Performances analysis of a Givens parametrized adaptive eigenspace algorithm

Jean-Pierre Delmas*

Institut National des Télécommunications, Département Signal et Image, 9 rue Charles Fourier, 91011 Evry Cedex, France

Abstract

In this paper, we address an adaptive estimation method for eigenspaces of covariance matrices. We are interested in a gradient procedure based on coupled maximizations or minimizations of Rayleigh quotients where the constraints are replaced by a Givens parametrization. This enables us to provide a canonic orthonormal eigenbasis estimator. We study the convergence of this algorithm with the help of the associated ordinary differential equation (ODE), and propose a performance evaluation by computing the variances of the estimated eigenvectors and of the estimated projection matrices on eigenspaces for fixed gain factors. In particular, we show that these misadjustments depend on whether the successive analyzed vector signals are correlated or not, and thus greatly depend on the origin of the covariance matrices of interest (spatial, temporal, spatio-temporal). More precisely, we show that these misadjustments can be smaller in the case of correlated observations than in the case of independent observations. Finally, we show that performance can be improved when the symmetric-centrosymmetric property of some of those covariance matrices is exploited. © 1998 Elsevier Science B.V. All rights reserved.

Zusammenfassung

In diesem Artikel wird eine adaptive Methode zur Schätzung von Eigenräumen von Kovarianzmatrizen behandelt. Wir interessieren uns für ein Gradientenverfahren, welches auf gekoppelten Maximierungen oder Minimierungen von Rayleigh-Quotienten beruht, wobei die Nebenbedingungen durch eine Givens-Parametrisierung ersetzt werden. Dies ermöglicht es, einen kanonischen Schätzer für orthonormale Eigenbasen anzugeben. Wir studieren die Konvergenz dieses Algorithmus mit Hilfe der zugehörigen gewöhnlichen Differentialgleichung. Zur Beurteilung der Leistungsfähigkeit schlagen wir vor, die Varianzen der geschätzten Eigenvektoren und der geschätzten Eigenraum-Projektionsmatrizen bei festen Verstärkungsfaktoren zu berechnen. Insbesondere zeigen wir, daß diese Fehlanpassungen davon abhängen, ob die nacheinander analysierten Vektorsignale korreliert sind, wodurch sich eine starke Abhängigkeit von der Herkunft der Kovarianzmatrizen (räumlich, zeitlich, räumlich-zeitlich) ergibt. Wir zeigen konkret, daß diese Fehlanpassungen im Fall korrelierter Beobachtungen kleiner sein können als im Fall unabhängiger Beobachtungen. Schließlich zeigen wir, daß die Leistungsfähigkeit erhöht werden kann, wenn die Symmetrie-Zentrosymmetrie-Eigenschaft einiger dieser Kovarianzmatrizen ausgenützt wird. © 1998 Elsevier Science B.V. All rights reserved.

* Tel.: + 33 1 60 76 46 32; fax: + 33 1 60 76 42 84; e-mail: delmas@int-evry.fr.

Résumé

Nous considérons dans cet article une méthode d'estimation de sous espaces propres de matrices de covariance. Nous nous intéressons à une méthode de gradient basée sur des minimisations ou des maximisations de quotients de Rayleigh dans lesquelles les contraintes sont remplacées par une paramétrisation de Givens. Cela permet de fournir de façon structurelle un estimateur orthonormé de bases orthonormées. Nous étudions la convergence de cet algorithme grâce à l'étude de son équation différentielle associée (ODE), et nous proposons une évaluation des performances par le calcul des variances des vecteurs propres et des matrices de projection associées estimées pour l'algorithme à pas fixe. Nous montrons, en particulier, que les erreurs quadratiques moyennes sont très sensibles à la corrélation des observations successives entre elles, donc dépendent fortement de l'origine des signaux observés (cas spatial, temporel ou spatio-temporel). De façon plus précise, nous montrons que ses erreurs quadratiques moyennes peuvent être plus petites dans le cas d'observations corrélées que dans le cas d'observations indépendantes. Nous montrons finalement que les performances peuvent être améliorées quand la structure symétrique-centrosymétrique de certaines matrices de covariance est prise en compte. © 1998 Elsevier Science B.V. All rights reserved.

1. Introduction

Over the past decade, adaptive estimation of subspaces of covariance matrices has been applied successfully to both temporal and spatial domain high-resolution spectral analysis. The interest for these methods, as a tool of outstanding importance in many fields of signal processing, has recently been renewed by the subspace approach used in blind identification of multi-channel finite impulse response filters [14]. Numerous solutions have been proposed to recursively update the eigendecomposition of a covariance matrix. Most of them can be gathered into five families. In the first one, classical batch eigendecomposition or singular value decomposition methods like the QR algorithms, Jacobi rotation methods, power iteration methods have been rendered adaptive. In the second family, variations of Bunch's rank-one updating method [3] have been proposed. The third family considers a first order perturbation analysis [5] and the fourth family stems from stochastic approximations of power method, (see [15] and the references therein). Finally, the last family relies on either unconstrained or constrained optimizations. In this last family some algorithms are derived from unconstrained optimizations of a specific cost function. In particular, a recursive least square algorithm [22] (respectively a Newton-based adaptive algorithm [13]) enables one to estimate a dominant (respectively a minorant eigendecomposition). As for the constrained

optimizations, they can be performed adaptively by a stochastic gradient algorithm where the constraints are taken into account by a Gram–Schmidt orthogonalization at each iteration [21]. To get rid of these constraints, an alternate solution consists in using an appropriate parametrization [18].

One can find in the literature many papers dealing with convergence analyses, but comparatively few papers concerning the performance analysis of adaptive eigenspace estimation are available by now. Among them, Larimore and Calvert [11] present a study of the convergence rate and the steady-state variance of the Thompson algorithm. Then, Yang and Kaveh made an analysis of the convergence rate and stability of their constrained gradient search procedure, under the classical independence assumption. An analysis of the parametrized stochastic gradient algorithm by Regalia [18] was sketched out in [6] and [7]. Finally, a deflation algorithm for tracking dominant or minorant eigensubspaces [19] and some algorithms tracking dominant eigensubspaces from a least square-like approach (see [23,24]) were presented and studied by the same tools. The main aim of this paper is to study the convergence and performances of a parametrized adaptive algorithm that gives a canonic orthonormal eigenbasis by introducing the necessary methodology and exploiting some of the results that can be derived therefrom.

This paper is organized as follows. After introducing some notations and describing the

parametrization of the orthonormal eigenvectors of the covariance matrices in Section 2, we study the convergence of the coupled stochastic gradient algorithms with the help of the associated ODE in Section 3. A methodology for evaluating the performance by computing the variances of the estimated eigenvectors and of the estimated projection matrices on eigenspaces for fixed gain factors is given in Section 4. We take into account the origin of the covariance matrices, by studying the case where the successive vector signals are independent (spatial case), autoregressive or moving average processes of any order (temporal case). In Section 5, we show that the performance in terms of misadjustment and speed of convergence, can be improved when the symmetric-centrosymmetric property of some covariance matrices is exploited. Finally, numerical results on the asymptotic performance of the algorithm such as mean square errors of estimated eigenvectors and of estimated projection matrices on an eigensubspace, which are evaluated by using the analysis developed in Section 4 are presented in Section 6.

The following notations are used in the paper. Matrices and vectors are represented by bold upper case and bold lower case character, respectively. Vectors are by default in column orientation. T stands transpose and I is the identity matrix. $E(\cdot)$, $cov(\cdot)$, $Tr(\cdot)$, $\|\cdot\|_{Frob}$ denote the expectation, the covariance, the trace operator and the Frobenuis matrix norm, respectively. $Vec(\cdot)$ is the 'vectorization' operator that turns a matrix into a vector consisting of the columns of the matrix stacked one below another. $\perp$ means 'orthogonal to' and $sp\{v_1, \ldots, v_k\}$ denotes the vector space spanned by the vectors $v_1, \ldots, v_k$. $AR(p)$, $MA(q)$ and $ARMA(p,q)$ denote autoregressive, moving average and autoregressive moving average processes of order p, q and p,q respectively.

2. Parametrization of the problem

We tackle the problem of adaptively estimating m normalized eigenvectors $q_1, \ldots, q_m$ corresponding to the m largest [or smallest] distinct eigenvalues $(\lambda_1 > \lambda_2 > \cdots > \lambda_m)$ [respectively $\lambda_{n-m+1} > \cdots > \lambda_n$] of an $n \times n$ covariance matrix $\Gamma_x = E[xx^T]$ of a Gaussian distributed, zero mean real random vector

x. To solve this problem, a method was proposed in the real case in [18] and then extended to the complex case in [6][1] where the constrained maximizations [respectively minimizations] of Rayleigh quotients,

$$\max_{\|q_1\| = 1} q_1^T \Gamma_x q_1 \tag{1}$$

and

$$\max_{\|q_i\| = 1, q_i \perp sp\{q_1, \ldots, q_{i-1}\}} q_i^T \Gamma_x q_i, \quad \text{for } i = 2 \ldots, m, \tag{2}$$

or equivalently

$$\max_{\{q_1, \ldots, q_m \text{orthonormal}\}} \sum_{i=1}^{m} q_i^T \Gamma_x q_i \tag{3}$$

that are taken into account in [21] by a Gram–Schmidt orthogonalization are replaced by unconstrained maximizations [respectively minimizations] thanks to a Givens parametrization of the different constraints. q_1 is the last column of a orthogonal matrix Q_1 and the other vectors q_i can be written as

$$q_1 = Q_1 \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \qquad q_2 = Q_1 \begin{bmatrix} Q_2 \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ 0 \end{bmatrix}, \ldots,$$

$$q_m = Q_1 \begin{bmatrix} Q_2 \begin{bmatrix} Q_m \begin{bmatrix} 0 \\ 1 \end{bmatrix} \\ 0 \\ 0 \end{bmatrix} \end{bmatrix}, \tag{4}$$

where Q_i is the following orthogonal matrix of order $n - i + 1$:

$$Q_i = U_{i,1} \ldots U_{i,j} \ldots U_{i,n-i}$$

$$\text{with } U_{i,j} \overset{\text{def}}{=} \begin{bmatrix} I_{j-1} & 0 & 0 & 0 \\ 0 & -\sin\theta_{i,j} & \cos\theta_{i,j} & 0 \\ 0 & \cos\theta_{i,j} & \sin\theta_{i,j} & 0 \\ 0 & 0 & 0 & I_{n-i-j} \end{bmatrix} \tag{5}$$

and $\theta_{i,j}$ belongs to $]-\pi/2, +\pi/2[$. The existence of such a parametrization for all orthonormal sets

---

[1] The complex Givens parametrization gives a very similar algorithm, the convergence analysis of which can be studied by the same arguments as in the real case; however, as we shall show in Section 4, the performance analysis would unfortunately lead to cumbersome calculus. For this reason we consider the real parametrization only.

$\{q_1, \ldots, q_m\}$ is shown in [18]. It consists of $m(2n - m - 1)/2$ real parameters. Furthermore, this parametrization is unique if we add some constraints. The first component distinct from zero of the last vector of unit norm that appears in the deflation procedure (4), $\boldsymbol{Q}_i \begin{bmatrix} \mathbf{0} \\ 1 \end{bmatrix}$ for $i = 1, \ldots, m$ of Eq. (4), that is of $q_1, \boldsymbol{Q}_1^T q_2, [\boldsymbol{Q}_2^T, \mathbf{0}] \boldsymbol{Q}_1^T q_3, \ldots, [\boldsymbol{Q}_{m-1}, \mathbf{0}]$ $\ldots [\boldsymbol{Q}_2^T, \mathbf{0}] \boldsymbol{Q}_1^T q_m$ must be positive. We can derive from the maximization (3) a stochastic gradient algorithm. In order to simplify this algorithm, a deflation procedure, inspired by the maximizations (1) and (2) was proposed [18]. The maximization (1) is performed with the help of the classical stochastic gradient algorithm, in which the parameters are $\theta_{1,1}, \ldots, \theta_{1,n-1}$ whereas the maximizations (2) are realized thanks to stochastic gradient algorithms with respect to the parameters $\theta_{i,1}, \ldots, \theta_{i,n-i}$, in which the preceding parameters $\theta_{l,1}(k), \ldots, \theta_{l,n-l}(k)$ for $l = 1, \ldots, i - 1$ are injected from the $i - 1$ previous algorithms. The deflation procedure is achieved by coupled stochastic gradient algorithms. This rather intuitive part of the computational process was confirmed by simulation results [18]. However a formal analysis of the convergence and performances had not been performed yet, and this indeed is the main problem addressed in this paper.

3. Convergence of the coupled algorithms

The main difficulty in studying the convergence of the stochastic gradient algorithms derived from this deflation approach comes from the existence of coupled algorithms. In order to study their convergence, these coupled stochastic gradient algorithms need to be globally written as[2]

$$\begin{bmatrix} \boldsymbol{\Theta}_1(k+1) \\ \cdot \\ \boldsymbol{\Theta}_m(k+1) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Theta}_1(k) \\ \cdot \\ \boldsymbol{\Theta}_m(k) \end{bmatrix} + \gamma_k \begin{bmatrix} \boldsymbol{g}_1(\boldsymbol{\Theta}_1(k), \boldsymbol{x}_k) \\ \cdot \\ \boldsymbol{g}_m(\boldsymbol{\Theta}_1(k), \ldots, \boldsymbol{\Theta}_m(k), \boldsymbol{x}_k) \end{bmatrix}$$

(6)

with $\boldsymbol{\Theta}_i \stackrel{\text{def}}{=} [\theta_{i,1}, \ldots, \theta_{i,n-i}]^T$ and $\boldsymbol{g}_1(\boldsymbol{\Theta}_1, \boldsymbol{x}) \stackrel{\text{def}}{=} \frac{1}{2} \nabla_{\boldsymbol{\Theta}_1}(q_1^T \boldsymbol{x})^2, \ldots, \boldsymbol{g}_m(\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_m, \boldsymbol{x}) \stackrel{\text{def}}{=} \frac{1}{2} \nabla_{\boldsymbol{\Theta}_m}(q_m^T \boldsymbol{x})^2$, or more compactly:[3]

$$\boldsymbol{\Theta}(k+1) = \boldsymbol{\Theta}(k) + \gamma_k \boldsymbol{H}[\boldsymbol{\Theta}(k), \boldsymbol{x}_k] \quad (7)$$

with $\boldsymbol{\Theta} \stackrel{\text{def}}{=} [\boldsymbol{\Theta}_1^T, \ldots, \boldsymbol{\Theta}_m^T]^T$ and $\boldsymbol{H} \stackrel{\text{def}}{=} [\boldsymbol{g}_1^T, \ldots, \boldsymbol{g}_m^T]^T$. The study of the convergence of the coupled stochastic gradient algorithms (7) is intimately connected to the stability properties of the associated ODE introduced by Ljung [12]:

$$\frac{\mathrm{d}\boldsymbol{\Theta}(t)}{\mathrm{d}t} = \boldsymbol{h}[\boldsymbol{\Theta}(t)] \quad (8)$$

where $\boldsymbol{h}(\boldsymbol{\Theta})$ is the mean field, i.e. $\boldsymbol{h}(\boldsymbol{\Theta}) \stackrel{\text{def}}{=} \mathrm{E}[\boldsymbol{H}(\boldsymbol{\Theta}, \boldsymbol{x}_k)]$.

Stability of the ODE (8). We suppose that $m = 2$, as the extension to $m > 2$ is straightforward. We note that the solutions $\boldsymbol{\Theta}(t)$ of Eq. (8) are coupled in a 'triangular form'. That is to say, $\boldsymbol{\Theta}_j$ is dependent on $\boldsymbol{\Theta}_i$ for $i < j$ but not on $\boldsymbol{\Theta}_i$ for $i > j$. Therefore $\boldsymbol{\Theta}_1$ can be analyzed independently of the remaining $\boldsymbol{\Theta}_i$. Since $\boldsymbol{g}_1(\boldsymbol{\Theta}_1, \boldsymbol{x})$ is the derivative of a positive field, the set of the stationary points of the part of the Eq. (8) which is associated with $\boldsymbol{\Theta}_1$, is globally asymptotically stable for that equation. But, thanks to the stationary property of the Rayleigh quotient (1) and the parametrization (4), only the parameters $\boldsymbol{\Theta}_1$ solutions of the maximization (1) are globally asymptotically stable for that equation. According to a classical result on the stability of the ODE [1], if we linearize the part of Eq. (8) which is associated with $\boldsymbol{\Theta}_1$ around a stable stationary point, the locally linearized equation is stable. And as the stable stationary points of a linear ODE are exponentially stable, the kth component of $\mathrm{d}\boldsymbol{\Theta}_1(t)/\mathrm{d}t$ behaves as $t \rightarrow +\infty$ as

$$\left[\frac{\mathrm{d}\boldsymbol{\Theta}_1(t)}{\mathrm{d}t} \right]_k \sim \alpha_k \exp(-\mu_k t) \quad \text{with } \mu_k > 0, \quad (9)$$

---

[2] We can introduce a block-diagonal gain $\gamma_k \text{diag}[\text{diag}(a_1), \ldots, \text{diag}(a_m)]$ in place of the scalar gain γ_k in order to take into account a better tradeoff between the misadjustment and the speed of convergence of each eigenvector q_i. All the following developments can be easily extended with this block-diagonal gain.

[3] In case of a minimization, $\boldsymbol{g}_1(\boldsymbol{\Theta}_1, \boldsymbol{x}) \stackrel{\text{def}}{=} -\frac{1}{2} \nabla_{\boldsymbol{\Theta}_1}(q_1^T \boldsymbol{x})^2, \ldots, \boldsymbol{g}_m(\boldsymbol{\Theta}_1, \ldots, \boldsymbol{\Theta}_m, \boldsymbol{x}) \stackrel{\text{def}}{=} -\frac{1}{2} \nabla_{\boldsymbol{\Theta}_m}(q_m^T \boldsymbol{x})^2$. We consider throughout the paper the case of a maximization only, and the case of a minimization can be studied similarly.

meaning that the ratio between the left and right-hand sides tends to 1. Consider the Lyapunov function

$$W(t) \stackrel{\mathrm{def}}{=} -\mathrm{E}(\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t)^2 \geqslant -\lambda_1,$$

and its time derivative

$$\frac{\mathrm{d}W(t)}{\mathrm{d}t} = -\frac{\mathrm{d}\boldsymbol{\Theta}_1^{\mathrm{T}}(t)}{\mathrm{d}t}\mathrm{E}[\nabla_{\boldsymbol{\Theta}_1}(\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t)^2]$$

$$- \frac{\mathrm{d}\boldsymbol{\Theta}_2^{\mathrm{T}}(t)}{\mathrm{d}t}\mathrm{E}[\nabla_{\boldsymbol{\Theta}_2}(\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t)^2].$$

By hypothesis we have

$$\frac{\mathrm{d}\boldsymbol{\Theta}_2^{\mathrm{T}}(t)}{\mathrm{d}t}\mathrm{E}[\nabla_{\boldsymbol{\Theta}_2}(\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t)^2] = \|\mathrm{E}[\nabla_{\boldsymbol{\Theta}_2}(\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t)^2]\|^2,$$

and since $\boldsymbol{q}_2$ and $\mathrm{d}\boldsymbol{q}_2/\mathrm{d}\boldsymbol{\Theta}_1$ are bounded, $\mathrm{E}[\nabla_{\boldsymbol{\Theta}_1}(\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t)^2]$ is also bounded. So, thanks to Eq. (9), we have with $\alpha > 0$ and $\mu > 0$:

$$\left| \frac{\mathrm{d}\boldsymbol{\Theta}_1^{\mathrm{T}}(t)}{\mathrm{d}t}\mathrm{E}[\nabla_{\boldsymbol{\Theta}_1}(\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t)^2] \right|$$

$$\leqslant \left\| \frac{\mathrm{d}\boldsymbol{\Theta}_1(t)}{\mathrm{d}t} \right\| \|\mathrm{E}[\nabla_{\boldsymbol{\Theta}_1}(\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t)^2]\| \leqslant \alpha \exp(-\mu t).$$

Consequently,

$$\frac{\mathrm{d}W(t)}{\mathrm{d}t} \leqslant + \alpha \exp(-\mu t) - \|\mathrm{E}[\nabla_{\boldsymbol{\Theta}_2}|\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t|^2]\|^2.$$

Then $W(t) + (\alpha/\mu)\exp(-\mu t)$ is a decreasing function of t, so $\lim_{t\to+\infty}W(t)$ exists, which implies $\lim_{t\to+\infty}(\mathrm{d}W(t)/\mathrm{d}t) = 0$ and then

$$\lim_{t\to+\infty} \mathrm{E}[\nabla_{\boldsymbol{\Theta}_2}(\boldsymbol{q}_2^{\mathrm{T}}(t)\boldsymbol{x}_t)^2] = \boldsymbol{0}.$$

Therefore the stationary points of the part of the Eq. (8) associated to $\boldsymbol{\Theta}_2$, are globally asymptotically stable for that equation. And thanks to the stationary property of the Rayleigh quotient (2) and the parametrization (4), only the parameters $\boldsymbol{\Theta}_2$ which are the solutions of the maximization (2) are globally asymptotically stable for that equation. So the following result is established.

Result 1. The parameters $\boldsymbol{\Theta}_*$ that maximize (1) and (2) are globally asymptotically stable for its associated ODE (8).

Convergence of the stochastic gradient algorithm (7). Although the stochastic gradient algorithm (8) can be viewed as a discrete time approximation to its associated ODE (8), the question of the connection of their limiting behaviour is not straightforward because the algorithm may have a much more complex asymptotic behaviour than a given solution of the ODE. To induce a connection, we are firstly interested in the hypotheses of Benveniste et al. [2] which specify conditions under which the stochastic algorithm (7) converges almost surely to the asymptotically stable points of Eq. (8). Thus we suppose that the gain sequence γ_k satisfies the conditions $\sum_{k=1}^{\infty}\gamma_k = +\infty$ and $\lim_{k\to+\infty}\gamma_k = 0$. The state vector $\boldsymbol{x}_k$ of the investigated algorithm (7) must have a dynamic Markov representation controlled by $\boldsymbol{\Theta}$. This signal model by Benveniste et al. [2] is fulfilled in our case because we consider that the observations $\boldsymbol{x}_k$ are independent or derived from the specific correlation model $\boldsymbol{x}_k \stackrel{\mathrm{def}}{=} [x_k, x_{k-1}, \ldots, x_{k-n+1}]^{\mathrm{T}}$ with x_k a stationary ARMA process. Thus $\boldsymbol{x}_k = f(\boldsymbol{\xi}_k)$, where $\boldsymbol{\xi}_k$ is a Markov chain independent of $\boldsymbol{\Theta}$, $\boldsymbol{\xi}_k \stackrel{\mathrm{def}}{=} \boldsymbol{x}_k$ in the independent case and $\boldsymbol{\xi}_k \stackrel{\mathrm{def}}{=} (\boldsymbol{w}_{k-n}, u_k, u_{k-1}, \ldots, u_{k-n+1})$ is issued from the state representation of an ARMA process, viz.,

$$\boldsymbol{w}_k = \boldsymbol{A}\boldsymbol{w}_{k-1} + \boldsymbol{b}u_k, \quad x_k = \boldsymbol{c}^{\mathrm{T}}\boldsymbol{w}_{k-1} + \mathrm{d}u_k.$$

To apply the corollary 6 of [2] (p. 46), which states that $\boldsymbol{\Theta}(k)$ defined by Eq. (7) converges almost surely to one of the asymptotically stable points $\boldsymbol{\Theta}_*$, we encounter two difficulties if we refer to the original Kushner–Clark theorem. The equilibrium points $\boldsymbol{\Theta}_*$ must be unique and the trajectory of $\boldsymbol{\Theta}(k)$ must intersect a compact subset infinitely often. In our application, the equilibrium points $\boldsymbol{\Theta}_*$ are not unique, and proving that the trajectory of $\boldsymbol{\Theta}(k)$ intersects a compact subset infinitely often is very challenging, as many authors believe. To our best knowledge, this condition has been proved only by Oja and Karhunen [16], and under the hypothesis that $\boldsymbol{x}_k$ is uniformly bounded.

However, we must note at this point that $\boldsymbol{q}_i$, $i = 1, \ldots, m$, $\boldsymbol{H}$ and $\boldsymbol{h}$ are 2π-periodic functions of

each parameter $\theta_{i,j}$. So, the set of the asymptotically stable points is composed of isolated points that remain 'sparse' near infinity.[4] Furthermore, the stability of the ODE (8) followed from the existence of a very regular Lyapunov function. Using these two particular properties, we can use a recent result by Fort and Pagès (theorem 6 of [8]). This theorem transfers the convergence of the solution of the ODE to the algorithm when the ODE has a Lyapunov function and when the equilibrium points $\boldsymbol{\Theta}_*$ are isolated. Under these conditions, it specifies that the stochastic algorithm converges almost surely to one of the points $\boldsymbol{\Theta}_*$.

Fixed gain γ. Unfortunately, in nonstationary environments the gain sequences γ_k needs to be reduced to constant 'small' steps if we want our algorithm to be able to track the slow variations of the parameters. The convergence results cannot be applied in a strict sense. In this case, the algorithm no longer converges almost surely. However, the weak convergence approach developed by Kushner [10] suggests that, for γ 'small enough', the adaptive algorithm will oscillate around the theoretical limit of the decreasing step scheme.

4. Asymptotic performance analysis

Consider a globally asymptotically stable equilibrium point $\boldsymbol{\Theta}_*$ of the ODE (8) which corresponds to a limit to which the solution $\boldsymbol{\Theta}_k$ of Eq. (7) converges in the decreasing gain case, we can get the asymptotic distribution of $\boldsymbol{\Theta}_k$ for fixed gain factors in stationary situations by using a general result by Benveniste [2] (theorem 2, p. 108).[5] Con-

---

sider the continuous Lyapunov equation

$$G_* C_{\Theta} + C_{\Theta} G_*^{\mathrm{T}} + R_* = O, \tag{10}$$

where $\boldsymbol{G}$ and $\boldsymbol{R}$ are respectively the derivative of the mean field and the covariance of the field of the algorithm (7),

$$\boldsymbol{G}(\boldsymbol{\Theta}) \stackrel{\text{def}}{=} \frac{\mathrm{d}\boldsymbol{h}(\boldsymbol{\Theta})}{\mathrm{d}\boldsymbol{\Theta}}, \tag{11}$$

$$\boldsymbol{R}(\boldsymbol{\Theta}) \stackrel{\text{def}}{=} \sum_{k=-\infty}^{\infty} \mathrm{cov}[\boldsymbol{H}(\boldsymbol{\Theta},\boldsymbol{x}_k),\boldsymbol{H}(\boldsymbol{\Theta},\boldsymbol{x}_0)], \tag{12}$$

and where the subscript $*$ stands for the value of the functions calculated for the parameter $\boldsymbol{\Theta}_*$ that maximizes the expressions (1) and (2). If all the eigenvalues of the derivative of the mean field $\boldsymbol{G}_*$ have strictly negative real parts (condition proved in Section 4.3), then when $\gamma \to 0$ and $t_k \to +\infty$ with $t_k \stackrel{\text{def}}{=} k\gamma$, $1/\sqrt{\gamma}[\boldsymbol{\Theta}(k) - \boldsymbol{\Theta}(t_k)]$ converges in law to a zero mean Gaussian random vector of covariance matrix $\boldsymbol{C}_{\Theta}$, where $\boldsymbol{C}_{\Theta}$ is the unique symmetric solution of Eq. (10). Then, as $\boldsymbol{\Theta}(t_k)$ converges almost surely to $\boldsymbol{\Theta}_*$,

$$\frac{1}{\sqrt{\gamma}}(\boldsymbol{\Theta}(k) - \boldsymbol{\Theta}_*) \overset{\mathscr{L}}{\to} \mathscr{N}(0,\boldsymbol{C}_{\Theta}). \tag{13}$$

We now evaluate the derivative of the mean field and the convariance of the field of algorithm (7).

4.1. Derivative of the mean field

We consider the case $m = 2$. The case $m = 1$ is a byproduct of the case $m = 2$, while the extension to $m > 2$ is straightforward but tedious. Thanks to the property (obtained easily by a flowgraph, see [6,18]), we have

$$\frac{\mathrm{d}\boldsymbol{q}_1(\boldsymbol{\Theta}_1)}{\mathrm{d}\boldsymbol{\Theta}_1} = \boldsymbol{Q}_1'(\boldsymbol{\Theta}_1)\boldsymbol{D}_1(\boldsymbol{\Theta}_1) \tag{14}$$

and

$$\frac{\partial\boldsymbol{q}_2(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_1)}{\partial\boldsymbol{\Theta}_2} = \boldsymbol{Q}_1'(\boldsymbol{\Theta}_1)\boldsymbol{Q}_2'(\boldsymbol{\Theta}_2)\boldsymbol{D}_2(\boldsymbol{\Theta}_2), \tag{15}$$

with $\boldsymbol{Q}_1(\boldsymbol{\Theta}_1) \stackrel{\text{def}}{=} [\boldsymbol{Q}_1'(\boldsymbol{\Theta}_1),\boldsymbol{q}_1(\boldsymbol{\Theta}_1)]$, $\boldsymbol{Q}_2(\boldsymbol{\Theta}_2) \stackrel{\text{def}}{=} [\boldsymbol{Q}_2'(\boldsymbol{\Theta}_2),$ $\boldsymbol{v}(\boldsymbol{\Theta}_2)]$ and with $\boldsymbol{D}_1(\boldsymbol{\Theta}_1)$, $\boldsymbol{D}_2(\boldsymbol{\Theta}_2)$ respectively an $n-1 \times n-1$ and an $n-2 \times n-2$ diagonal

---

[4] In practical use, the parameters $\theta_{i,j}$ remain in $[-\pi/2, +\pi/2]$ when initialized at $\theta_{i,j} = 0$. A test is built into the algorithm to ensure that $|\theta_{i,j}| \leqslant \pi/2$ at each iteration and if for some k, $|\theta_{i,j}|$ becomes greater than $\pi/2$, the update for that parameter is bypassed, and only the remaining parameters are allowed to evolve.

[5] A thorough derivation of this result has been established only under the necessary assumption that the global attractor is unique. However, its practical use in more general situations is usually justified by a general diffusion approximation result [2] (theorem 1, p. 107). For instance in [24], this result was applied to a situation where the globally asymptotically stable set is the continuum $\{(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_m)\in\mathscr{R}^{n\times m}|(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_m)=(\boldsymbol{q}_1,\ldots,\boldsymbol{q}_m)\boldsymbol{U}$ with $\boldsymbol{U}^{\mathrm{T}}\boldsymbol{U}=\boldsymbol{I}_m\}$.

matrix where $D_1(\Theta_1)_{n-1,n-1} = D_2(\Theta_2)_{n-2,n-2} = 1$ and $D_1(\Theta_1)_{k,k} = \prod_{l=k+1}^{n-1}\cos(\theta_{1,l})$ for $1 \leqslant k \leqslant n-2$, $D_2(\Theta_2)_{k,k} = \prod_{l=k+1}^{n-2}\cos(\theta_{2,l})$ for $1 \leqslant k \leqslant n-3$. As such, the mean field $h(\Theta_1,\Theta_2)$ of the algorithm (7) can be partitioned as

$$\begin{bmatrix} h_1(\Theta_1) \\ h_2(\Theta_1,\Theta_2) \end{bmatrix}$$

$$= \begin{bmatrix} D_1(\Theta_1)Q_1'^{T}(\Theta_1)\Gamma_x q_1(\Theta_1) \\ D_2(\Theta_2)Q_2'^{T}(\Theta_2)Q_1'^{T}(\Theta_1)\Gamma_x q_2(\Theta_1,\Theta_2) \end{bmatrix}.$$

Consequently,

$$G_* = \begin{bmatrix} G_{11} & O \\ G_{21} & G_{22} \end{bmatrix}. \tag{16}$$

It is shown in Appendix A that G_{11}, G_{22} and G_{21} are respectively given by

$$G_{11} = D_1(\Theta_1*)Q_1'^{T}(\Theta_1*)(\Gamma_x - \lambda_1 I_n)Q_1'(\Theta_1*)D_1(\Theta_1*), \tag{17}$$

$$G_{22} = D_2(\Theta_2*)Q_2'^{T}(\Theta_2*)Q_1'^{T}(\Theta_1*)(\Gamma_x - \lambda_2 I_n)$$
$$\times Q_1'(\Theta_1*)Q_2'(\Theta_2*)D_2(\Theta_2*), \tag{18}$$

$$G_{21} = D_2(\Theta_2*)Q_2'^{T}(\Theta_2*)Q_1'^{T}(\Theta_1*)(\Gamma_x - \lambda_2 I_n)$$
$$\times Q_{1,2}(\Theta_1*,\Theta_2*), \tag{19}$$

where $Q_{1,2}(\Theta_1,\Theta_2)$ is defined in Appendix A.

Lastly, let us note that the performance analysis could be extended to complex data if the relation (14) could be easily extended. In fact Eq. (14) becomes in the complex case

$$\frac{dq_1(\Theta_1)}{d\Theta_1} = [Q_1'(\Theta_1)D_1(\Theta_1), Q_1(\Theta_1)K_1(\Theta_1)], \tag{20}$$

where now Θ_1 denotes the parameter $(\psi_{1,1},\ldots,\psi_{1,n-1},\ldots,\phi_{1,1},\ldots,\phi_{1,n-1})$ of the complex parametrization [6,7], $D_1(\Theta_1)$ is the $n-1 \times n-1$ diagonal matrix with $D_1(\Theta_1)_{n-1,n-1} = 1$ and $D_1(\Theta_1)_{k,k} = \prod_{l=k+1}^{n-1}\cos(\psi_{1,l})\exp(\phi_{1,l})$ for $1 \leqslant k \leqslant n-2$ and K_1 an $n \times n$ non-diagonal matrix. Because the relation (20) is much more complicated than the relation (14), the performance analysis in the complex case would be much more cumbersome.

4.2. Covariance of the field

4.2.1. Independent observations

For independent observations x_k which generally correspond to spatial situations, we also consider the case $m = 2$ for the same reason as for the derivative. $H(\Theta_1,\Theta_2,x_k)$ can be partitioned as

$$\begin{bmatrix} H_1(\Theta_1,x_k) \\ H_2(\Theta_1,\Theta_2,x_k) \end{bmatrix} = \begin{bmatrix} \frac{dq_1^{T}(\Theta_1)}{d\Theta_1} x_k x_k^{T} q_1(\Theta_1) \\ \frac{\partial q_2^{T}(\Theta_1,\Theta_2)}{\partial\Theta_2} x_k x_k^{T} q_2(\Theta_1,\Theta_2) \end{bmatrix}.$$

Because $\mathrm{E}[H_i(\Theta_*,x_k)] = (\partial q_i^{T}(\Theta)/\partial\Theta)_* \Gamma_x q_i(\Theta_*) = \frac{1}{2}\lambda_i(\partial\|q_i(\Theta)\|^2/\partial\Theta)_* = 0$, for $i = 1,2$ the expression (12) becomes

$$R_* = \sum_{k=-\infty}^{+\infty} \mathrm{E}[H(\Theta_{1*},\Theta_{2*},x_k)H^{T}(\Theta_{1*},\Theta_{2*},x_0)]$$

$$= \mathrm{E}[H(\Theta_{1*},\Theta_{2*},x_0)H^{T}(\Theta_{1*},\Theta_{2*},x_0)].$$

Therefore, R_* can be partitioned as

$$R_* = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}, \tag{21}$$

where

$$R_{ij} = \frac{\partial q_i^{T}(\Theta)_*}{\partial\Theta_i} \mathrm{E}[x_0 x_0^{T} q_i(\Theta_*) q_j^{T}(\Theta_*) x_0 x_0^{T}] \frac{\partial q_j(\Theta)_*}{\partial\Theta_j}. \tag{22}$$

Is is shown in Appendix B that

$$R_{21} = R_{12}^{T} = O, \tag{23}$$

$$R_{11} = D_1(\Theta_1*)Q_1'^{T}(\Theta_1*)\lambda_1\Gamma_x Q_1'(\Theta_1*)D_1(\Theta_1*), \tag{24}$$

$$R_{22} = D_2(\Theta_2*)Q_2'^{T}(\Theta_2*)Q_1'^{T}(\Theta_1*)\lambda_2\Gamma_x Q_1'(\Theta_1*)$$
$$\times Q_2'(\Theta_2*)D_2(\Theta_2*). \tag{25}$$

4.2.2. Correlated observations

We consider the specific correlation model $x_k \overset{\text{def}}{=} [x_k, x_{k-1}, \ldots, x_{k-n+1}]^{T}$ with x_k being an MA(q), an AR(p) or an ARMA(p,q) stationary process which in general corresponds to temporal situations. In this case R_* is no longer block diagonal. It is shown in Appendix C that

$$R_{11} = D_1(\Theta_1*)Q_1'^{T}(\Theta_1*)(\lambda_1\Gamma_x + \Gamma_{11})Q_1'(\Theta_1*)$$
$$\times D_1(\Theta_1*), \tag{26}$$

$$R_{22} = D_2(\Theta_2*)Q_2'^{T}(\Theta_2*)Q_1'^{T}(\Theta_1*)(\lambda_2\Gamma_x + \Gamma_{22})$$
$$\times Q_1'(\Theta_1*)Q_2'(\Theta_2*)D_2(\Theta_2*), \tag{27}$$

$$R_{21} = R_{12}^T = D_2(\Theta_2*)Q_2'^T(\Theta_2*)Q_1'^T(\Theta_1*)$$
$$\times \Gamma_{21}Q_1'(\Theta_1*)D_1(\Theta_1*), \qquad (28)$$

where, for an MA(q) process, $\Gamma_{i,j}$ takes the value

$$\Gamma_{i,j} = \sum_{k=1}^{q+n-1} \Gamma_k q_j q_i^T \Gamma_k + \Gamma_k^T q_j q_i^T \Gamma_k^T$$
$$+ (q_i^T \Gamma_k q_j)\Gamma_k + (q_i^T \Gamma_k^T q_j)\Gamma_k^T, \qquad (29)$$

where Γ_k denotes the cross-correlation matrix $E[x_k x_0^T]$, and respectively for an AR(p) process

$$\Gamma_{i,j} = \sum_{k=1}^{n-2} \Gamma_k q_j q_i^T \Gamma_k + \Gamma_k^T q_j q_i^T \Gamma_k^T$$
$$+ (q_i^T \Gamma_k q_j)\Gamma_k + (q_i^T \Gamma_k^T q_j)\Gamma_k^T$$
$$+ \sum_{k=1}^{p} \sum_{l=1}^{p} \alpha_k \alpha_l \frac{(\rho_k \ \rho_l)^{n-1}}{1 - \rho_k \ \rho_l}[\Gamma_{\rho_k} q_j q_i^T \Gamma_{\rho_l}$$
$$+ \Gamma_{\rho_k}^T q_j q_i^T \Gamma_{\rho_l}^T + (q_i^T \Gamma_{\rho_k} q_j)\Gamma_{\rho_l}$$
$$+ (q_i^T \Gamma_{\rho_k}^T q_j)\Gamma_{\rho_l}^T]. \qquad (30)$$

For an ARMA(p,q) process, it is shown that the relation (30) also holds provided n is replaced by $n + \text{Sup}(0, q + 1 - p)$.

4.3. Exploitation of the Lyapunov equation (10)

As $G_{ii} = (\partial q_i^T/\partial \Theta_i)(\Gamma_x - \lambda_i I)(\partial q_i/\partial \Theta_i)$, $i = 1, \ldots, m$, and thanks to the orthogonal properties of $\{q_1(\Theta_1), \ldots, q_i(\Theta_1, \ldots, \Theta_i)\}$ which implies $q_j^T(\partial q_i/\partial \Theta_i) = 0^T$ for $j = 1, \ldots, i$, the symmetric matrices G_{ii} are negative definite. The eigenvalues of the block triangular matrix G thus have strictly negative real parts. The condition required in Section 4 is thus fulfilled. The Lyapunov equation (10) cannot be solved in a closed form expression. But since (10) is of triangular form, it can be solved numerically step by step for successive values of m.

The application of a continuity theorem directly adapted from the theorem 6.2a [17], (p. 387) to the differentiable mapping $\Theta \to Q = (q_1, \ldots, q_m)$ gives the asymptotic distribution of eigenvector estimators as

$$\frac{1}{\sqrt{\gamma}}(\text{Vec}(Q(k)) - \text{Vec}(Q_*)) \xrightarrow{\mathscr{L}} \mathscr{N}(0, C_Q), \qquad (31)$$

where

$$C_Q = \frac{d\text{Vec}(Q)}{d\Theta} C_\Theta \frac{d^T\text{Vec}(Q)}{d\Theta},$$
$$\text{with} \quad \frac{d\text{Vec}(Q)}{d\Theta} = \left(\frac{d^T q_1}{d\Theta}, \ldots, \frac{d^T q_m}{d\Theta}\right)^T.$$

In particular,

$$\frac{1}{\sqrt{\gamma}}(q_i(k) - q_{i*}) \xrightarrow{\mathscr{L}} \mathscr{N}(0, C_{q_i}), \qquad (32)$$

where $C_{q_i} = (dq_i/d\Theta) C_\Theta (d^T q_i/d\Theta)$, $i = 1, \ldots, m$.

Then, applying a second time the same theorem of continuity to the differentiable mapping $Q = (q_1, \ldots, q_m) \to P = \sum_{i=1}^m q_i q_i^T$ gives the asymptotic distribution of subspace projector estimators $P(k)$:

$$\frac{1}{\sqrt{\gamma}}(\text{Vec}(P(k)) - \text{Vec}(P_*)) \xrightarrow{\mathscr{L}} \mathscr{N}(0, C_P), \qquad (33)$$

where

$$C_P = \frac{d\text{Vec}(P)}{d\text{Vec}(Q)} C_Q \frac{d^T\text{Vec}(P)}{d\text{Vec}(Q)},$$

$$\text{with} \quad \frac{d\text{Vec}(P)}{d\text{Vec}(Q)} = I_n \otimes Q + (q_1 \otimes I_n, \ldots, q_m \otimes I_n).$$

Thanks to the hypothesis of boundedness of the parameters $\Theta(k)$, the convergence in law of $(1/\sqrt{\gamma})(\Theta(k) - \Theta_*)$ (see Eq. (13)) implies the convergence of the first two moments. So $\|E(\Theta(k)) - \Theta_*\|_{\text{Frob}}^2 = o(\gamma)$ and $\text{cov}(\Theta(k)) \sim \gamma C_\Theta$ when $\gamma \to 0$ and $k \to +\infty$. Consequently, by expanding Q and P around Θ_*, we obtain the mean square error of eigenvectors and subspace projection matrix estimators:

$$\|Q(k) - Q_*\|_{\text{Fro}}^2 \sim \gamma \text{Tr}(C_Q) = \gamma \text{Tr}\left[\sum_{i=1}^m \frac{dq_i}{d\Theta} C_\Theta \frac{dq_i^T}{d\Theta}\right]_*. \qquad (34)$$

In particular,

$$\|q_i(k) - q_{i*}\|_{\text{Fro}}^2 \sim \gamma \text{Tr}(C_{q_i}) = \gamma \text{Tr}\left[\frac{dq_i}{d\Theta} C_\Theta \frac{dq_i^T}{d\Theta}\right]_* \qquad (35)$$

and

$$\|P(k) - P_*\|_{\text{Fro}}^2 \sim \gamma \text{Tr}(C_P)$$

$$= 2\gamma(\text{Tr}(C_Q) + \sum_{1 \leqslant i,j \leqslant m} q_{i*}^T(C_Q)_{j,i} q_{j*})$$

$$= 2\gamma \text{Tr}\left(\sum_{i=1}^m \frac{dq_i}{d\Theta} C_\Theta \frac{dq_i^T}{d\Theta}\right)_*$$

$$+ 2\gamma\left(\sum_{1 \leqslant i \neq j \leqslant m} q_i^T\left[\frac{dq_j}{d\Theta} C_\Theta \frac{dq_i^T}{d\Theta}\right] q_j\right)_*. \tag{36}$$

5. Special case of symmetric-centrosymmetric covariance matrices

To improve the accuracy of the subspace estimation, we can exploit the symmetric-centrosymmetric or block-symmetric-centrosymmetric property of some covariance matrices. This property occurs in important applications: temporal covariance matrices obtained from a temporal sampling of a stationary signal, and spatial covariance matrices issued from uncorrelated and band-limited sources observed on a symmetric-centrosymmetric sensor array (for example on uniform linear arrays) [20] are centro-symmetric; spatio-temporal covariance matrices used in subspace methods for blind identification of multichannel FIR filters [14] are block-symmetric-centrosymmetric.

In the real case, we use the property that an orthonormal eigenbasis of a symmetric centro-symmetric matrix can be obtained from orthonormal eigenbases of two half-size symmetric real matrices [4]. For example if n is even, Γ can be partitioned as follows:

$$\Gamma = \begin{bmatrix} \Gamma_1 & \Gamma_2^T \\ \Gamma_2 & J\Gamma_1 J \end{bmatrix}, \tag{37}$$

where J is an $n/2 \times n/2$ matrix with ones on its anti-diagonal and zeroes elsewhere, and $\Gamma_1^T = \Gamma_1$, $J\Gamma_2^T = \Gamma_2^T J$. Then we may determine $n/2$ symmetric [respectively $n/2$ skew symmetric] orthonormal eigenvectors q_i of Γ and corresponding eigenvalues λ_i from the $n/2$ orthonormal eigenvectors u_i of $\Gamma_1 + \varepsilon_i J\Gamma_2$, that is from the solutions of $(\Gamma_1 + \varepsilon_i J\Gamma_2)u_i = \lambda_i u_i$ with $\varepsilon_i = +1$ [respectively

$\varepsilon_i = -1$], where

$$q_i = \frac{1}{\sqrt{2}}\begin{bmatrix} u_i \\ \varepsilon_i J u_i \end{bmatrix} \quad \text{with } \varepsilon_i = +1 \text{ [resp. } \varepsilon_i = -1\text{]}. \tag{38}$$

If we note that $\Gamma_1 + \varepsilon_i J\Gamma_2 = \frac{1}{2}E((x_k' + \varepsilon_i Jx_k'') \times (x_k' + \varepsilon_i Jx_k'')^T)$ with $x \stackrel{\text{def}}{=} [x' \quad x'']^T$, we can use the Givens adaptive method described previously and we can follow the same steps until an equation similar to (10) is solved. As in Section 4, we only consider the case $m = 2$ for the same reason. The eigenvectors q_1 and q_2 of Γ have the structure (38) with the correct signs of ε_1 and ε_2. If $\varepsilon_1 = \varepsilon_2 \stackrel{\text{def}}{=} \varepsilon$, the formulas (16), (17), (18) and (19) for the derivative of the mean field and Eqs. (21) and (22) for the covariance of the field of the algorithm still hold, provided q_1, q_2 and Γ_x are replaced by u_1, u_2 and $2(\Gamma_1 + \varepsilon J\Gamma_2)$. But for $\varepsilon_1 = -\varepsilon_2$, since $(x_k' + \varepsilon_1 Jx_k'')$ is used for computing q_1 and $(x_k' - \varepsilon_1 Jx_k'')$ is used for computing q_2, the analysis of Section 4 is no longer valid. However, thanks to the parameterization of $u_1(\Theta_1)$ and $u_2(\Theta_2)$ we have

$$\frac{dq_i(\Theta_i)}{d\Theta_i} = \frac{1}{\sqrt{2}}\begin{bmatrix} I \\ \varepsilon_i J \end{bmatrix} \frac{du_i(\Theta_i)}{d\Theta_i}$$

$$= \frac{1}{\sqrt{2}}\begin{bmatrix} I \\ \varepsilon_i J \end{bmatrix} Q_1'(\Theta_i) D_1(\Theta_i). \tag{39}$$

The matrix H of Eq. (7) becomes

$$H = \begin{bmatrix} H_1(\Theta_1, x_k' + \varepsilon_1 Jx_k'') \\ H_2(\Theta_2, x_k' + \varepsilon_2 Jx_k'') \end{bmatrix},$$

with

$$H_i = D_1(\Theta_i)Q_1'^T(\Theta_i)(x_k' + \varepsilon_i Jx_k'')(x_k' + \varepsilon_i Jx_k'')^T u_1(\Theta_i),$$

$$\text{for } i = 1,2. \tag{40}$$

Eq. (16) becomes immediately

$$G_* = \begin{bmatrix} G_{11} & O \\ O & G_{22} \end{bmatrix}, \tag{41}$$

with

$$G_{ii} = 2D_1(\Theta_i*)Q_1'^T(\Theta_i*)(\Gamma_1 + \varepsilon_i J\Gamma_2 - \lambda_i I)$$

$$\times Q_1'(\Theta_i*)D_1(\Theta_i*), \quad \text{for } i = 1,2, \tag{42}$$

and for independent observations $\boldsymbol{x}_k$, it is shown after some manipulations of a relation similar to the relation (48), reported in Appendix D, that

$$\boldsymbol{R}_* = \begin{bmatrix} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{bmatrix}, \tag{43}$$

with

$$\boldsymbol{R}_{21} = \boldsymbol{R}_{12}^{\mathrm{T}} = \boldsymbol{O}, \tag{44}$$

$$\boldsymbol{R}_{ii} = 4\lambda_i \boldsymbol{D}_1(\boldsymbol{\Theta}_i*)\boldsymbol{Q}_1'^{\mathrm{T}}(\boldsymbol{\Theta}_i*)(\boldsymbol{\Gamma}_1 + \varepsilon_i \boldsymbol{J}\boldsymbol{\Gamma}_2)$$
$$\times \boldsymbol{Q}_1'(\boldsymbol{\Theta}_i*)\boldsymbol{D}_1(\boldsymbol{\Theta}_i*), \quad \text{for } i = 1, 2. \tag{45}$$
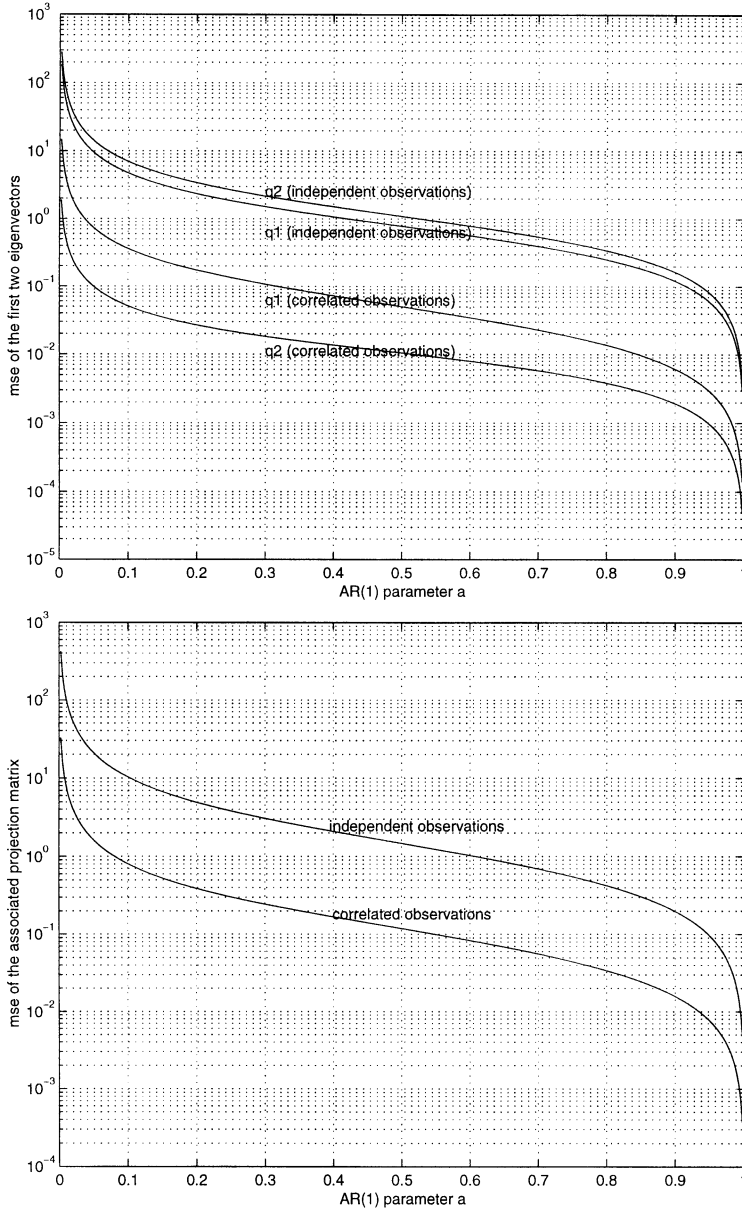
Fig. 1. Mean square errors of the first two eigenvectors and of the associated projection matrix (normalized by the gain factor γ) estimated from independent or correlated AR(1) observations, as a function of the AR parameter a.
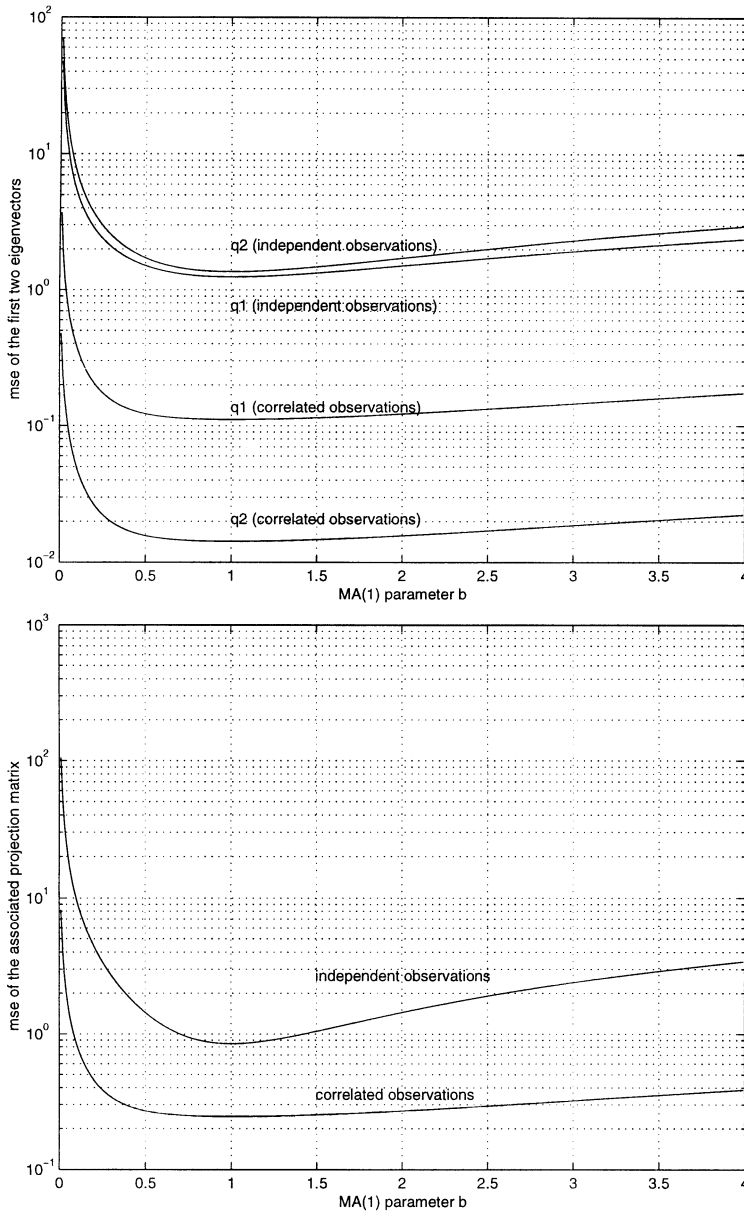
Fig. 2. Mean square errors of the first two eigenvectors and of the associated projection matrix (normalized by the gain factor γ) estimated from independent or correlated MA(1) observations, as a function of the MA parameter b.

By solving the block diagonal system of Eq. (10) and obtain $C_\Theta = \begin{bmatrix} C_{\Theta_1} & \mathbf{0} \\ \mathbf{0} & C_{\Theta_2} \end{bmatrix}$. The estimated parameter $\Theta_1(k)$ and $\Theta_2(k)$ and thus the estimated eigenvectors $q_1(k)$ and $q_2(k)$ are asymptotically uncorrelated. From Eqs. (35) and (39), we can deduce the mean square error of the estimated eigenvectors $q_i(k)$:

$$\|q_i(k) - q_{i*}\|_{\mathrm{Fro}}^2 \sim \gamma \mathrm{Tr}\left[\frac{\mathrm{d}q_i}{\mathrm{d}\boldsymbol{\Theta}_i} C_{\Theta_i} \frac{\mathrm{d}q_i^{\mathrm{T}}}{\mathrm{d}\boldsymbol{\Theta}_i}\right]_*$$

$$= \gamma \mathrm{Tr}\left[\frac{\mathrm{d}u_i}{\mathrm{d}\boldsymbol{\Theta}_i} C_{\Theta_i} \frac{\mathrm{d}u_i^{\mathrm{T}}}{\mathrm{d}\boldsymbol{\Theta}_i}\right]_*, \quad \text{for } i = 1,2, \tag{46}$$
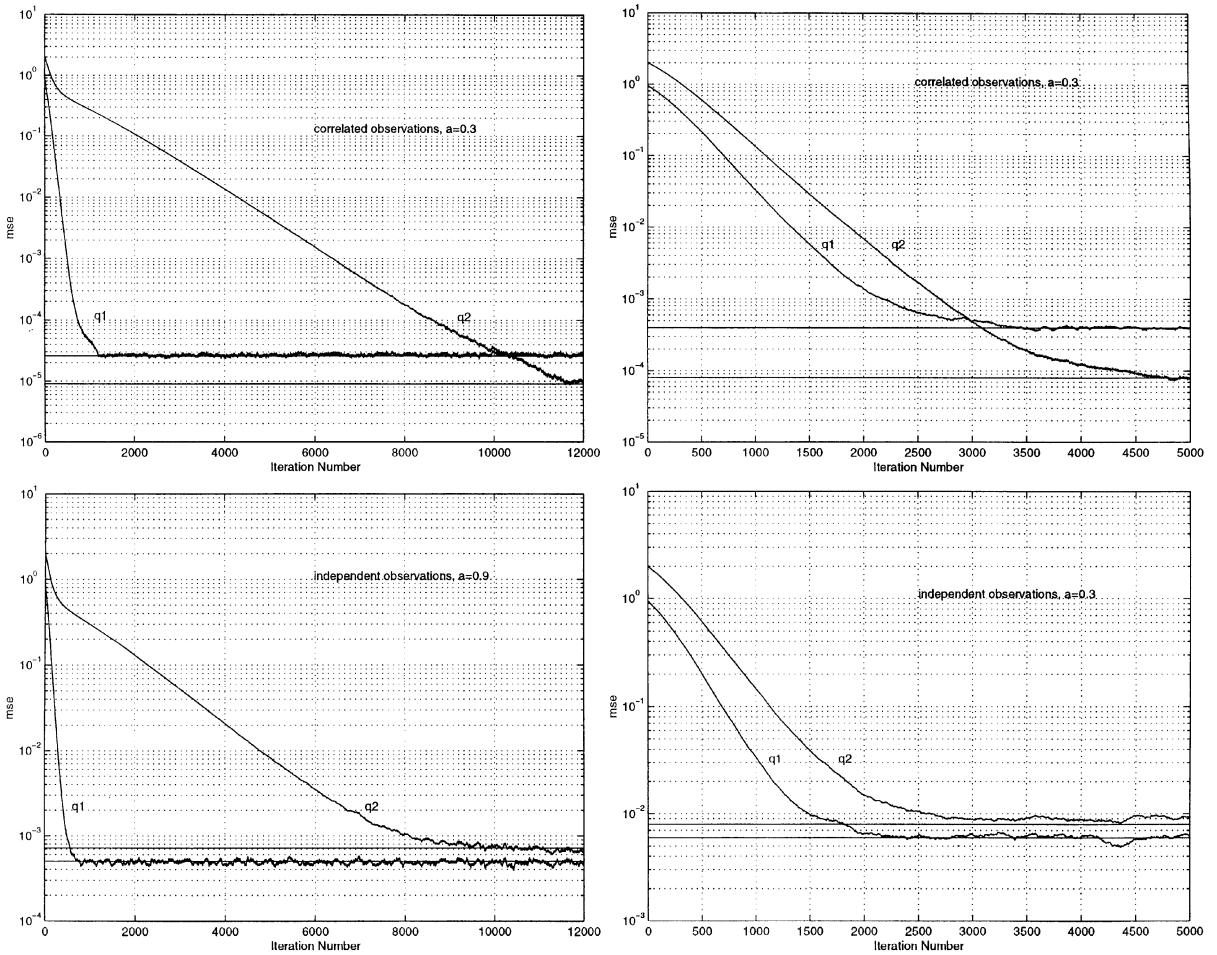
Fig. 3. Learning curves of the mean square error $\mathrm{E}\|\boldsymbol{q}_1(k) - \boldsymbol{q}_{1*}\|_{\mathrm{Fro}}^2$ and $\mathrm{E}\|\boldsymbol{q}_2(k) - \boldsymbol{q}_{2*}\|_{\mathrm{Fro}}^2$ compared to $\gamma\mathrm{Tr}\boldsymbol{C}_{q_1}$ and to $\gamma\mathrm{Tr}\boldsymbol{C}_{q_2}$ averaging 400 independent runs for correlated or independent observations $\boldsymbol{x}_k$, the covariance matrix $\boldsymbol{\Gamma}_x$ being issued from an AR(1) with coefficient $a = 0.9$ or $a = 0.3$.

and the mean square error of the associated estimated projection matrix deduced from Eqs. (36) and (39) and the decorrelation of $\boldsymbol{q}_1(k)$ and $\boldsymbol{q}_2(k)$:

$$\|\boldsymbol{P}(k) - \boldsymbol{P}_*\|_{\mathrm{Fro}}^2 \sim 2\gamma \sum_{i=1}^{2} \mathrm{Tr}\left[\frac{\mathrm{d}\boldsymbol{u}_i}{\mathrm{d}\boldsymbol{\Theta}_i}\boldsymbol{C}_{\Theta_i}\frac{\mathrm{d}\boldsymbol{u}_i^{\mathrm{T}}}{\mathrm{d}\boldsymbol{\Theta}_i}\right]_*. \qquad (47)$$

By taking into account the structure of the eigenvectors, we can expect a better trade-off between the misadjustments and the speeds of convergence because we increase the ratio of the successive eigenvalues of the analyzed covariance matrices

$\boldsymbol{\Gamma}_1 + \varepsilon_i\boldsymbol{J}\boldsymbol{\Gamma}_2$ and we decrease the number of parameters to update. The results and the simulations presented in the following section confirm this conjecture.

6. Results and simulations

We now examine two cases where we compare the results given by numerical solutions of the Lyapunov equation (10) with computer simulations of the algorithm (7).
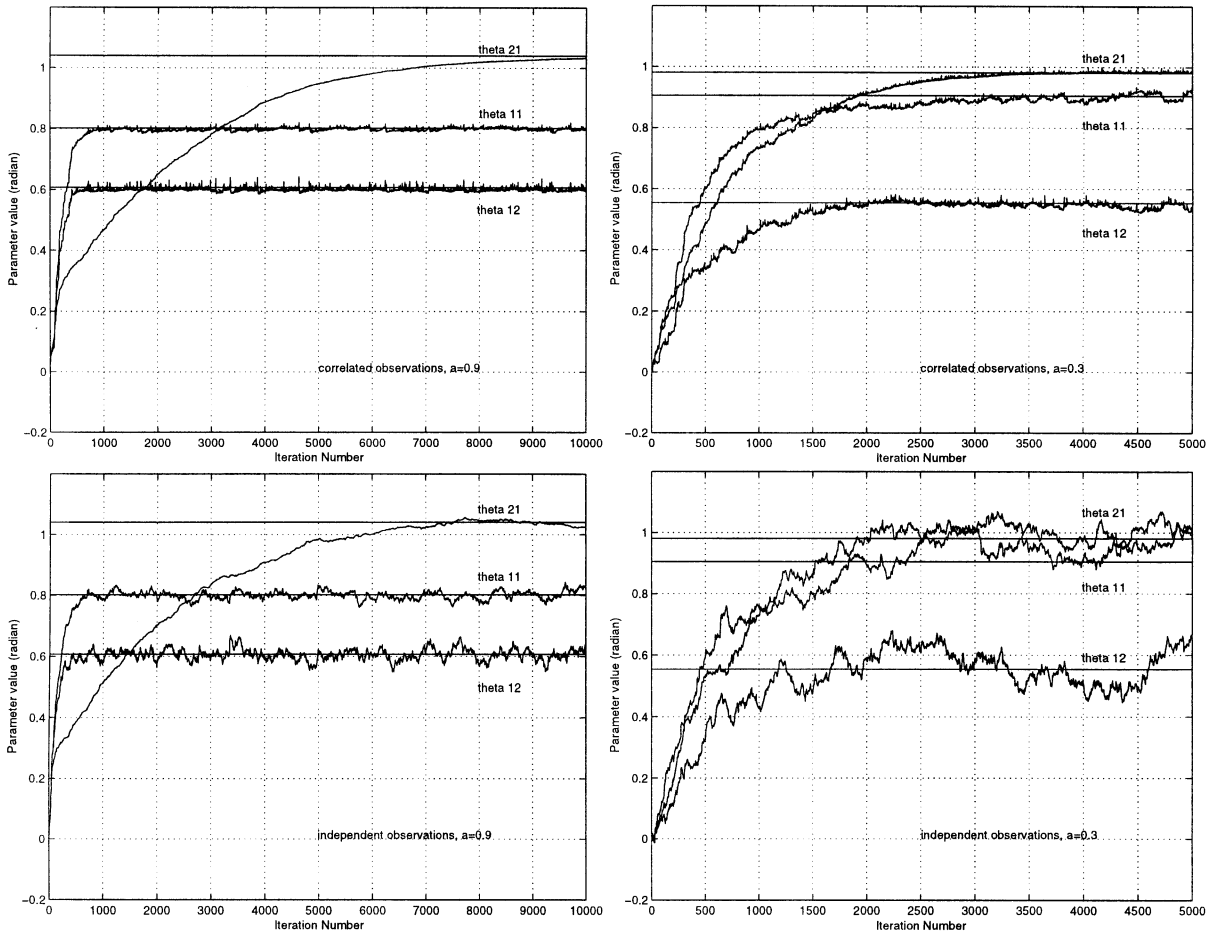
Fig. 4. Learning curves of the parameters $\theta_{1,1}$, $\theta_{1,2}$ and $\theta_{2,1}$ from the algorithm (7) for correlated or independent observations $\mathbf{x}_k$, the covariance matrix $\mathbf{\Gamma}_x$ being issued from an AR(1) with coefficient $a = 0.9$ or $a = 0.3$.

Firstly, let us present the case of a 3×3 covariance matrix $\mathbf{\Gamma}_x$ obtained from independent or correlated observations $\mathbf{x}_k$. Fig. 1 shows the mean square errors (35) of the first two estimated eigenvectors and of the associated estimated projection matrix (36) (both normalized by the gain factor γ). In each of these two figures, two distinct cases are considered: the underlying covariance matrix $\mathbf{\Gamma}_x$ is always that of an AR(1) process of power unity of parameter a, the estimates of the eigenvectors and of the associated projection matrix are obtained either from independent observations $\mathbf{x}_k$ or from correlated observations $\mathbf{x}_k = [x_k, \ldots, x_{k-n+1}]^{\mathrm{T}}$ with x_k is an AR(1) process. The same is performed for the case

of MA consecutive observations in Fig. 2 as a function of the parameter b of the MA model of order 1.

We observe that these errors are a function of the eigenvalue spread. These misadjustments increase when the eigenvalue spread decreases: for the AR(1) model, these errors decrease when a increases, since the eigenvalue spread increases; and for the MA(1) model these errors are minimum when the eigenvalue spread $(1 + b^2 + b\sqrt{2})/(1 + b^2 - b\sqrt{2})$ is maximum, that is for $b = 1$. The values of the errors are between 10 dB and 20 dB worse for independent observations, than for AR or MA consecutive observations. For a given covariance matrix $\mathbf{\Gamma}_x$, the results are thus very sensitive to the

Fig. 5. Mean square error (normalized by the gain factor γ) of the first two eigenvectors as a function of the parameter r for independent observations, the 4×4 covariance $\boldsymbol{\Gamma}_x$ matrix being issued from an AR(2), in two cases: the symmetric-centrosymmetric structure of $\boldsymbol{\Gamma}_x$ is taken into account, and is not.
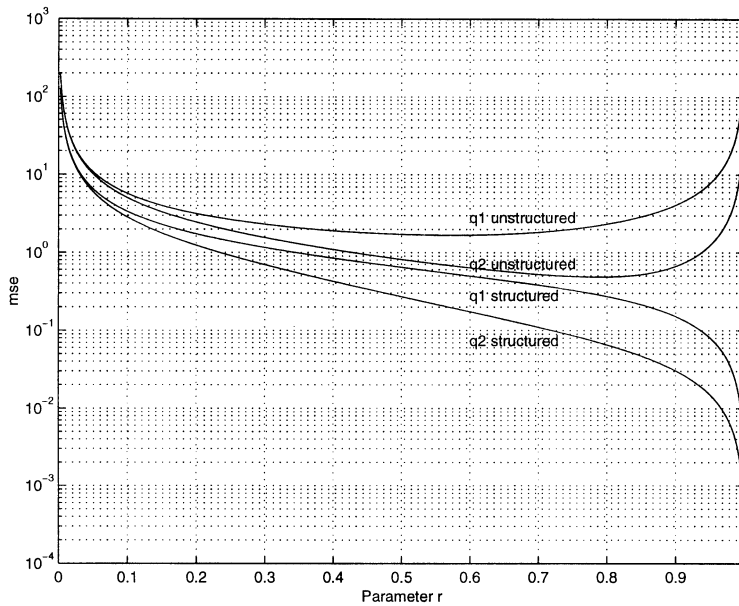
independence or correlation property of the consecutive observations $\boldsymbol{x}_k$ which gave rise to it. We observe that this misadjustment increases very slowly for the successive eigenvectors in case of independent observations, whereas it decreases in case of correlated observations.

A simulation of the algorithm (7) is presented where $\gamma = 0.004$. Fig. 3 shows the learning curves of the mean square error of the first two eigenvectors for $a = 0.9$ or $a = 0.3$ and for correlated or independent observations $\boldsymbol{x}_k$. It confirms the preceding results, in particular, it perfectly agrees with the theoretical values predicted by Fig. 1. The speed of convergence of the vectors $\boldsymbol{q}_1(\theta_{1,1},\theta_{1,2})$ and $\boldsymbol{q}_2(\theta_{1,1},\theta_{1,2},\theta_{2,1})$ can be explained by examining the different speeds of convergence of the parameters $\theta_{1,1}$, $\theta_{1,2}$ and $\theta_{2,1}$ shown in Fig. 4. The parameters $\theta_{1,1}$ and $\theta_{1,2}$ associated with $\boldsymbol{q}_1(\theta_{1,1},\theta_{1,2})$ converge faster with increasing a and the opposite happens for $\theta_{2,1}$ associated with $\boldsymbol{q}_2(\theta_{1,1},\theta_{1,2},\theta_{2,1})$. Because $\lambda_1 - \lambda_2 \approx 0.46$ and $\lambda_2 - \lambda_3 \approx 0.29$ for $a = 0.3$ [respectively 2.55 and 0.13 for $a = 0.9$], $\lambda_1 - \lambda_2$ increases but $\lambda_2 - \lambda_3$ decreases when a increases from 0.3 to 0.9. Thus, as far as the speed of convergence is

concerned, the eigenvalue spread is too global a parameter. The speed of convergence depends on the gaps between successive eigenvalues. It increases when the gaps between successive eigenvalues increase. This result is intuitive: the larger the gap between successive eigenvalues, the better the conditioning of eigenvectors and the faster the convergence. We note that, despite the different values of misadjusment, the speed of convergence is not affected by the origin of the covariance matrix. This latter result is confirmed by the interpretation of the associated ODE. Since the gain factor γ is 'small', the algorithm (7) follows its ODE (8) from the start, so that the transient regime is completely described by its ODE which is invariant to the type of correlation between successive observations. This result is also in accordance with the study of the convergence speed performed by Yang and Kaveh [21] where they distinguish only the origin of the covariance matrices by switches that are set in their eigenspace-linear combiner to one position for sensor array data (independent observations) and to some other position for time series data (correlated observations).
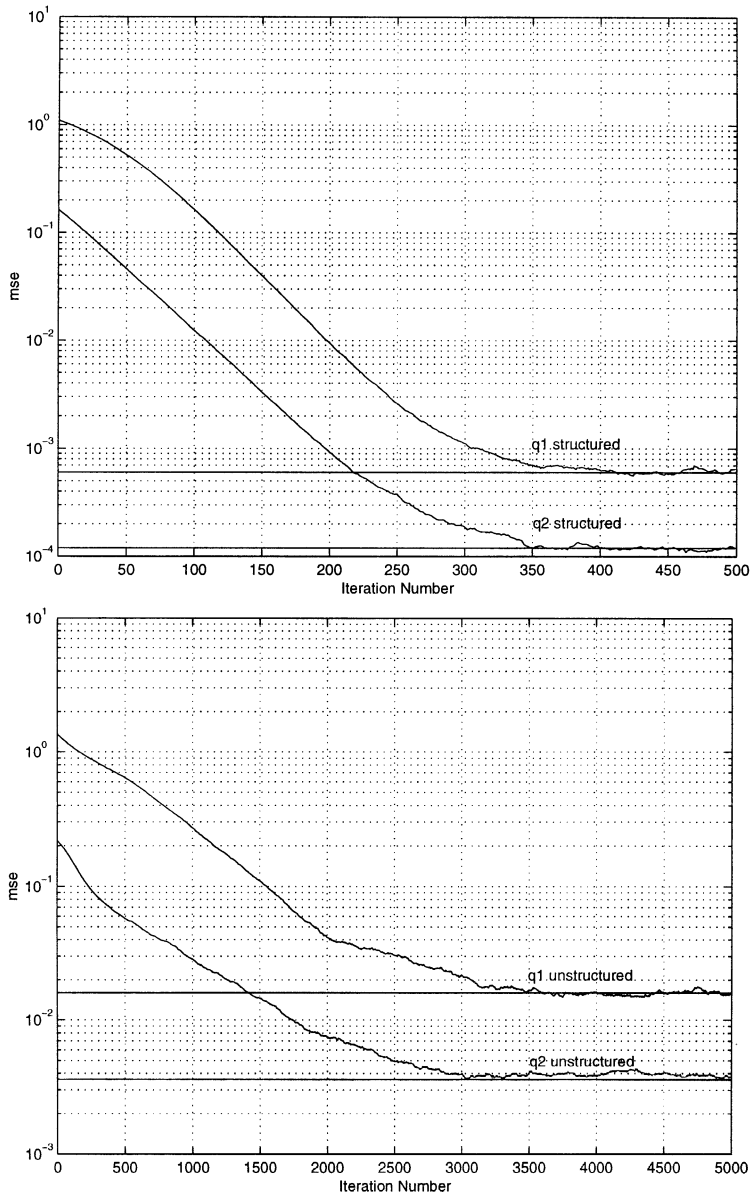
Fig. 6. Learning curves of $\mathrm{E}\|\boldsymbol{q}_1(k) - \boldsymbol{q}_{1*}\|_{\mathrm{Fro}}^2$ and $\mathrm{E}\|\boldsymbol{q}_2(k) - \boldsymbol{q}_{2*}\|_{\mathrm{Fro}}^2$ (averaged on 400 runs) from the algorithm (7) with $\gamma = 0.004$ for consecutive independent observations $\boldsymbol{x}_k$, the 4×4 covariance $\boldsymbol{\Gamma}_x$ matrix being issued from an AR(2) parameterized by $r = 0.9$ and asymptotical mean square error when the symmetric-centrosymmetric structure of $\boldsymbol{\Gamma}_x$ is taken into account or not.

Next, we present the case of a 4×4 covariance matrix $\boldsymbol{\Gamma}_x$ of an AR(2) processs. The AR model of order 2 has two poles $r \exp(ib)$ and $r \exp(-ib)$ with $b = \pi/4$. Fig. 5 shows the mean square error (normalized by the gain factor γ) of the first two eigen-

vectors, as a function of the parameter r, for independent observations, when the symmetric-centrosymmetric structure of $\boldsymbol{\Gamma}_x$ is taken into account or not. We notice that when the structure of $\boldsymbol{\Gamma}_x$ is not taken into account, the mean square error of
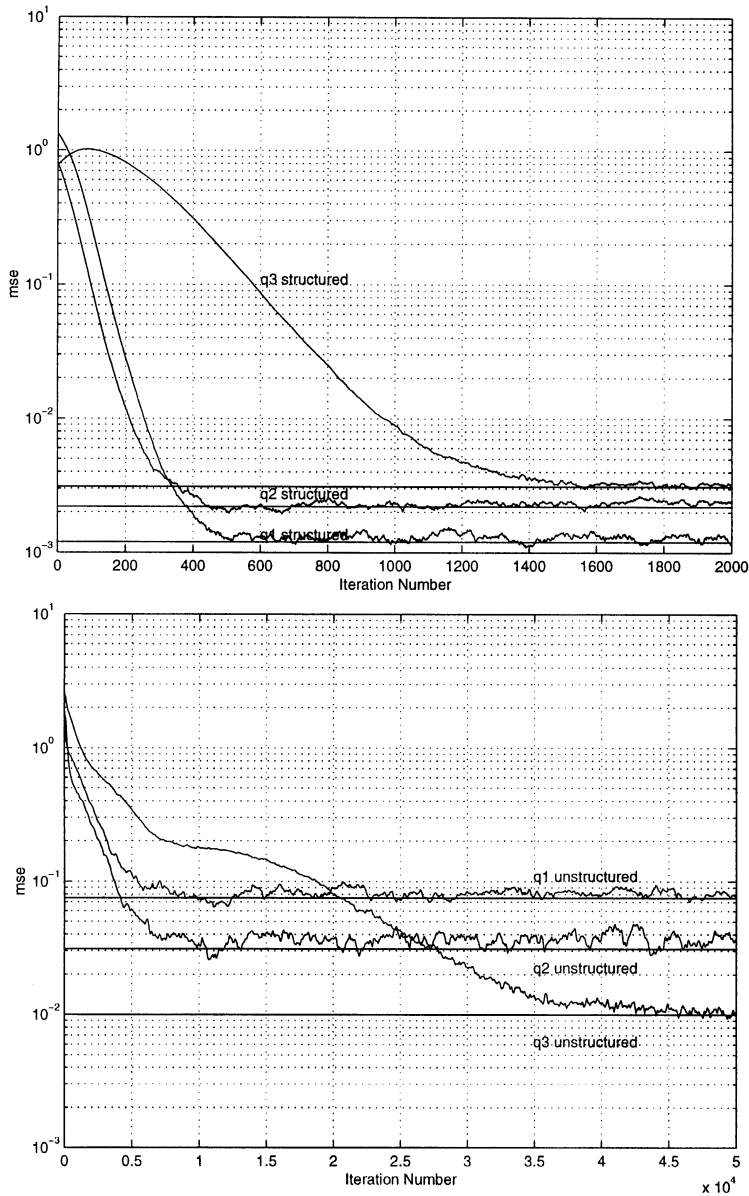
Fig. 7. Learning curves of $E\|\boldsymbol{q}_1(k) - \boldsymbol{q}_{1*}\|^2_{\text{Fro}}$, $E\|\boldsymbol{q}_2(k) - \boldsymbol{q}_{2*}\|^2_{\text{Fro}}$ and $E\|\boldsymbol{q}_3(k) - \boldsymbol{q}_{3*}\|^2_{\text{Fro}}$ (averaged on 400 runs) from the algorithm (7) with $\gamma = 0.004$ for consecutive independent observations $\boldsymbol{x}_k$, the 8×8 covariance $\boldsymbol{\Gamma}_x$ matrix being issued from an AR(2) parameterized by $r = 0.9$ and asymptotical mean square error when the symmetric-centrosymmetric structure of $\boldsymbol{\Gamma}_x$ is taken into account or not.

the first eigenvector is minimum for $r \approx 0.55$, which corresponds to the maximum of λ_1/λ_2, whereas that of the second eigenvector is minimum for $r \approx 0.80$, which corresponds to a trade-off between the contribution of $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, the covariances of which

are related respectively to λ_1/λ_2 and to λ_2/λ_3. And when the structure of $\boldsymbol{\Gamma}_x$ is taken into account, since the eigenvalues of $\boldsymbol{\Gamma}_1 + \boldsymbol{J}\boldsymbol{\Gamma}_2$ (respectively $\boldsymbol{\Gamma}_1 - \boldsymbol{J}\boldsymbol{\Gamma}_2$) are λ_1 and λ_3 (respectively λ_2 and λ_4), the mean square errors of $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ are decreasing functions

of the ratio λ_1/λ_3 (respectively λ_2/λ_4). As these ratios are an increasing function of the parameter r, these errors are a decreasing function of r. We see that the mean square errors are smaller in this latter case. This advantage increases with r. This fact is explained by the substitution of the ratios λ_1/λ_2 and λ_2/λ_3 by λ_1/λ_3 and λ_2/λ_4, respectively, which determines the behavior of the covariances of $\boldsymbol{\Theta}_1$ and $\boldsymbol{\Theta}_2$, and by the reduction of the number of parameters $\boldsymbol{\Theta}$ (here 2 versus 5). Fig. 6, obtained for the same value of $\gamma = 0.004$, shows that the misadjustments agree with the theoretical values predicted by Fig. 5 and that the speed of convergence is improved when we take into account the structure of the eigenvectors induced by the symmetric-centrosymmetric structure of the covariance matrix. This advantage carries over to higher values of n provided that the ratios of the successive eigenvalues of $\boldsymbol{\Gamma}_1 + \boldsymbol{J\Gamma}_2$ and of $\boldsymbol{\Gamma}_1 - \boldsymbol{J\Gamma}_2$ are larger than the associated ratios of the successive eigenvalues of $\boldsymbol{\Gamma}_x$. So Fig. 7 shows the learning curves of the first three eigenvectors of the 8×8 covariance matrix $\boldsymbol{\Gamma}_x$ of the same AR(2) process for which the first three eigenvectors of $\boldsymbol{\Gamma}_1 + \boldsymbol{J\Gamma}_2$ are 3.10 0.83 0.12) and the first two eigenvectors of $\boldsymbol{\Gamma}_1 - \boldsymbol{J\Gamma}_2$ are (3.39 0.45).

In conclusion of these simulations, we can draw a contrast between the effect of eigenvalue spread on the convergence of our stochastic gradient algorithm against the stochastic gradient algorithm used for estimating the linear prediction parameters. An increasing of the gaps between successive eigenvalues improves both the mean square error and the speed of convergence, while in the linear prediction stochastic gradient algorithm, it has no direct effect on the mean square error (which is $n\gamma\sigma^2/2$ where σ^2 denotes the variance of the error prediction) and worsens the speed of convergence.

7. Conclusion

In this paper, we presented a convergence study of a Givens parametrized adaptive eigensubspace algorithm, based on the stability of the associated ODE. Using a convergence rate result of Benveniste et al. and a continuity theorem, we gave the asymptotic distribution of the estimated eigenvectors and projection matrices on eigenspaces and evaluated their misadjustments. We analysed the effect of the eigenvalue spread on the mean square error and on convergence speed by simulations.

We showed that these misadjustments are sensitive to the correlation between successive observations. In particular, we found that these are smaller when the observations are correlated than when the observations are uncorrelated for a covariance matrix of an AR(1) or MA(1) stationary process. We observed the same surprising result in block estimation. Unfortunately, this result cannot be extended to an arbitrary stationary process. On the other hand, simulations showed that the convergence speed is not affected by the correlation between successive observations.

Finally, we proposed to improve the tradeoff between the misadjustment and the convergence speed by exploiting the symmetric-centrosymmetric property of some covariance matrices. These results are confirmed by simulations.

Appendix A. Proof of the relation (17), (18) and (19)

$\boldsymbol{G}_{11}$, $\boldsymbol{G}_{22}$, $\boldsymbol{G}_{21}$ are respectively equal to

$$\boldsymbol{G}_{11} = \frac{\mathrm{d}\boldsymbol{q}_1^{\mathrm{T}}(\boldsymbol{\Theta}_1)_*}{\mathrm{d}\boldsymbol{\Theta}_1}\boldsymbol{\Gamma}_x\frac{\mathrm{d}\boldsymbol{q}_1(\boldsymbol{\Theta}_1)_*}{\mathrm{d}\boldsymbol{\Theta}_1}$$
$$+ \left[\frac{\partial \boldsymbol{A}_1}{\partial \theta_{1,1}}\boldsymbol{\Gamma}_x\boldsymbol{q}_1(\boldsymbol{\Theta}_1), \ldots, \frac{\partial \boldsymbol{A}_1}{\partial \theta_{1,n-1}}\boldsymbol{\Gamma}_x\boldsymbol{q}_1(\boldsymbol{\Theta}_1)\right]_*,$$

$$\boldsymbol{G}_{22} = \frac{\partial \boldsymbol{q}_2^{\mathrm{T}}(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2)_*}{\partial \boldsymbol{\Theta}_2}\boldsymbol{\Gamma}_x\frac{\partial \boldsymbol{q}_2(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2)_*}{\partial \boldsymbol{\Theta}_2}$$
$$+ \left[\frac{\partial \boldsymbol{A}_2}{\partial \theta_{2,1}}\boldsymbol{\Gamma}_x\boldsymbol{q}_2(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2), \ldots, \right.$$
$$\left. \frac{\partial \boldsymbol{A}_2}{\partial \theta_{2,n-2}}\boldsymbol{\Gamma}_x\boldsymbol{q}_2(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2)\right]_*,$$

$$\boldsymbol{G}_{21} = \frac{\partial \boldsymbol{q}_2^{\mathrm{T}}(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2)_*}{\partial \boldsymbol{\Theta}_2}\boldsymbol{\Gamma}_x\frac{\partial \boldsymbol{q}_2(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2)_*}{\partial \boldsymbol{\Theta}_1}$$
$$+ \left[\frac{\partial \boldsymbol{A}_2}{\partial \theta_{1,1}}\boldsymbol{\Gamma}_x\boldsymbol{q}_2(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2), \ldots, \right.$$
$$\left. \frac{\partial \boldsymbol{A}_2}{\partial \theta_{1,n-1}}\boldsymbol{\Gamma}_x\boldsymbol{q}_2(\boldsymbol{\Theta}_1,\boldsymbol{\Theta}_2)\right]_*,$$

with $A_1 \stackrel{\text{def}}{=} \mathrm{d}q_1^{\mathrm{T}}(\Theta_1)/\mathrm{d}\Theta_1 = D_1(\Theta_1)Q_1'^{\mathrm{T}}(\Theta_1)$ and $A_2 \stackrel{\text{def}}{=} \partial q_2^{\mathrm{T}}(\Theta_1,\Theta_2)/\partial\Theta_2 = D_2(\Theta_2)Q_2'^{\mathrm{T}}(\Theta_2)Q_1'^{\mathrm{T}}(\Theta_1)$. Since $\|q_1(\Theta_1)\|^2 = \|q_2(\Theta_1,\Theta_2)\|^2 = 1$ holds for all Θ_1 and Θ_2, this implies $A_1 q_1 = 0$ and $A_2 q_2 = 0$ also holds for all Θ_1 and Θ_2 and therefore

$$A_i \frac{\mathrm{d}q_i}{\mathrm{d}\Theta_j} + \left[\frac{\partial A_i}{\partial\theta_{j,1}}q_i,\dots,\frac{\partial A_i}{\partial\theta_{j,n-1}}q_i\right] = O$$

$$\Rightarrow \left[\frac{\partial A_i}{\partial\theta_{j,1}}\Gamma_x q_i,\dots,\frac{\partial A_i}{\partial\theta_{j,n-1}}\Gamma_x q_i\right]_* = -\lambda_i A_{i*}\frac{\mathrm{d}q_i}{\mathrm{d}\Theta_j}*$$

for respectively $i, j = 1,1, 2,2$ and $2,1$. Putting all the pieces together, we get the expressions (17), (18) and (19), where $Q_{1,2}(\Theta_1,\Theta_2) = \partial q_2(\Theta_1,\Theta_2)/\partial\Theta_1$ with

$$\frac{\partial q_2(\Theta_1,\Theta_2)}{\partial\Theta_1} = \left[q_2\left(\theta_{1,1}+\frac{\pi}{2},\theta_{1,2},\dots,\theta_{1,n-1},\Theta_2\right),\dots,\right.$$

$$\left. q_2\left(\theta_{1,1},\dots,\theta_{1,n-2},\theta_{1,n-1}+\frac{\pi}{2},\Theta_2\right)\right].$$

We note that unlike $\partial q_i(\Theta_1,\dots,\Theta_i)/\partial\Theta_i$, the term $\partial q_i(\Theta_1,\dots,\Theta_i)/\partial\Theta_j$ for $j < i$ cannot be easily expressed as a product of matrices.

Appendix B. Proof of the relation (23), (24) and (25)

We begin with the general Gaussian property

$$E\{x_k[x_k^{\mathrm{T}}q_i][q_j^{\mathrm{T}}x_0]x_0^{\mathrm{T}}\}$$

$$= E\{x_k[x_k^{\mathrm{T}}q_i]\}E\{q_j^{\mathrm{T}}x_0]x_0^{\mathrm{T}}\}$$

$$+ E\{x_k[x_0^{\mathrm{T}}q_j]\}E\{q_i^{\mathrm{T}}x_k]x_0^{\mathrm{T}}\}$$

$$+ E\{[q_i^{\mathrm{T}}x_k][x_0^{\mathrm{T}}q_j]\}E\{x_k x_0^{\mathrm{T}}\} \qquad (48)$$

applied with $k = 0$ and $i = 1,2$ and $j = 1,2$. To simplify notations, we denote $q_1(\Theta_{1*})$ and $q_2(\Theta_{1*},\Theta_{2*})$ as q_1 and q_2, respectively, where

$$E\{x_0[x_0^{\mathrm{T}}q_i]\}E\{[q_j^{\mathrm{T}}x_0]x_0^{\mathrm{T}}\} = \lambda_i\lambda_j q_i q_j^{\mathrm{T}}$$

and

$$E\{[q_i^{\mathrm{T}}x_0][x_0^{\mathrm{T}}q_j]\}E\{x_0 x_0^{\mathrm{T}}\} = \delta_{i,j}\lambda_i\Gamma_x$$

with $\delta_{i,j} = 0$ if $i = j$ and 1 elsewhere.

We then obtain

$$E\{x_0[x_0^{\mathrm{T}}q_i][q_j^{\mathrm{T}}x_0]x_0^{\mathrm{T}}\} = \lambda_i\lambda_j(q_i q_j^{\mathrm{T}} + q_j q_i^{\mathrm{T}}) + \delta_{i,j}\lambda_i\Gamma_x.$$

Therefore, thanks to the orthonormal properties of $\{q_1(\Theta_1),q_2(\Theta_1,\Theta_2)\}$, which imply $q_1^{\mathrm{T}}(\partial q_2/\partial\Theta_2) = 0$ and $q_i^{\mathrm{T}}(\partial q_i/\partial\Theta_i) = 0$ for $i = 1,2$, the relations (22), (14) and (15) gives the relations (23), (24) and (25).

Appendix C. Proof of the relations (29) and (30)

In the case of correlated observations x_k, the relations (12) and (21), (22) imply

$$R_{ij} = \frac{\partial q_i^{\mathrm{T}}(\Theta)_*}{\partial\Theta_i}\sum_{k=-\infty}^{+\infty}E[x_k x_k^{\mathrm{T}}q_i(\Theta_*)q_j^{\mathrm{T}}(\Theta_*)x_0 x_0^{\mathrm{T}}]$$

$$\times\frac{\partial q_j(\Theta)_*}{\partial\Theta_j}, \qquad (49)$$

where from the Gaussian property (48) gives

$$E\{x_k[x_k^{\mathrm{T}}q_i][q_j^{\mathrm{T}}x_0]x_0^{\mathrm{T}}\}$$

$$= \lambda_i\lambda_j q_i q_j^{\mathrm{T}} + \Gamma_k q_j q_i^{\mathrm{T}}\Gamma_k + [q_i^{\mathrm{T}}\Gamma_k q_j]\Gamma_k,$$

where Γ_k denotes the cross-correlation matrix $E[x_k x_0^{\mathrm{T}}]$. Since $\Gamma_{-k} = \Gamma_k^{\mathrm{T}}$, the terms $\Gamma_{i,j}$ of Eqs. (26)–(28) become

$$\Gamma_{i,j} = \sum_{k=1}^{+\infty}\Gamma_k q_j q_i^{\mathrm{T}}\Gamma_k + \Gamma_k^{\mathrm{T}}q_j q_i^{\mathrm{T}}\Gamma_k^{\mathrm{T}} + (q_i^{\mathrm{T}}\Gamma_k q_j)\Gamma_k$$

$$+ (q_i^{\mathrm{T}}\Gamma_k^{\mathrm{T}}q_j)\Gamma_k^{\mathrm{T}}. \qquad (50)$$

This infinite sum reduces to the finite sum (29) for an MA process, whereas for an AR process we use the property that the correlation coefficients $\gamma_i \stackrel{\text{def}}{=} E(x_k x_{k-i})$ can be written as ([9] p. 88 (2.64))

$$\gamma_k = \sum_{l=1}^{p}\alpha_l\rho_l^{|k|}, \qquad (51)$$

where $\rho_1,\dots,\rho_p$ are the roots (we have supposed that these roots are simple) of the characteristic equation associated to the AR(p) process, and $\alpha_1,\dots,\alpha_p$ are constants determined by the values of $\gamma_0,a_1,\dots,a_p$ ($a_k, k = 1,\dots,p$ denote the coefficients of the AR(p) process). Since

$$\Gamma_k = \sum_{l=1}^{p}\alpha_l\rho_l^k\Gamma_{\rho_l}, \quad \text{for } k > n-2,$$

with $|\rho_l| < 1$, where Γ_{ρ_l} denotes the $n \times n$ matrix, the entries of which are $(\Gamma_{\rho_l})_{i,j} = \rho_l^{j-i}$, the term (50) also reduces to the finite sum (30). As for an ARMA(p,q)

process, the relation (51) still applies, but only for $|k| \geqslant \text{Sup}(0, q + 1 - p)$. Therefore the relation (30) also holds, provided n is replaced by $n + \text{Sup}(0, q + 1 - p)$.

Appendix D. Proof of the relations (44) and (45)

For independent observations x_k, we have

$$R_{ij} = \frac{\mathrm{d}u_i^{\mathrm{T}}(\Theta_i)_*}{\mathrm{d}\Theta_i} \mathrm{E}[y_i y_i^{\mathrm{T}} u_i(\Theta_{i*}) u_j^{\mathrm{T}}(\Theta_{j*}) y_j y_j^{\mathrm{T}}] \frac{\mathrm{d}u_j(\Theta_j)_*}{\mathrm{d}\Theta_j},$$

$$\text{for } i = 1, 2, \ j = 1, 2, \tag{52}$$

with $y_i \stackrel{\text{def}}{=} x_k' + \varepsilon_i J x_k'$ and $\varepsilon_1 = -\varepsilon_2$. From the Gaussian property (48), using the same development of Appendix B there follows

$$\mathrm{E}\{y_i[y_i^{\mathrm{T}} u_i][u_i^{\mathrm{T}} y_i] y_i^{\mathrm{T}}\}$$

$$= 8\lambda_i^2 u_i u_i^{\mathrm{T}} + 4\lambda_i(\Gamma_1 + \varepsilon_i J \Gamma_2), \quad \text{for } i = 1, 2,$$

and as $\mathrm{E}(y_2 y_1^{\mathrm{T}}) = (1 + \varepsilon_1 \varepsilon_2)\Gamma_1 + (\varepsilon_1 + \varepsilon_2) J \Gamma_2 = O$,

$$\mathrm{E}\{y_i[y_i^{\mathrm{T}} u_i][u_j^{\mathrm{T}} y_j] y_j^{\mathrm{T}}\} = 4\lambda_i \lambda_j u_i u_j^{\mathrm{T}}, \quad \text{for } i \neq j.$$

Therefore, thanks to the unit norm of $u_1(\Theta_1)$ and $u_2(\Theta_2)$, the relations (44) and (45) are deduced.

References

[1] R. Bellman, Stability Theory of Differential Equations, McGraw Hill, New York, 1953.

[2] A. Benveniste, M. Métivier, P. Priouret, Adaptive Algorithms and Stochastic Approximations, Springer, Berlin, 1990.

[3] J.R. Bunch, C.P. Nielsen, D. Sorenson, Rank-one modification of a symmetric eigenproblem, Numerische Mathematik 31 (1978) 31–48.

[4] A. Cantoni, P. Butler, Eigenvalues and eigenvectors of symmetric centrosymmetric matrices, Linear Algebra and its Applications 13 (1976) 275–288.

[5] B. Champagne, A new adaptive eigendecomposition algorithm based on a first order perturbation, in: Proc. ICASSP, San Fransisco, May 1992, pp. 409–412.

[6] J.P. Delmas, A complex adaptive eigensubspace algorithm for DOA or frequency estimation and tracking, in: Proc. EUSIPCO 1992, pp. 657–660.

[7] J.P. Delmas, Performance analysis of parametrized adaptive eigensubpace algorithms, in: Proc. ICASSP, Detroit, May 1995, pp. 2056–2059.

[8] J.C. Fort, G. Pagès, Convergence of stochastic algorithms: from the Kushner and Clark theorem to the Lyapunov functional method, Advances in Applied Probability 28 (December 1996) 1072–1094.

[9] C.S. Haykin, Adaptive Filter Theory, Prentice-Hall, Englewood Cliffs, 1991.

[10] H.J. Kushner, Weak convergence methods and singularly perturbed stochastic control and filtering problems, systems and control: Foundations and Applications, Vol. 3, Birkhauser, Basel, 1990.

[11] M.G. Larimore, R.J. Calvert, Convergence studies of Thompson's unbiased adaptive spectral estimator, in: Proc. 14th Asilomar Conf. Circuits Syst. Comput., Pacific Grove, CA, November 1980, pp. 258–262.

[12] L. Ljung, T. Söderström, Theory and Practice of Recursive Identification, MIT Press, Cambridge, MA, 1983.

[13] G. Mathew, V.U. Reddy, S. Dasgupta, Adaptive estimation of eigensubspace, IEEE Trans. Signal Process. 43 (2) (February 1995) 401–411.

[14] E. Moulines, P. Duhamel, J.F. Cardoso, S. Mayrargue, Subspace methods for blind identification of multichannel FIR filters, IEEE Trans. Signal Process. 43 (2) (February 1995) 516–525.

[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, Journal of Math. Analysis and Applications 106 (1985) 69–84.

[17] C.R. Rao, Linear Statistical Inference and Its Applications, Wiley, New York, 1973.

[18] P.A. Regalia, An adaptive unit norm filter with applications to signal analysis and Karhunen Loéve tranformations, IEEE Trans. on Circuits and Systems 37 (5) (May 1990) 646–649.

[19] C. Riou, T. Chonavel, P.Y. Cochet, Adaptive subspace estimation – Application to moving sources localization and blind channel identification, Proc. ICASSP, Atlanta, May 1996, pp. 1649–1652.

[20] G. Xu, R.H. Roy, T. Kailath, Detection of number of sources via exploitation of centro-symmetry property, IEEE Trans. Signal Process. 42 (1) (January 1994) 102–112.

[21] J.F. Yang, M. Kaveh, Adaptive eigenspace algorithms for direction or frequency estimation and tracking, IEEE Trans. on ASSP 36 (2) (February 1988) 241–251.

[22] B. Yang, Projection approximation subspace tracking, IEEE Trans. Signal Process. 43 (1) (January 1995) 95–107.

[23] B. Yang, Convergence analysis of the subspace tracking algorithms PAST and PASTd, in: Proc. ICASSP, Atlanta, May 1996, pp. 1760–1763.

[24] B. Yang, F. Gersemsky, Asymptotic distribution of recursive subspace estimators, in: Proc. ICASSP, Atlanta, May 1996, pp. 1764–1767.