive subspace estimators," *Proc. ICASSP Atlanta, GA*, May 1996, pp. 1764-1767.
- [5] J.P. Delmas, F. Alberge, "Asymptotic performance analysis of subspace adaptive algorithms introduced in the neural network literature," *IEEE Trans. on Signal Process.*, vol. 46, no. 1, pp. 170-182, January 1998.
- [6] J.P. Delmas, J.F. Cardoso, "Performance analysis of an adaptive algorithm for tracking dominant subspace," *IEEE Trans. on Signal Process.*, vol. 46, no. 11, pp. 3045-3057, November 1998.
- [7] E. Oja, H. Ogawa and J. Wangviwattana, "Principal component analysis by homogeneous neural networks, Part I: The weighted subspace criterion," *IEICE Trans. Inform. and Syst.*, vol.E75-D, pp. 366-375, 1992.
- [8] S. Bartelmaos, K. Abed-Meraim, "Fast adaptive algorithms for minor component analysis using Householder transformation," submitted to *Digital Signal Processing*, 2008.
- [9] J.F. Yang, M. Kaveh, "Adaptive eigensubspace algorithms for direction or frequency estimation and tracking," *IEEE Trans. on ASSP*, vol. 36, no. 2, pp. 241-251, February 1988.
- [10] A. Cichocki, S. Amari, *Adaptive blind signal and image processing, learning algorithms and applications*, John Wiley and Sons, 2002.
- [11] L. Xu, "Least mean square error reconstruction principle for self organizing neural net," *Neural Networks*, vol. 6, no.3, pp. 627-448, 1993.
- [12] 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. Inform. Syst.*, vol.E75-D, pp. 376-382, 1992.
- [13] A. Benveniste, M. Métivier and P. Priouret, *Adaptive algorithms and stochastic approximations*, New York: Springer Verlag, 1990.
- [14] T.W. Anderson, *An introduction to multivariate statistical analysis*. 2nd ed., New York: Wiley, 1984.
- [15] J.P. Delmas, V. Gabillon, "Asymptotic performance analysis of PCA algorithms issued from the weighted subspace criterion," in preparation to *IEEE Trans. on Signal Process.* January 2009.