PERFORMANCES ANALYSIS OF PARAMETERIZED ADAPTIVE EIGENSUBSPACE ALGORITHMS

Jean Pierre Delmas

Département SIM, Institut National des Télécommunications,
9 rue Charles Fourier, 91011 Evry cedex, FRANCE
e-mail: delmas@int-evry.fr

ABSTRACT

In this paper, we address adaptive estimation methods of eigenspaces of covariance matrices. We are interested in methods based on several coupled maximizations or minimizations of Rayleigh ratios where the constraints are replaced by appropriate parameterizations (Givens and mixed Givens/Householder). We prove the convergence of these algorithms with the help of the associated Ordinary Differential Equation, and we propose an evaluation of the performances by computing the variances of the estimated eigenvectors for fixed gain factors. We show that these variances are very sensitive to the difference between two consecutive eigenvalues. Moreover, they also 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). Finally we show that the performances can be improved when the centro-symmetric property of some of those covariance matrices is taken into account.

1.INTRODUCTION

For the last ten years, adaptive estimation of covariance matrices has been applied successfully to both temporal and spatial domain high-resolution spectral analysis. The interest for these methods, a tool of outstanding importance in many fields of Signal Processing, has recently been renewed by the subspace approach used in blind identification of multichannel FIR filters [1]. Among the solutions that propose to recursively update the eigendecomposition of a covariance matrice, we are interested here in methods derived from constrained optimizations. These constrained optimizations can be performed adaptively by a stochastic gradient search over time where the constraints are taken into account by a Gram-Schmidt orthogonalization at each iteration [2]. To get rid of these constraints, an alternate solution consists in using an appropriate parameterization [3]. Until now, only simulations attested the convergence of the stochastic coupled gradient based adaptive algorithms constructed on these parameters. We essentially propose in this paper to study the convergence and the performances of these methods by introducing the necessary methodology and exploiting some of the results that can be derived from it.

This paper is organized as follows. After introducing some notations and describing the parameterization of the orthonormal eigenvectors of the covariance matrices in Section 2, we prove the convergence of the coupled stochastic gradient algorithms with the help of the associated Ordinary Differential Equation in Section 3. An evaluation of the performances by computing the variances of the estimated eigenvectors for fixed gain factors is given in Section 4. And finally, we show that the performances can be improved when the centro-symmetric property of some of those covariance matrices is taken into account in Section 5.

2.PARAMETERIZATION OF THE PROBLEM

We tackle the problem of an adaptive estimation of the m eigenvectors $\mathbf{q}_1,..,\mathbf{q}_m$ corresponding to the m largest (or smallest) eigenvalues ($\lambda_1\geq\lambda_2..\geq\lambda_m$) of a $n\times n$ covariance matrix $\mathbf{\Gamma}_x=\mathrm{E}[\mathbf{x}^*(k)\mathbf{x}^T(k)]$ from different realizations $\mathbf{x}(k)$ of a random vector $\mathbf{x}$ ($*,T,H$ stand respectively for conjugate, transpose and conjugate transpose). To solve this problem, a method was proposed in [3] and then extended to the complex case in [4], where the constrained maximizations of Rayleigh ratios:

$$\underset{\|\mathbf{q}_i\|=1}{\mathrm{Max}}\ \mathbf{q}_i^H\mathbf{\Gamma}_x\mathbf{q}_i \quad \text{and} \quad \underset{\|\mathbf{q}_i\|=1\ \mathbf{q}_i\perp \mathrm{sp}\{\mathbf{q}_1,..,\mathbf{q}_{i-1}\}}{\mathrm{Max}}\ \mathbf{q}_i^H\mathbf{\Gamma}_x\mathbf{q}_i$$

for $i=2,..,m$, (1a) and (1b)

that are taken into account in [2] by a Gram-Schmidt orthogonalization are replaced by unconstrained maximizations thanks to a Givens parameterization of the different constraints. $\mathbf{q}_1$ is the last column of a unitary matrix $\mathbf{Q}_1$ and the other vectors $\mathbf{q}_i$ can be written:

$$\mathbf{q}_1=\mathbf{Q}_1\begin{bmatrix}0\\1\end{bmatrix},\ \mathbf{q}_2=\mathbf{Q}_1\begin{bmatrix}\mathbf{Q}_2\begin{bmatrix}0\\1\end{bmatrix}\\0\end{bmatrix},..$$

$$..,\mathbf{q}_m=\mathbf{Q}_1\begin{bmatrix}\mathbf{Q}_2\begin{bmatrix}\begin{bmatrix}\mathbf{Q}_m\begin{bmatrix}0\\1\end{bmatrix}\\0\end{bmatrix}\\0\end{bmatrix}\\0\end{bmatrix} \tag{2a}$$

where $\mathbf{Q}_i$ is the following unitary matrix of order $n-i+1$:
$\mathbf{Q}_i=\mathbf{U}_{i,1}..\mathbf{U}_{i,j}..\mathbf{U}_{i,n-i}$

$$\text{with } \mathbf{U}_{i,j}=\begin{bmatrix}\mathbf{I}_{j-1} & 0 & 0 & 0\\ 0 & -\sin\psi_{i,j} & \cos\psi_{i,j} & 0\\ 0 & e^{i\phi_{i,j}}\cos\psi_{i,j} & e^{i\phi_{i,j}}\sin\psi_{i,j} & 0\\ 0 & 0 & 0 & \mathbf{I}_{n-i-j}\end{bmatrix} \tag{2b}$$

and $\psi_{i,j}$ and $\phi_{i,j}$ belong to $]-\frac{\pi}{2},+\frac{\pi}{2}]$. This parameterization is unique except when the first component that appear in

$$\mathbf{Q}_i\begin{bmatrix}0\\1\end{bmatrix} \text{ for } i=1,..,m \text{ is non zero}$$

The maximization (1a) is performed with the help of the classical stochastic gradient algorithm, in which the parameters are $\psi_{1,1},\phi_{1,1},..,\psi_{1,n-1},\phi_{1,n-1}$, whereas the maximizations (1b) are realized thanks to stochastic gradient algorithms with respect to the parameters $\psi_{i,1},\phi_{i,1},..,\psi_{i,n-1}$ $\phi_{i,n-i-1}$, in which the preceding parameters $\psi_{l,1}(k),\phi_{l,1}(k),..,\psi_{l,n-l}(k),\phi_{l,n-l}(k)$ for $l=1,..i-1$ 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 simulations results [3]. 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 Section 2 comes from the existence of coupled algorithms. In order to study the convergence, the coupled stochastic gradient algorithms need to be globally written as :

$$\begin{bmatrix} \theta_1 \\ \cdot \\ \theta_m \end{bmatrix}_{k+1} = \begin{bmatrix} \theta_1 \\ \cdot \\ \theta_m \end{bmatrix}_k + \gamma_k \begin{bmatrix} g_1(\theta_1(k),x(k)) \\ \cdot \\ g_m(\theta_1(k),..,\theta_m(k),x(k)) \end{bmatrix} \quad (3a)$$

with $\theta_i \triangleq [\psi_{i,1},\phi_{i,1},..,\psi_{i,n-i},\phi_{i,n-i}]^T$ and

$$g_1(\theta_1,x) \triangleq \nabla_{\theta_1}(|q_1^T x|^2),..,g_m(\theta_1,..,\theta_m,x) \triangleq \nabla_{\theta_m}(|q_m^T x|^2),$$

or more compactly :

$$\theta(k+1) = \theta(k) + \gamma_k H[\theta(k),x(k)] \text{ with } \theta \triangleq [\theta_1^T,..,\theta_m^T]^T \quad (3b)$$

The study of the convergence of the coupled stochastic gradient algorithms (3b) is intimately connected to the stability properties of the associated Ordinary Differential Equation (ODE) introduced by Ljung [5]:

$$\frac{d\theta(t)}{dt} = h[\theta(t)] \quad (4)$$

with $h(\theta) \triangleq E_\theta[H(\theta,x(k))]$.
More precisely, if the gain sequence γ_k satisfies the conditions

$$\sum_{k=1}^{+\infty} \gamma_k = +\infty \text{ and } \lim_{k\to+\infty} \gamma_k = 0,$$

we can apply a convergence result deduced from [6] (see theorem 2.3.1 p.39): if the ODE (4) admits a globally asymptotically stationary stable point θ_*, then $\theta(k)$ defined by (3b) converges almost surely to θ_*.
We can prove this result if we suppose that $m=2$, and the extension to $m>2$ is straightforward. Since $g_1(\theta_1,x)$ is the derivative of a positive gradient field, the stationary point of the block of (4) associated to θ_1, which is the solution of the maximization (1a), is globally asymptotically stationary stable for that equation. Furthermore the k^{th} component of $\frac{d\theta_1}{dt}$ verifies for $t\to+\infty$:

$$\left[\frac{d\theta_1}{dt}\right]_k \sim \alpha_k e^{-\mu_k t} \ (\mu_k > 0) \quad (5)$$

Consider the Lyapunov function:

$$W(t) \triangleq E|q_2^T(t)x(t)|^2 \leq \lambda_1$$

$$\frac{dW(t)}{dt} = \frac{d\theta_1^T}{dt} E[\nabla_{\theta_1}(|q_2^T(t)x(t)|^2)] + \frac{d\theta_2^T}{dt} E[\nabla_{\theta_2}(|q_2^T(t)x(t)|^2)]$$

by hypothesis we have:

$$\frac{d\theta_2^T}{dt} E[\nabla_{\theta_2}(|q_2^T(t)x(t)|^2)] = \|E[\nabla_{\theta_2}(|q_2^T(t)x(t)|^2)]\|^2$$

and since θ_1 and θ_2 are bounded, $E[\nabla_{\theta_1}(|q_2^T(t)x(t)|^2)]$ is also bounded. So that, thanks to (5), we have with $\alpha>0$ and $\mu>0$:

$$\left|\frac{d\theta_1^T}{dt} E[\nabla_{\theta_1}(|q_2^T(t)x(t)|^2)]\right| \leq \left\|\frac{d\theta_1}{dt}\right\| \, \|E[\nabla_{\theta_1}(|q_2^T(t)x(t)|^2)]\| \leq \alpha e^{-\mu t}$$

Consequently,

$$\frac{dW(t)}{dt} \geq -\alpha e^{-\mu t} + \|E[\nabla_{\theta_2}(|q_2^T(t)x(t)|^2)]\|^2$$

Then $W(t) - \frac{\alpha}{\mu} e^{-\mu t}$ is an increasing function of t, so

$$\lim_{t\to+\infty} W(t) \text{ exists,}$$

which implies

$$\lim_{t\to+\infty} \frac{dW(t)}{dt} = 0 \text{ and then } \lim_{t\to+\infty} E[\nabla_{\theta_2}(|q_2^T(t)x(t)|^2)] = 0.$$

Therefore the stationary point of the block of (4) associated to θ_2, which is the solution of the maximization (1b), is globally asymptotically stationary stable for that equation. We showed that the parameters that maximize the expressions (1) are globally asymptotically stable for its associated ODE, so that the convergence of the coupled stochastic algorithms (3a) is ensured.
Unfortunately, in nonstationary environments the gain sequences γ_k are 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 stricto senso. In this case, the algorithm does not converge almost surely any longer. However, the weak convergence approach developed by Kushner [7] suggests that, for γ "small enough", the adaptive algorithm will oscillate around the theoretical limit of the decreasing step scheme.
The key point in the above derivation is that the parameters are bounded, so that all these results still hold for similar algorithms using alternative parameterization of the orthonormal constraints, provided the boundedness condition is fulfilled. In particular, a similar class of algorithms can be derived exploiting an Householder factorization of unitary matrices [8]. The vectors $q_1,q_2..,q_m$ can be considered as the m last columns of the unitary matrix Q:

$$Q = \left[I_n - 2a_1a_1^H\right]\begin{bmatrix} I_{n-1}-2a_2a_2^H & 0 \\ 0^T & 1 \end{bmatrix} .. \begin{bmatrix} I_{n-m+1}-2a_ma_m^H & 0 \\ 0^T & I_{m-1} \end{bmatrix} \quad (6)$$

where each vector $a_i \in \mathbb{C}^{n+1-i}$ of unity norm is given by the Givens parameterization described previously. The maximizations (1) are also performed with the help of coupled stochastic gradient algorithms, but with a completely different deflation procedure, the convergence of which can be proved exactly along the same lines.

4.ASYMPTOTICAL VARIANCE CALCULUS

The variances of the estimated eigenvectors of the Givens and the Givens/Householder algorithms can be computed for fixed gain factors and in gaussian situations, thanks to the general result (see [9] theorem 2, p104) : when $\gamma \to 0$ and $t_k \to +\infty$ with $t_k \triangleq k\gamma$, then $1/\sqrt{\gamma}\,[\theta(k)-\theta(t_k)]$ converges in law to a zero mean gaussian random variable of covariance matrix $\mathbf{P}$ that is the unique symmetric solution of the continuous Lyapunov equation:

$$\mathbf{G}_*\mathbf{P}+\mathbf{P}\mathbf{G}_*^T+\mathbf{R}_*=\mathbf{0} \tag{7}$$

where

$$\mathbf{G} \triangleq \frac{\mathrm{d}\mathbf{h}(\theta)}{\mathrm{d}\theta} \text{ and } \mathbf{R} \triangleq \sum_{k=-\infty}^{+\infty} \mathrm{cov}_{\theta}[\mathbf{H}(\theta,\mathbf{x}(k)),\mathbf{H}(\theta,\mathbf{x}(0))] \tag{8}$$

and the subscript $_*$ stands for the value of the functions calculated for the parameter θ_* that maximizes the expressions (1). Then, if γ is "small enough" and k "large enough", $\theta(k)$ is unbiased gaussian and has a covariance matrix approximately equal to $\gamma\mathbf{P}$.

In the real case and for $m=1$, we have, thanks to the property:

$$\frac{\mathrm{d}\mathbf{q}_1}{\mathrm{d}\theta_1}=\mathbf{Q}_1'(\theta_1)\mathbf{D}_1(\theta_1) \tag{9}$$

with $\mathbf{Q}_1(\theta_1) \triangleq [\mathbf{Q}_1'(\theta_1),\mathbf{q}_1(\theta_1)]$ and with $\mathbf{D}_1(\theta_1)$ a $n-1\times n-1$ diagonal matrix where $\mathbf{D}_1(\theta_1)_{n-1,n-1}=1$ and $\mathbf{D}_1(\theta_1)_{k,k}=\prod_{l=k+1}^{n-1}\cos(\psi_{1,l})$ for $1\le k\le n-2$, the closed-form expression:

$$\mathbf{G}=\mathbf{D}_1(\theta_1)\mathbf{Q}_1'^T(\theta_1)(\Gamma_x-\lambda_1\mathbf{I}_n)\mathbf{Q}_1'(\theta_1)\mathbf{D}_1(\theta_1) \tag{10}$$

A closed-form expression for $\mathbf{R}$ is obtained for independent observations $\mathbf{x}$ (which corresponds, most of the time, to spatial situations):

$$\mathbf{R}=\mathbf{D}_1(\theta_1)\mathbf{Q}_1'^T(\theta_1)\lambda_1\Gamma_x\mathbf{Q}_1'(\theta_1)\mathbf{D}_1(\theta_1) \tag{11}$$

In the case of correlated observations $\mathbf{x}$ (which on the other hand corresponds to temporal situations), we obtain a different expression by examining particular cases. For a MA(q) process the term $\lambda_1\Gamma_x$ contained in (11) is replaced by:

$$\sum_{k=1}^{q+n-1}\Gamma_k\mathbf{q}_1\mathbf{q}_1^T\Gamma_k+\Gamma_k^T\mathbf{q}_1\mathbf{q}_1^T\Gamma_k^T+(\mathbf{q}_1^T\Gamma_k\mathbf{q}_1)(\Gamma_k+\Gamma_k^T) \tag{12}$$

where Γ_k denotes the cross-correlation matrix $\mathrm{E}(\mathbf{x}_k\mathbf{x}_0^T)$, where $\mathbf{x}_k \triangleq [x_k,x_{k-1},\ldots,x_{k-n+1}]^T$ and where $\mathbf{q}_1 \triangleq \mathbf{q}_1(\theta_1)$. And for an AR(1) process the same term is replaced by:

$$\sum_{k=1}^{n-2}\Gamma_k\mathbf{q}_1\mathbf{q}_1^T\Gamma_k+\Gamma_k^T\mathbf{q}_1\mathbf{q}_1^T\Gamma_k^T+(\mathbf{q}_1^T\Gamma_k\mathbf{q}_1)(\Gamma_k+\Gamma_k^T)+$$
$$\frac{\gamma_0^2a^{2n-2}}{1-a^2}[\Gamma_a\mathbf{q}_1\mathbf{q}_1^T\Gamma_a+\Gamma_a^T\mathbf{q}_1\mathbf{q}_1^T\Gamma_a^T+(\mathbf{q}_1^T\Gamma_a\mathbf{q}_1)(\Gamma_a+\Gamma_a^T)] \tag{13}$$

where $\gamma_0 \triangleq \mathrm{E}(x_k^2)$, $\gamma_1 \triangleq \mathrm{E}(x_kx_{k-1}) = a\gamma_0$ and Γ_a denotes the $n\times n$ matrix whose entries are $(\Gamma_a)_{i,j}=a^{j-i}$.

These results can be extended to $m=2$, the extension to $m>2$ is straightforward but tedious. The expression (10) becomes:

$$\mathbf{G}=\begin{bmatrix}\mathbf{G}_{11} & \mathbf{0}\\ \mathbf{G}_{21} & \mathbf{G}_{22}\end{bmatrix} \tag{14}$$

where $\mathbf{G}_{11}$ is given by (10), $\mathbf{G}_{22}$ by:

$$\mathbf{D}_2(\theta_2)\mathbf{Q}_2'^T(\theta_2)\mathbf{Q}_1'^T(\theta_1)(\Gamma_x-\lambda_2\mathbf{I}_n)\mathbf{Q}_1'(\theta_1)\mathbf{Q}_2'(\theta_2)\mathbf{D}_2(\theta_2)$$

with $\mathbf{Q}_2(\theta_2) \triangleq [\mathbf{Q}_2'(\theta_2),\mathbf{v}(\theta_2)]$ and with $\mathbf{D}_2(\theta_2)$ a $n-2\times n-2$ diagonal matrix where $\mathbf{D}_2(\theta_2)_{n-2,n-2}=1$ and $\mathbf{D}_2(\theta_2)_{k,k}=\prod_{l=k+1}^{n-2}\cos(\psi_{2,l})$ for $1\le k\le n-3$. And $\mathbf{G}_{21}$ is given by:

$$\mathbf{D}_2(\theta_2)\mathbf{Q}_2'^T(\theta_2)\mathbf{Q}_1'^T(\theta_1)(\Gamma_x-\lambda_2\mathbf{I}_n)\mathbf{Q}_1(\theta_1)\Delta(\theta_1,\theta_2)$$

with:

$$\frac{\mathrm{d}\mathbf{q}_2}{\mathrm{d}\theta_1}=\mathbf{Q}_1(\theta_1)\Delta(\theta_1,\theta_2)$$

For independent observations $\mathbf{x}$, the expression (11) becomes:

$$\mathbf{R}=\begin{bmatrix}\mathbf{R}_{11} & \mathbf{0}\\ \mathbf{0} & \mathbf{R}_{22}\end{bmatrix} \tag{15}$$

where $\mathbf{R}_{11}$ is given by (11) and where $\mathbf{R}_{22}$ is

$$\lambda_2\mathbf{D}_2(\theta_2)\mathbf{Q}_2'^T(\theta_2)\mathbf{Q}_1'^T(\theta_1)(\Gamma_x+2\lambda_2\mathbf{q}_2\mathbf{q}_2^T)\mathbf{Q}_1'(\theta_1)\mathbf{Q}_2'(\theta_2)\mathbf{D}_2(\theta_2) \tag{16}$$

with $\mathbf{Q}_2(\theta_2) \triangleq [\mathbf{Q}_2'(\theta_2),\mathbf{v}(\theta_2)]$ and with $\mathbf{D}_2(\theta_2)$ a $n-2\times n-2$ diagonal matrix where $\mathbf{D}_2(\theta_2)_{n-2,n-2}=1$ and $\mathbf{D}_1(\theta_1)_{k,k}=\prod_{l=k+1}^{n-2}\cos(\psi_{2,l})$ for $1\le k\le n-3$.

In the case of correlated observations $\mathbf{x}$, $\mathbf{R}$ is no longer block diagonal . We have:

$$\mathbf{R}=\begin{bmatrix}\mathbf{R}_{11} & \mathbf{R}_{12}\\ \mathbf{R}_{21} & \mathbf{R}_{22}\end{bmatrix} \tag{17}$$

with:

$$\mathbf{R}_{21}=\mathbf{R}_{12}^T=\mathbf{D}_2(\theta_2)\mathbf{Q}_2'^T(\theta_2)\mathbf{Q}_1'^T(\theta_1)\Gamma_{21}\mathbf{Q}_1'(\theta_1)\mathbf{D}_1(\theta_1) \tag{18}$$

where Γ_{21} takes respectively the values:

$$\sum_{k=1}^{q+n-1}\Gamma_k\mathbf{q}_1\mathbf{q}_2^T\Gamma_k+\Gamma_k^T\mathbf{q}_1\mathbf{q}_2^T\Gamma_k^T+(\mathbf{q}_2^T\Gamma_k\mathbf{q}_1)\Gamma_k+(\mathbf{q}_2^T\Gamma_k^T\mathbf{q}_1)\Gamma_k^T$$

for a MA(q) process, and the value:

$$\sum_{k=1}^{n-2}\Gamma_k\mathbf{q}_1\mathbf{q}_2^T\Gamma_k+\Gamma_k^T\mathbf{q}_1\mathbf{q}_2^T\Gamma_k^T+(\mathbf{q}_2^T\Gamma_k\mathbf{q}_1)\Gamma_k+(\mathbf{q}_2^T\Gamma_k^T\mathbf{q}_1)\Gamma_k^T+$$
$$\frac{\gamma_0^2a^{2n-2}}{1-a^2}[\Gamma_a\mathbf{q}_1\mathbf{q}_2^T\Gamma_a+\Gamma_a^T\mathbf{q}_1\mathbf{q}_2^T\Gamma_a^T+(\mathbf{q}_2^T\Gamma_a\mathbf{q}_1)\Gamma_a+(\mathbf{q}_2^T\Gamma_k^T\mathbf{q}_1)\Gamma_a^T]$$

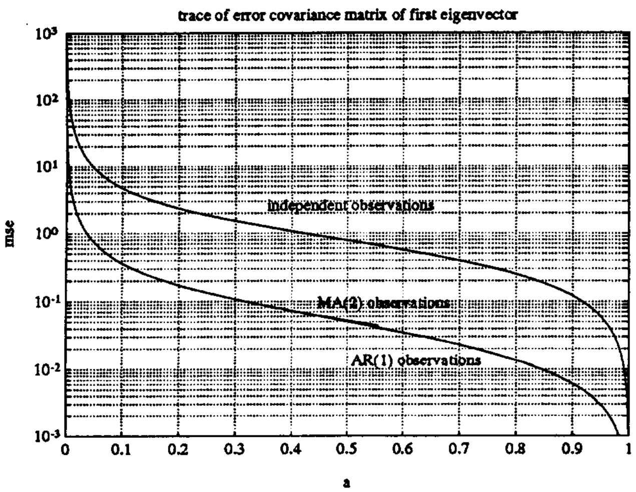
for an AR(1) process. For a MA(q) process (resp. AR(1)), $\mathbf{R}_{11}$ is given by the expression deduced from (12) (resp. (13)) and (11), while $\mathbf{R}_{22}$ is obtained by replacing in (16) the term $\Gamma_x+2\lambda_2\mathbf{q}_2\mathbf{q}_2^T$ by $\Gamma_x+2\lambda_2\mathbf{q}_2\mathbf{q}_2^T+\Gamma_{22}$, in which Γ_{22} is given by (12) (resp. (13)), provided $\mathbf{q}_1$ is replaced by $\mathbf{q}_2$.

Finally, we solve equation (7) in the situations of interest. We note that (7) is of triangular form for independent observations $\mathbf{x}$ in our deflation procedure. Then we can evaluate the precision of the estimated eigenvectors or of the estimated eigenspaces by derivation of a continuity theorem [10]. So if γ is "small enough" and k "large enough", the error covariance matrix of an eigenvector and the mean square error of the projection matrix on a eigenspace are approximately given by:

$$E[(\mathbf{q}_i(\boldsymbol{\theta}(k))-\mathbf{q}_i(\boldsymbol{\theta}_*))(\mathbf{q}_i(\boldsymbol{\theta}(k))-\mathbf{q}_i(\boldsymbol{\theta}_*))^T] \sim \gamma \frac{d\mathbf{q}_i(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \mathbf{P} \frac{d\mathbf{q}_i^T(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \quad (19)$$

$$E\|\sum_{i=1}^{m} \mathbf{q}_i(\boldsymbol{\theta}(k))\mathbf{q}_i^T(\boldsymbol{\theta}(k))-\mathbf{q}_i(\boldsymbol{\theta}_*)\mathbf{q}_i^T(\boldsymbol{\theta}_*)\|^2 \sim 2\gamma \mathrm{tr}[\sum_{i=1}^{m} \frac{d\mathbf{q}_i(\boldsymbol{\theta})}{d\boldsymbol{\theta}} \mathbf{P} \frac{d\mathbf{q}_i^T(\boldsymbol{\theta})}{d\boldsymbol{\theta}}]$$

We present on the figure below the case of a 3×3 covariance matrix issued from independent, AR(1) or MA(2) consecutive observations $\mathbf{x}$. The trace of (19) parameterized by the parameter a of the AR(1) process is displayed for $\mathbf{q}_1$. We observe that this error increases when a decreases i.e. when the two first eigenvalues get nearer. A simulation of the algorithm shows that the speed of convergence also decreases when a decreases. The values of the error are very close for AR(1) and MA(2) processes but is about 12dB worse for independent observations, thus are very sensitive to the relations between the different observations $\mathbf{x}(k)$.



5.SPECIAL CASE OF CENTRO-SYMMETRIC COVARIANCE MATRICES

To improve the accuracy of the subspaces estimation, we can exploit the centro-symmetric or block-centro-symmetric property of some covariance matrices. This property occurs in important applications: temporal covariance matrices issued from a temporal sampling of a stationary signal, and spatial covariance matrices issued from uncorrelated and band-limited sources observed on a centro-symmetric sensor array (for example on uniform linear arrays) [11] are centro-symmetric ; spatio-temporal covariance matrices used in subspace methods for blind identification of multichannel FIR filters [1] are block-centro-symmetric.

In the real case, we can use the property that an eigenvectors orthonormal basis of a symmetric centro-symmetric matrix can be obtained from eigenvectors orthonormal bases of two half size symmetric real matrices [12]. For example if n is even, $\boldsymbol{\Gamma}$ can be partitioned as follows:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{\Gamma}_1 & \boldsymbol{\Gamma}_2^T \\ \boldsymbol{\Gamma}_2 & \mathbf{J}\boldsymbol{\Gamma}_1\mathbf{J} \end{bmatrix},$$

where $\mathbf{J}$ is a matrix with ones on its anti-diagonal and zeroes elsewhere. And we may determine $n/2$ symmetric $(\mathbf{q}=\mathbf{Jq})$ [resp. $n/2$ skew symmetric $(\mathbf{q}=-\mathbf{Jq})$] orthonormal eigenvectors of $\boldsymbol{\Gamma}$ from the orthonormal eigenvectors of $\boldsymbol{\Gamma}_1+\mathbf{J}\boldsymbol{\Gamma}_2$ [resp. $\boldsymbol{\Gamma}_1-\mathbf{J}\boldsymbol{\Gamma}_2$]. Then, we use the Givens or Givens/Householder adaptive methods described previously and we can show that the variances of the estimated eigenvectors are reduced with respect to those obtained without using the centro-symmetric structure of $\boldsymbol{\Gamma}$.

In the complex case, the preceding property does not exist any longer : we only know that all eigenvectors $\mathbf{q}$ associated to a simple eigenvalue are structured, of the form : $\mathbf{q} = e^{i\omega}\mathbf{J}\mathbf{q}^*$, where the real number ω is a function of the eigenvector. But nevertheless this property can be used in this context. For example for n even we use the structure

$$\mathbf{q} = 1/\sqrt{2} \begin{bmatrix} \mathbf{u} \\ e^{i\omega}\mathbf{J}\mathbf{u}^* \end{bmatrix} \text{ with } \|\mathbf{u}\| = 1$$

to adaptively estimate the eigenvector associated to the smallest or largest eigenvalue or $\boldsymbol{\Gamma}$ according to an algorithm derived from the Givens algorithm, and we can show that the variance of the estimated eigenvector is also reduced when compared to those obtained without using the centro-symmetric structure of $\boldsymbol{\Gamma}$.

REFERENCES

[1] E.Moulines, P.Duhamel, JF.Cardoso and S.Mayrargue, *Subspace Methods for Blind Identification of Multichannel FIR Filters*, Proceeding of ICCASP 1994, pp.573-576.

[2] J.P.Yang and M.Kaveh, *Adaptive Eigensubspace Algorithms for Direction or Frequency Estimation and Tracking*, Trans.on ASSP, vol.36, N°2 Feb.1988.

[3] P.A.Regalia, *An Adaptive Unit Norm Filter with Applications to Signal Analysis and Karhunen-Loève Transformations*, Trans.on Circuits and Systems, vol.37, N°5 pp.646-649 May 1990.

[4] JP.Delmas, *A complex Adaptive Eigensubspace Algorithm for DOA or frequency Estimation and Tracking*, Proceeding of EUSIPCO 1992, pp.657-660.

[5] L.Ljung and T.Söderström, *Theory and Practice of Recursive Identification*, MIT Press 1983.

[6] H.Kushner, D.Clark, *Stochastic approximation methods for constrained and unconstrained systems*, Springer Verlag, 1978.

[7] H.Kushner, *Weak Convergence Methods and Singularly Perturbated Stochastic Control and Filtering Problems*, Systems and control: Foundations and Applications, vol.3, Birkhauser, 1990.

[8] P.Vaidyanathan, *Multirate Systems and Filter Banks*, Prentice Hall, 1993.

[9] A.Benveniste, M.Métivier, P.Priouret, *Adaptive Algorithms and Stochastic Approximation*, Springer Verlag, 1990.

[10] J.Billingsley, *Convergence of Probability Measures*, John Wiley, New York 1968.

[11] X.Guanghan, R.Roy and T.Kailath, *Detection of Number of Sources via Exploitation of Centro-Symmetry Property*, IEEE Trans.on Signal Processing, vol.42, N°1 Jan.1994.

[12] A.Cantoni and P.Butler, *Eigenvalues and Eigenvectors of Symmetric Centrosymmetric Matrices*, Linear Algebra and its Applications 13, pp.275-288, 1976.