On Adaptive EVD Asymptotic Distribution of Centro-Symmetric Covariance Matrices

Jean-Pierre Delmas

V. Conclusion

In this correspondence, a blind adaptive (BA) FRESH filter for the extraction of a desired signal from spectrally overlapped interference has been proposed. It has been shown that this BA-FRESH filter converges to the same optimum value as the LMS-FRESH filter. Furthermore, the rates of convergence of the two filters are of the same order $O(\frac{1}{N})$. Simulation results show that after a reasonable length of samples, the performance of the BA-FRESH filter is very close to that of the LMS-FRESH filter. The main advantage of the BA-FRESH filter is that it does not require knowledge of the statistics of the desired signal, nor does it require a copy of the signal. It only needs knowledge of the cyclostationary properties of the signals. This advantage makes the use of the BA-FRESH filter an attractive alternative.

References

[1] S. Haykin, *Adaptive Filter Theory*, 2nd ed. New York: Wiley, 1991, ch. 1, 9, 10, 15.

[2] W. A. Gardner, "Cyclic Wiener filtering: Theory and method," *IEEE Trans. Commun.*, vol. 41, pp. 151–163, Jan. 1993.

[3] A. Chevreuil and P. Loubaton, "On the use of conjugate cyclostationarity: A blind second-order multi-user equalization method," in *Proc. ICASSP*, 1996, pp. 2439–2442.

[4] G. B. Giannakis, "Filterbanks for blind channel identification and equalization," *IEEE Signal Processing Lett.*, vol. 4, pp. 184–187, June 1997.

[5] A. Chevreuil and P. Loubaton, "Blind second-order identification of FIR channels: Forced cyclostationarity and structured subspace method," *IEEE Signal Processing Lett.*, vol. 4, pp. 204–206, July 1997.

[6] E. Serpedin and G. B. Giannakis, "Blind channel identification and equalization with modulation induced cyclostationarity," in *Proc. Conf. Inform. Sci. Syst.*, 1997, pp. 792–797.

[7] J. Zhang and K. M. Wong, "A new kind of adaptive frequency shift filter," in *Proc. ICASSP*, 1995, vol. 2, pp. 913–916.

[8] W. A. Gardner and L. E. Franks, "Characterization of cyclostationary random signal processes," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 4–14, Jan. 1975.

[9] W. A. Gardner, *Introduction to Random Processes*, 2nd ed. New York: McGraw-Hill, 1990, ch. 12.

[10] W. A. Gardner, *Statistical Spectral Analysis: A Non-Probabilistic Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1987, ch. 14.

[11] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Baltimore, MD: Johns Hopkins Univ. Press, 1996, ch. 2.

[12] S. Haykin, *Communication Systems*, 2nd ed. New York: Wiley, 1983, ch. 9.

On Adaptive EVD Asymptotic Distribution of Centro-Symmetric Covariance Matrices

Jean-Pierre Delmas

***Abstract*—This correspondence investigates the gain in statistical performance/complexity of the adaptive estimation of the eigenvalue decomposition (EVD) of covariance matrices when the centrosymmetric (CS) structure of such matrices is utilized. After deriving the asymptotic distribution of the EVD estimators, it is shown, in particular, that the closed-form expressions for the asymptotic covariance of batch and adaptive EVD estimators are very similar, provided that the number of samples is replaced by the inverse of the step size.**

I. Introduction

Signal processing applications often lead to structured matrix problems. Algorithms that take this structure into account usually require fewer computations and have better numerical properties [1]. An important matrix structure is the centro-symmetric structure of covariance matrices of stationary signals, for which the symmetric Toeplitz structure is a particular case. This structure is instrumental in the realm of EVD problems. It is well known [2] that an orthonormal eigenbasis of a symmetric CS matrix can be obtained from orthonormal eigenbases of two half-sized symmetric real matrices [2]. This property has already been used in [3] and [4] for, respectively, a parameterized adaptive eigenspace algorithm and an adaptive eigenfilter bank. However, no asymptotic performance analysis has yet been done. The purpose of this correspondence is to specify the gain in statistical performance/complexity of the adaptive EVD when the CS structure of covariance matrices is used. For that, we choose, as an example, the stochastic gradient ascent algorithm (SGA), and we exhibit the asymptotic distribution of its EVD estimator.

This correspondence is organized as follows. In Section II, we recall the property that an orthonormal eigenbasis of a CS matrix can be obtained from orthonormal eigenbases of half-sized symmetric matrices. In Section III, we study the asymptotic distribution of an adaptive estimator of EVD of CS covariance matrices. In particular, a theorem gives a closed-form expression of the covariance of the limiting distribution of such an estimated EVD. Finally, in Section IV, we present some simulations.

II. Eigenvalue Decomposition Structure

We consider an $n \times n$ CS covariance matrix $\mathbf{R}_x = E(\mathbf{x}\mathbf{x}^T)$ of a Gaussian distributed, zero mean, real random vector $\mathbf{x}$, and we denote by $\lambda_1 > \cdots > \lambda_n$ the distinct eigenvalues of $\mathbf{R}_x$ and by $\mathbf{v}_1, \ldots, \mathbf{v}_n$ the corresponding normalized eigenvectors. The EVD estimators that we propose stem from the property that an orthonormal eigenbasis of $\mathbf{R}_x$ can be obtained from orthonormal eigenbases of half-sized symmetric matrices [2]. This property is recalled here for convenience of the reader and in order to fix notations. $\mathbf{R}_x$ can be reduced to a block diagonal form by a data independent orthogonal transformation

Manuscript recevied October 6, 1997; revised August 11, 1998. The associate editor coordinating the review of this paper and approving it for publication was Dr. Yingbo Hua.

The author is with the Institut National des Télécommunications, Evry Cedex, France (e-mail: delmas@int-evry.fr).

Publisher Item Identifier S 1053-587X(99)03240-7.

$\mathbf{K}$

$$\mathbf{R}_x = \mathbf{K}\begin{bmatrix}\mathbf{R}^- & \mathbf{O}\\ \mathbf{O} & \mathbf{R}^+\end{bmatrix}\mathbf{K}^T \tag{2.1}$$

with, respectively, for n even and n odd

$$\mathbf{K} = \frac{1}{\sqrt{2}}\begin{bmatrix}\mathbf{I} & \mathbf{I}\\ -\mathbf{J} & \mathbf{J}\end{bmatrix} \quad \text{and} \quad \mathbf{K} = \frac{1}{\sqrt{2}}\begin{bmatrix}\mathbf{I} & \mathbf{0} & \mathbf{I}\\ \mathbf{0}^T & \sqrt{2} & \mathbf{0}^T\\ -\mathbf{J} & \mathbf{0} & \mathbf{J}\end{bmatrix} \tag{2.2}$$

where $\mathbf{J}$ is a $\lfloor n/2\rfloor \times \lfloor n/2\rfloor$ matrix with ones on its antidiagonal and zeros elsewhere, and $\mathbf{I}$ is the $\lfloor n/2\rfloor \times \lfloor n/2\rfloor$ identity matrix. Therefore, $\lfloor n/2\rfloor$ skew-symmetric and $\lceil n/2\rceil$ symmetric orthonormal eigenvectors of $\mathbf{R}_x$ (denoted,[1] respectively, by $\mathbf{v}_1^-, \ldots, \mathbf{v}_{\lfloor n/2\rfloor}^-$ and $\mathbf{v}_1^+, \ldots, \mathbf{v}_{\lceil n/2\rceil}^+$), and corresponding eigenvalues (denoted, respectively, by $\lambda_1^-, \ldots, \lambda_{\lfloor n/2\rfloor}^-$ and $\lambda_1^+, \ldots, \lambda_{\lceil n/2\rceil}^+$) are determined from the solutions of the equations

$$\begin{aligned}\mathbf{R}^-\mathbf{u}_i^- &= \lambda_i^-\mathbf{u}_i^- \quad i = 1, \ldots, \lfloor n/2\rfloor \quad \text{and}\\ \mathbf{R}^+\mathbf{u}_i^+ &= \lambda_i^+\mathbf{u}_i^+ \quad i = 1, \ldots, \lceil n/2\rceil\end{aligned} \tag{2.3}$$

where $\mathbf{v}_i^s$ are connected to $\mathbf{u}_i^s$, respectively, for n even and odd by

$$\begin{aligned}&\mathbf{v}_i^s = \mathbf{K}_e^s\mathbf{u}_i^s, \quad \left[\text{resp. } \mathbf{v}_i^s = \mathbf{K}_o^s\mathbf{u}_i^s\right]\\ &s = -, \quad i = 1, \ldots, \lfloor n/2\rfloor, \quad s = +, \quad i = 1, \ldots, \lceil n/2\rceil\end{aligned} \tag{2.4}$$

with $\mathbf{K}_e^- \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}\mathbf{I}\\ -\mathbf{J}\end{smallmatrix}\right]$, $\mathbf{K}_e^+ \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}\mathbf{I}\\ \mathbf{J}\end{smallmatrix}\right]$ and $\mathbf{K}_o^- \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}\mathbf{I} & \mathbf{0}^T\\ -\mathbf{J} & \end{smallmatrix}\right]$, $\mathbf{K}_o^+ \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}}\left[\begin{smallmatrix}\sqrt{2} & \mathbf{0}^T\\ \mathbf{0} & \mathbf{J}\end{smallmatrix}\right]$. Moreover, the set $\{\mathbf{v}_1^-, \ldots, \mathbf{v}_{\lfloor n/2\rfloor}^-, \mathbf{v}_1^+, \ldots, \mathbf{v}_{\lceil n/2\rceil}^+\}$ forms an orthonormal set that therefore spans the eigenspace of $\mathbf{R}_x$.

III. Adaptive Estimator

A. Adaptive Algorithm

Every adaptive EVD algorithm can be adapted to our CS structured situation. Because $E\left[\begin{smallmatrix}\mathbf{y}_t^-\\ \mathbf{y}_t^+\end{smallmatrix}\right]\left[\mathbf{y}_t^{-T}\ \mathbf{y}_t^{+T}\right] = \left[\begin{smallmatrix}\mathbf{R}^- & \mathbf{O}\\ \mathbf{O} & \mathbf{R}^+\end{smallmatrix}\right]$ with $\left[\begin{smallmatrix}\mathbf{y}_t^-\\ \mathbf{y}_t^+\end{smallmatrix}\right] \stackrel{\text{def}}{=} \mathbf{K}^T\mathbf{x}_t$, each adaptive EVD algorithm can be split into two decoupled algorithms. To give prominence to this improvement brought by this splitting, we take, as an example, an adaptive algorithm introduced in the neural network literature by Oja [the so-called stochastic gradient ascent algorithm (SGA)] because of the simplicity of its asymptotic distribution [9]. Its convergence is studied in [7] and, as was shown in [9], it achieves a good convergence speed/misadjustment tradeoff among a family of numerically simple algorithms.

$$\begin{aligned}\mathbf{u}_{k,t+1}^s = \mathbf{u}_{k,t}^s + \alpha_k^s\gamma\Bigg[&\mathbf{I}_{n^s} - \mathbf{u}_{k,t}^s\mathbf{u}_{k,t}^{sT}\\ &- \sum_{i=1}^{k-1}\left(1 + \frac{\alpha_i^s}{\alpha_k^s}\right)\mathbf{u}_{i,t}^s\mathbf{u}_{i,t}^{sT}\Bigg]\mathbf{y}_t^s\mathbf{y}_t^{sT}\mathbf{u}_{k,t}^s\end{aligned} \tag{3.5}$$

$$\lambda_{k,t+1}^s = \lambda_{k,t}^s + \gamma\left[\mathbf{u}_{k,t}^{sT}\mathbf{y}_t^s\mathbf{y}_t^{sT}\mathbf{u}_{k,t}^s - \lambda_{k,t}^s\right] \tag{3.6}$$

for $s = -, +$ and $k = 1, \ldots, n^s$ ($n^- \stackrel{\text{def}}{=} \lfloor n/2\rfloor$ and $n^+ \stackrel{\text{def}}{=} \lceil n/2\rceil$). $\mathbf{u}_{k,t}^-$ [resp. $\mathbf{u}_{k,t}^+$] is associated with the $\lfloor n/2\rfloor$ skew-symmetric eigenvectors $\mathbf{v}_i$, [resp. the $\lceil n/2\rceil$ symmetric eigenvectors $\mathbf{v}_i$]. The parameters α_k^s ($\alpha_1^s = 1$ and $\alpha_k^s > 0$, $k = 1, \ldots, n^s$) afford a better tradeoff between the convergence speed and misadjustment [9], and γ is the step size. As the computational cost of the SGA algorithm is $O(n^2)$ flops by iteration, the number of operations of our split procedure is roughly halved. As this split SGA algorithm (3.5) and (3.6) can be globally written in the form (we write Θ_t^γ for the seqeunce of estimates to emphasize the dependence on γ)

$$\Theta_{t+1}^\gamma = \Theta_t^\gamma + \gamma g\left(\Theta_t^\gamma, \mathbf{x}_t\right) \tag{3.7}$$

with $\Theta_t = \text{Vec}(\Theta_t^-, \Theta_t^+)$ where $\Theta_t^s \stackrel{\text{def}}{=} \text{Vec}(\mathbf{u}_{1,t}^s, \ldots, \mathbf{u}_{n^s,t}^s, \lambda_{1,t}^s, \ldots, \lambda_{n^s,t}^s)$, $s = -, +$, we can use a general approximation result [8, th. 2, p. 108], which is shortly recalled in [9, Sec. III.A] to evaluate the asymptotic distribution of this EVD estimator. Of course, the study set out in this section could be immediately extended to other gradient-type (for example, algorithms studied in [9]) or RLS-type algorithms.

B. Asymptotic Distribution

1) Local Caracterization of the Field: According to [8, Th. 2, p. 108], we need to characterize two local properties of the field $g(\Theta_t, \mathbf{x}_t)$: the mean value of its derivative $\mathbf{D}$ and its covariance $\mathbf{G}$, both evaluated at the point $\Theta_t = \Theta_* = \text{Vec}(\Theta_*^-, \Theta_*^+)$ with $\Theta_*^s \stackrel{\text{def}}{=} \text{Vec}(\mathbf{u}_1^s, \ldots, \mathbf{u}_{n^s}^s, \lambda_1^s, \ldots, \lambda_{n^s}^s)$, $s = -, +$.

Derivative of the field: It is straightforward to see that $\mathbf{D}$ can be partitioned as

$$\mathbf{D} = \begin{bmatrix}\mathbf{D}_{u^-} & \mathbf{O} & \mathbf{O} & \mathbf{O}\\ \mathbf{D}_{u^-,\lambda^-} & -\mathbf{I}_{n^-} & \mathbf{O} & \mathbf{O}\\ \mathbf{O} & \mathbf{O} & \mathbf{D}_{u^+} & \mathbf{O}\\ \mathbf{O} & \mathbf{O} & \mathbf{D}_{u^+,\lambda^+} & -\mathbf{I}_{n^+}\end{bmatrix} \tag{3.8}$$

with

$$\mathbf{D}_{u^s,\lambda^s} = 2\,\text{Diag}\left(\lambda_1^s\mathbf{u}_1^{sT}, \ldots, \lambda_{n^s}^s\mathbf{u}_{n^s}^{sT}\right), \quad s = -, +. \tag{3.9}$$

$\mathbf{D}_{u^s}$ are $n^s \times n^s$ block matrices, the block $(\mathbf{D}_{u^s})_{i,j}$ of which is given in [9] by (3.10), shown at the bottom of the page.

Covariance of the field: The field of the algorithm (3.5) and (3.6) can be globally written in the form

$$\begin{aligned}g(\Theta_t, \mathbf{x}_t) &= \begin{bmatrix}g^-\left(\Theta_t^-, \mathbf{y}_t^-\mathbf{y}_t^{-T}\right)\\ g^+\left(\Theta_t^+, \mathbf{y}_t^+\mathbf{y}_t^{+T}\right)\end{bmatrix} = \begin{bmatrix}g_{u^-}\left(\Theta_t^-, \mathbf{y}_t^-\mathbf{y}_t^{-T}\right)\\ g_{\lambda^-}\left(\Theta_t^-, \mathbf{y}_t^-\mathbf{y}_t^{-T}\right)\\ g_{u^+}\left(\Theta_t^+, \mathbf{y}_t^+\mathbf{y}_t^{+T}\right)\\ g_{\lambda^+}\left(\Theta_t^+, \mathbf{y}_t^+\mathbf{y}_t^{+T}\right)\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{A}^- & \mathbf{O}\\ \mathbf{B}^- & \mathbf{O}\\ \mathbf{O} & \mathbf{A}^+\\ \mathbf{O} & \mathbf{B}^+\end{bmatrix}\begin{bmatrix}\text{Vec}\left(\mathbf{y}_t^-\mathbf{y}_t^{-T}\right)\\ \text{Vec}\left(\mathbf{y}_t^+\mathbf{y}_t^{+T}\right)\end{bmatrix} - \begin{bmatrix}\mathbf{O}\\ \Lambda^-\\ \mathbf{O}\\ \Lambda^+\end{bmatrix}\end{aligned} \tag{3.11}$$

with

$$\begin{aligned}\mathbf{A}^s &\stackrel{\text{def}}{=} \begin{bmatrix}\mathbf{u}_1^{sT} \otimes \mathbf{A}_1^s\\ \vdots\\ \mathbf{u}_{n^s}^{sT} \otimes \mathbf{A}_{n^s}^s\end{bmatrix}, \quad \mathbf{B}^s \stackrel{\text{def}}{=} \begin{bmatrix}\mathbf{u}_1^{sT} \otimes \mathbf{u}_1^{sT}\\ \vdots\\ \mathbf{u}_{n^s}^{sT} \otimes \mathbf{u}_{n^s}^{sT}\end{bmatrix} \quad \text{and}\\ \Lambda^s &\stackrel{\text{def}}{=} \begin{bmatrix}\lambda_1^s\\ \vdots\\ \lambda_{n^s}^s\end{bmatrix} \quad s = -, +\end{aligned} \tag{3.12}$$

[1] We introduce this notation because, in general, we have no *a priori* information on the order of the eigenvalues associated to skew-symmetric and symmetric eigenvectors.

$$(\mathbf{D}_{u^s})_{i,j} = \begin{cases}-\alpha_i^s\left[\sum_{k=1}^{i-1}\left(\lambda_i^s + \frac{\alpha_k^s}{\alpha_i^s}\lambda_k^s\right)\mathbf{u}_k^s\mathbf{u}_k^{sT} + 2\lambda_i^s\mathbf{u}_i^s\mathbf{u}_i^{sT} + \sum_{k=i+1}^{n^s}\left(\lambda_i^s - \lambda_k^s\right)\mathbf{u}_k^s\mathbf{u}_k^{sT}\right], & i = j\\ \mathbf{O}, & i < j\\ -\alpha_i^s\left(1 + \frac{\alpha_j^s}{\alpha_i^s}\right)\lambda_i^s\mathbf{u}_j^s\mathbf{u}_i^{sT}, & i > j.\end{cases} \tag{3.10}$$

and

$$\mathbf{A}_k^s = \alpha_k^s \left[\mathbf{I}_{n^s} - \mathbf{u}_{k,t}^s \mathbf{u}_{k,t}^{s^T} - \sum_{i=1}^{k-1} \left(1 + \frac{\alpha_i^s}{\alpha_k^s} \right) \mathbf{u}_{i,t}^s \mathbf{u}_{i,t}^{s^T} \right] \quad k = 1, \ldots, n^s, \quad s = -, + \tag{3.13}$$

thanks to the classic relation $\mathrm{Vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\mathrm{Vec}(\mathbf{B})$. The covariance $\mathbf{G}$ of the field evaluated at $\Theta = \Theta_* = \mathrm{Vec}(\mathbf{v}^-, \Lambda^-, \mathbf{v}^+, \Lambda^+)$, in the case where the observations $\mathbf{x}_t$ are independent, can be partitioned as

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_{u-} & \mathbf{G}_{u-,\lambda-}^T & \mathbf{O} & \mathbf{O} \\ \mathbf{G}_{u-,\lambda-} & \mathbf{G}_{\lambda-} & \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{G}_{u+} & \mathbf{G}_{u+,\lambda+}^T \\ \mathbf{O} & \mathbf{O} & \mathbf{G}_{u+,\lambda+} & \mathbf{G}_{\lambda+} \end{bmatrix}. \tag{3.14}$$

We prove in the Appendix that

$$\mathbf{G}_{u^s,\lambda^s} = \mathbf{O}, \quad \text{and} \quad \mathbf{G}_{\lambda^s} = 2\,\mathrm{Diag}\left(\left(\lambda_1^s\right)^2, \ldots, \left(\lambda_{n^s}^s\right)^2 \right) \tag{3.15}$$

for $s = -, +$. Here, $\mathbf{G}_{u^s}$ are the $n^s \times n^s$ matrices, the block $(\mathbf{G}_{u^s})_{i,j}$ of which is given in [9] by (3.16), shown at the bottom of the page.

2) Solution of the Lyapunov Equation For independent observations $\mathbf{x}_t$ and for the investigated algorithm for which the eigenvalues of the derivative (3.8) of the mean field have strictly negative real parts (see [9] for the eigenvalues of $\mathbf{D}_{u^s}$), the hypotheses of the model of Benveniste *et al.* ([8, th. 2, p. 108]) are fulfilled. However, the underlying assumption for the results by Benveniste *et al.* is that the solution of the corresponding stochastic approximation type algorithm with decreasing step size almost surely converges to the unique asymptotically stable point of the associated ODE. Since the normalized eigenvectors are defined up to a sign, the global attractor Θ_* is not unique. However, the practical use of the Benveniste results in such a situation is usually justified (for example, in [10]) by using formally a general approximation result ([8, th. 1, p. 107]). Furthermore, the almost-sure (a.s.) convergence of the associated decreasing step-size algorithms are not strictly fulfilled for the SGA algorithm. This a.s. convergence would need a boundedness condition whose satisfaction is a challenging problem. However, as discussed in [11], this condition was proved for only the Oja learning rule [12] designed for extracting the most dominant eigenvector by means of a single linear unit neuron network, where Oja *et al.* [13] showed that if this algorithm is used with uniformly bounded inputs $\mathbf{x}_t$, then $\mathbf{v}_{1,t}$ remains inside some bounded subset. If we allow ourselves the Benveniste results in our situation, the Lyapunov continuous equations ([9, Eq. (3.16)]) can be solved exactly. Since the matrices $\mathbf{D}$ and $\mathbf{G}$ are 2×2 block diagonal, this Lyapunov equation can be reduced to two decoupled equations. Thus, the covariance matrix $\mathbf{C}_\Theta$ of the asymptotic distribution of $\frac{1}{\sqrt{\gamma}}(\Theta_t^\gamma - \Theta_*)$ when $t \to \infty$ and $\gamma \to 0$ is

$$\mathbf{C}_\Theta = \mathrm{Diag}(\mathbf{C}_{\Theta-}, \mathbf{C}_{\Theta+}) \tag{3.17}$$

where $\mathbf{C}_{\Theta^s} = \begin{bmatrix} \mathbf{C}_{u^s} & \mathbf{C}_{u^s,\lambda^s}^T \\ \mathbf{C}_{u^s,\lambda^s} & \mathbf{C}_{\lambda^s} \end{bmatrix}$ are solutions of the Lyapunov equation

$$\begin{bmatrix} \mathbf{D}_{u^s} & \mathbf{O} \\ \mathbf{D}_{u^s,\lambda^s} & -\mathbf{I}_{n^s} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{u^s} & \mathbf{C}_{u^s,\lambda^s}^T \\ \mathbf{C}_{u^s,\lambda^s} & \mathbf{C}_{\lambda^s} \end{bmatrix} + \begin{bmatrix} \mathbf{C}_{u^s} & \mathbf{C}_{u^s,\lambda^s}^T \\ \mathbf{C}_{u^s,\lambda^s} & \mathbf{C}_{\lambda^s} \end{bmatrix} \begin{bmatrix} \mathbf{D}_{u^s}^T & \mathbf{D}_{u^s,\lambda^s}^T \\ \mathbf{O} & -\mathbf{I}_{n^s} \end{bmatrix} = -\begin{bmatrix} \mathbf{G}_{u^s} & \mathbf{O} \\ \mathbf{O} & \mathbf{G}_{\lambda^s} \end{bmatrix}. \tag{3.18}$$

Therefore, $\mathbf{C}_{u^s}$ are solutions of the Lyapunov equation $\mathbf{D}_{u^s}\mathbf{C}_{u^s} + \mathbf{C}_{u^s}\mathbf{D}_{u^s}^T + \mathbf{G}_{u^s} = \mathbf{O}$, the block $(\mathbf{C}_{u^s})_{i,j}$ of which is given in [9] by

$$(\mathbf{C}_{u^s})_{i,j} = \begin{cases} \sum_{1 \le k \ne i \le n^s} \frac{\alpha_{\min(i,k)}^s \lambda_i^s \lambda_k^s}{2\left|\lambda_i^s - \lambda_k^s\right|} \mathbf{u}_k^s \mathbf{u}_k^{s^T}, & i = j \\ -\frac{\alpha_{\min(i,j)}^s \lambda_i^s \lambda_j^s}{2\left|\lambda_i^s - \lambda_j^s\right|} \mathbf{u}_j^s \mathbf{u}_i^{s^T}, & i \ne j \end{cases} \tag{3.19}$$

for $s = -, +$. It is proved in the Appendix that

$$\mathbf{C}_{u^s,\lambda^s} = \mathbf{O}, \quad \text{and} \quad \mathbf{C}_{\lambda^s} = \mathrm{Diag}\left(\left(\lambda_1^s\right)^2, \ldots, \left(\lambda_{n^s}^s\right)^2 \right) \tag{3.20}$$

for $s = -, +$. Last, if we apply the linear mapping deduced from (2.4), in which $\mathbf{v}^s \stackrel{\text{def}}{=} \mathrm{Vec}(\mathbf{v}_1^s, \ldots, \mathbf{v}_{n^s}^s)$ is equal for $s = -, +$

$$\mathbf{v}^s = \begin{cases} \mathrm{Diag}\left(\mathbf{K}_e^s, \ldots, \mathbf{K}_e^s\right)\mathbf{u}^s, & \text{for } n \text{ even} \\ \mathrm{Diag}\left(\mathbf{K}_o^s, \ldots, \mathbf{K}_o^s\right)\mathbf{u}^s, & \text{for } n \text{ odd} \end{cases} \tag{3.21}$$

the following theorem is proved.

Theorem 1: $\frac{1}{\sqrt{\gamma}}\left(\mathrm{Vec}(\mathbf{v}_t^-, \Lambda_t^-, \mathbf{v}_t^+, \Lambda_t^+) - \mathrm{Vec}(\mathbf{v}^-, \Lambda^-, \mathbf{v}^+, \Lambda^+)\right)$ converges in distribution ($t \to \infty$ and $\gamma \to 0$) to the zero mean Gaussian distribution of covariance $\mathbf{C}_{v,\lambda}$ with $\mathbf{C}_{v,\lambda} = \mathrm{Diag}(\mathbf{C}_{v-}, \mathbf{C}_{\lambda-}, \mathbf{C}_{v+}, \mathbf{C}_{\lambda+})$

$$\mathbf{C}_{\lambda^s} = \mathrm{Diag}\left(\left(\lambda_1^s\right)^2, \ldots, \lambda_{n^s}^s \right)^2), \quad s = -, + \tag{3.22}$$

$$\mathbf{C}_{v^s} = \sum_{1 \le i \ne j \le n^s} \frac{\alpha_{\min(i,j)}^s \lambda_i^s \lambda_j^s}{2\left|\lambda_i^s - \lambda_j^s\right|} \mathbf{e}_i \mathbf{e}_i^T \otimes \mathbf{v}_j^s \mathbf{v}_j^{s^T} - \sum_{1 \le i \ne j \le n^s} \frac{\alpha_{\min(i,j)}^s \lambda_i^s \lambda_j^s}{2\left|\lambda_i^s - \lambda_j^s\right|} \mathbf{e}_i \mathbf{e}_j^T \otimes \mathbf{v}_j^s \mathbf{v}_i^{s^T}, \quad s = -, + \tag{3.23}$$

where $\mathbf{e}_i^s$ is the ith unit vector in $\mathcal{R}^{n^s}$. Therefore, if γ is "small enough" and t "large enough," the mean square error of Θ_t is approximately equal to $\gamma \mathrm{Tr}\mathbf{C}_\Theta$; therefore, in particular

$$E\|\Lambda_t - \Lambda\|_{\mathrm{Fro}}^2 \sim \gamma \sum_{k=1}^n \lambda_k^2 \tag{3.24}$$

$$E\|\mathbf{v}_t - \mathbf{v}\|_{\mathrm{Fro}}^2 \sim \gamma \left(\sum_{1 \le j \ne k \le \lfloor n/2 \rfloor} \frac{\alpha_{\min(j,k)}^- \lambda_j^- \lambda_k^-}{2\left|\lambda_j^- - \lambda_k^-\right|} + \sum_{1 \le j \ne k \le \lceil n/2 \rceil} \frac{\alpha_{\min(j,k)}^+ \lambda_j^+ \lambda_k^+}{2\left|\lambda_j^+ - \lambda_k^+\right|} \right). \tag{3.25}$$

We note that the asymptotic MSE of the estimated eigenvectors are reduced when the CS structure is taken into account [9, rel. (3.62)]. The number of terms in the summations (3.25) are roughly halved, and the difference between two successive eigenvalues λ_k^s is generally larger than between successive eigenvalues λ_k. In particular, if successive eigenvalues λ_k interlace (i.e., $\lambda_{2k} = \lambda_k^-$ and $\lambda_{2k+1} = \lambda_k^+$), the asymptotic MSE can be considerably reduced. Necessary conditions for this interlaced distribution are given in [2] for the general CS structure and in [6] for the Toeplitz structure. As far as the asymptotic distribution is concerned, similar results could be derived from other gradient-like algorithms such as the generalized Hebbian algorithm, the weighted subspace algorithm, and the optimal fitting analyzer. It would be sufficient to use the asymptotic distributions of their unstructured eigenvectors estimators given in [9].

$$(\mathbf{G}_{u^s})_{i,j} = \begin{cases} \sum_{k=1}^{i-1} \left(\alpha_k^s\right)^2 \lambda_i^s \lambda_k^s \mathbf{u}_k^s \mathbf{u}_k^{s^T} + \sum_{k=i+1}^{n^s} \left(\alpha_i^s\right)^2 \lambda_i^s \lambda_k^s \mathbf{u}_k^s \mathbf{u}_k^{s^T}, & i = j \\ -\left(\alpha_{\min(i,j)}^s\right)^2 \lambda_i^s \lambda_j^s \mathbf{u}_j^s \mathbf{u}_i^{s^T}, & i \ne j. \end{cases} \tag{3.16}$$

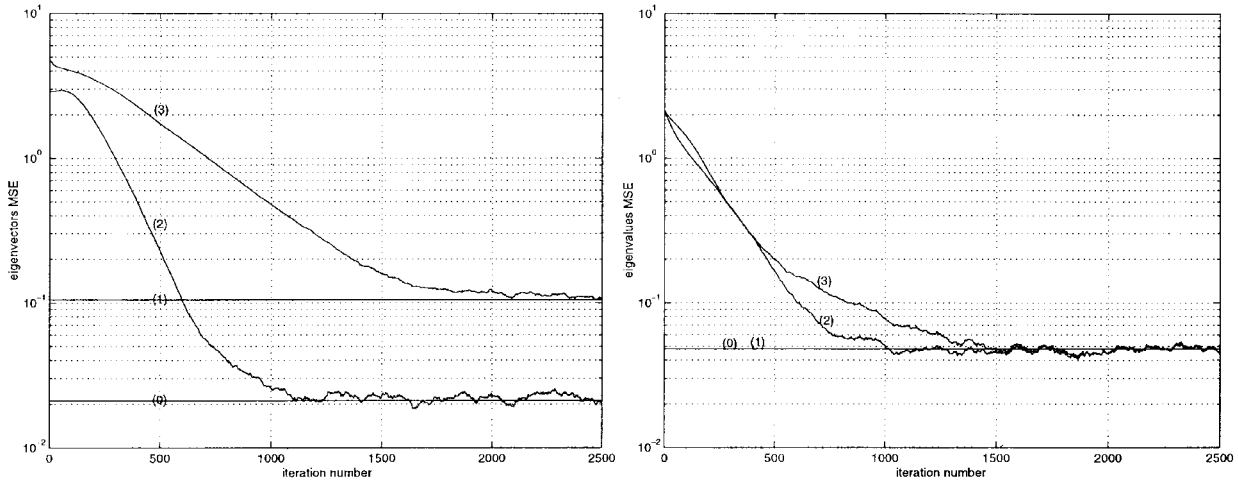

Fig. 1. Learning curves of the MSE $E\|\mathbf{v}_t - \mathbf{v}_*\|^2_{\mathrm{Fro}}$ and $E\|\Lambda_t - \Lambda_*\|^2$ averaging 100 independent runs for, respectively, the SGA algorithm when the CS structure is taken into account (2) or not (3), compared with the theoretical asymptotic values $\gamma\,\mathrm{Tr}(\mathbf{C}_v)$ and $\gamma\,\mathrm{Tr}(\mathbf{C}_\lambda)$ when the CS structure is taken into account (0) or not (1).

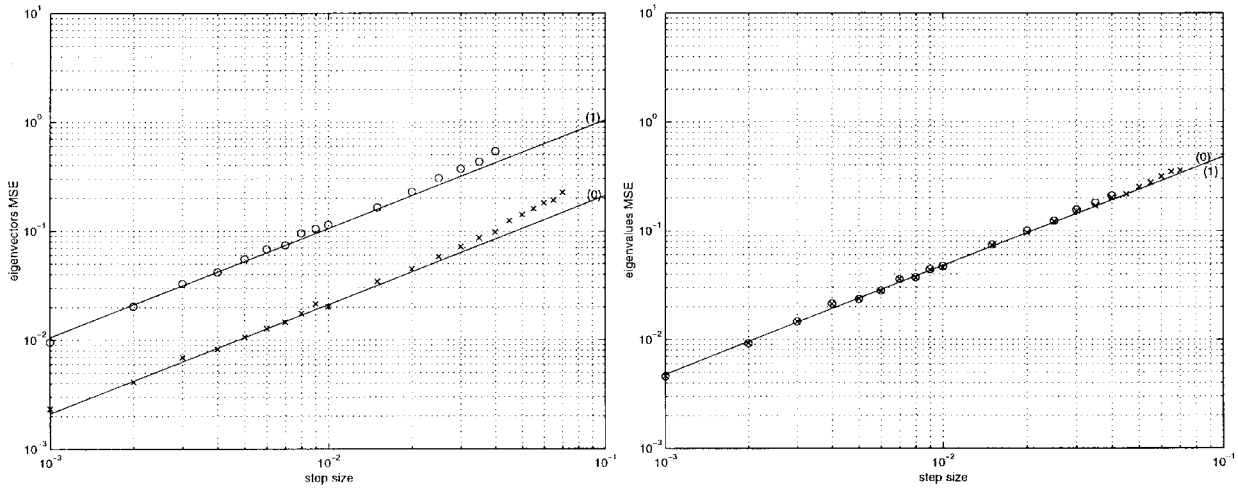

Fig. 2. Estimated [respectively, theoretical asymptotic] eigenvectors and eigenvalues MSE as a function of the step size γ when the CS structure is taken into account $(\times)$ [respectively, (0)] or not (o) [respectively, (1)].

C. Comparison Between Batch and Adaptive EVD Estimators

It is easy to show that the ML batch EVD estimator and the SGA adaptive estimator have very similar asymptotic distributions. With $\alpha_i = 1, i = 1, \ldots, n$, these distributions are equivalent[2] if we substitute $\frac{2}{t}$ (t is the sample number) by γ and the differences $(\lambda_j - \lambda_k)^2$ by $|\lambda_j - \lambda_k|$.

It is worth noticing that the ML batch estimators and the SGA adaptive estimators derive from the same cost functions, which is undoubtedly the reason for such similar asymptotic properties. On the one hand, $\mathbf{U}_k^s \stackrel{\mathrm{def}}{=} (\mathbf{u}_1^s, \ldots, \mathbf{u}_{n^s}^s)$ derives from the successive constrained minimizations of $\mathrm{Tr}(\mathbf{U}_k^{s\,T}\mathbf{R}_t^s\mathbf{U}_k^s)$, $k = 1, \ldots, n^s$ with respect to $\mathbf{u}_k^s$ under the constraint that $\mathbf{U}_k^{s\,T}\mathbf{U}_k^s = \mathbf{I}_k$ in ML batch estimation. On the other hand, $\mathbf{U}_{n^s}^s$ derives from the minimization of $\mathrm{Tr}(\mathbf{U}_{n^s}^{s\,T}\mathbf{R}_t^s\mathbf{U}_{n^s}^s)$ from a projected gradient-like procedure in SGA instantaneous adaptive estimation. The projection on the constraint $\mathbf{U}_{n^s}^{s\,T}\mathbf{U}_{n^s}^s = \mathbf{I}_{n^s}$ is realized thanks to an expansion of a Gram–Schmidt orthogonalization [9].

[2]The proof is omitted for want of space.

IV. Simulations

We consider throughout this section independent observations $\mathbf{x}_t$ in $\mathcal{R}^4$ associated with the symmetric Toeplitz matrix $\mathbf{R}_x = \mathrm{Toeplitz}(1, -0.3633, 0.0209, -0.0043)$ obtained from an ARMA process generated by the linear filter $F(z) = 57.7293(1 - 0.03z^{-1})/(1 - 0.03z^{-1} - 0.01z^{-2})$ driven by an unit variance noise. $\mathbf{R}_x$ has the following eigenvalues: $\lambda_1 = \lambda_1^- = 1.6079, \lambda_2 = \lambda_1^+ = 1.2028, \lambda_3 = \lambda_2^- = 0.7597, \lambda_4 = \lambda_2^+ = 0.4296$. This experiment presents the case of SGA adaptive estimation $(\alpha_i^s = 1, s = -, +, i = 1, \ldots, n^s)$. Fig. 1 shows the learning curves (averaged over 100 independent runs) of the eigenvalue MSE and eigenvector MSE when the CS structure is taken into account or not, with the common step size $\gamma = 0.01$. These MSE's tend to values in excellent agreement with the theoretical values predicted by (3.24) and (3.25). We observe a reduction of the eigenvector MSE of 7 db when the CS structure is taken into account. Furthermore, in this latter case, the convergence speed is improved as well. Fig. 2 shows the theoretical asymptotic and the estimated eigenvalue and eigenvector MSE's as a function of γ. Our present asymptotic analysis is seen to be valid over a large range of γ $(\gamma < 0.03)$, and

$$(\mathbf{G}_{u^s,\lambda^s})_{i,j} = \left(\mathbf{u}_i^{s^T} \otimes \mathbf{u}_i^{s^T}\right)(\mathbf{R}^s \otimes \mathbf{R}^s + (\mathbf{R}^s \otimes \mathbf{R}^s)\mathbf{P})\left(\mathbf{u}_j^s \otimes \mathbf{A}_j^{s^T}\right) = \left(\mathbf{u}_i^{s^T} \otimes \mathbf{u}_i^{s^T}\right)(\mathbf{R}^s \otimes \mathbf{R}^s)\left(\mathbf{u}_j^s \otimes \mathbf{A}_j^{s^T} + \mathbf{A}_j^{s^T} \otimes \mathbf{u}_j^s\right)$$
$$= \left(\mathbf{u}_i^{s^T}\mathbf{R}^s\mathbf{u}_j^s\right)\left(\mathbf{u}_i^{s^T}\mathbf{R}^s\mathbf{A}_j^{s^T}\right) + \left(\mathbf{u}_i^{s^T}\mathbf{R}^s\mathbf{A}_j^{s^T}\right)\left(\mathbf{u}_i^{s^T}\mathbf{R}^s\mathbf{u}_j^s\right) = \lambda_i^s\delta_{i,j}\left(\lambda_i^s\mathbf{u}_i^{s^T}\mathbf{A}_j^{s^T}\right) + \left(\lambda_i^s\mathbf{u}_i^{s^T}\mathbf{A}_j^{s^T}\right)\lambda_j^s\delta_{i,j}$$
$$= \begin{cases}\mathbf{0}^T, & \text{for } i \neq j \\ 2\left(\lambda_i^s\right)^2\mathbf{u}_i^{s^T}\mathbf{A}_i^{s^T} = 2\left(\lambda_i^s\right)^2\alpha_i^s\mathbf{u}_i^{s^T}\left(\mathbf{I}_{n^s} - \mathbf{u}_i^s\mathbf{u}_i^{s^T}\right) = \mathbf{0}^T, & \text{for } i = j.\end{cases} \tag{A.3}$$

the domain of "stability" is $\gamma < 0.07$, for which we observe good agreement between the theoretical and estimated MSE's.

V. Conclusion

In this correspondence, we have shown that when the CS structure of covariance matrices is taken into account, the EVD estimation can be split into two independent EVD estimations. As a result, we have proved, taking the SGA algorithm as an example, that the asymptotic MSE is reduced, and the complexity of the EVD is roughly halved. Finally, numerical simulations confirm the accuracy of our asymptotic analysis and show that for the SGA adaptive estimation, the convergence speed is improved, yielding a better tradeoff between convergence speed and misadjustment.

Appendix A
Proof of the Relations (3.15)

From (3.11) and (3.14), we have $\mathbf{G}_{u^s,\lambda^s} = \mathbf{B}^s\mathrm{Cov}(\mathrm{Vec}(\mathbf{y}_t^s\mathbf{y}_t^{s^T}))\mathbf{A}^{s^T}$ and $\mathbf{G}_{\lambda^s} = \mathbf{B}^s\mathrm{Cov}(\mathrm{Vec}(\mathbf{y}_t^s\mathbf{y}_t^{s^T}))\mathbf{B}^{s^T}$. For a Gaussian vector $\mathbf{y}_t^s$, we have ([5, p. 57])

$$\mathrm{Cov}\left(\mathrm{Vec}\left(\mathbf{y}_t^s\mathbf{y}_t^{s^T}\right)\right) = \mathbf{R}^s \otimes \mathbf{R}^s + (\mathbf{R}^s \otimes \mathbf{R}^s)\mathbf{P} \tag{A.1}$$

where $\mathbf{P}$ is an $(n^s)^2 \times (n^s)^2$ block matrix acting as a permutation operator in the sense that for any vector $\mathbf{a}$ or matrix $\mathbf{A}$ and vector $\mathbf{b}$, we have

$$\mathbf{P}(\mathbf{a} \otimes \mathbf{b}) = \mathbf{b} \otimes \mathbf{a} \quad \text{and} \quad \mathbf{P}(\mathbf{A} \otimes \mathbf{b}) = \mathbf{b} \otimes \mathbf{A}. \tag{A.2}$$

On the one hand, it follows that the block $(\mathbf{G}_{u^s,\lambda^s})_{i,j}$ is given by (A.3), shown at the top of the page. The first and second equalities use, respectively, (A.1) and (A.2), whereas the third equality stems from the classic property $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC} \otimes \mathbf{BD})$, the fourth equality uses (2.3), and final equality uses (3.13). On the other hand, the $(\mathbf{G}_{\lambda^s})_{i,j}$ entries are given by

$$(\mathbf{G}_{\lambda^s})_{i,j} = \left(\mathbf{u}_i^{s^T} \otimes \mathbf{u}_i^{s^T}\right)(\mathbf{R}^s \otimes \mathbf{R}^s + (\mathbf{R}^s \otimes \mathbf{R}^s)\mathbf{P})\left(\mathbf{u}_j^s \otimes \mathbf{u}_j^s\right)$$
$$= 2\left(\mathbf{u}_i^{s^T} \otimes \mathbf{u}_i^{s^T}\right)(\mathbf{R}^s \otimes \mathbf{R}^s)\left(\mathbf{u}_j^s \otimes \mathbf{u}_j^{s^T}\right)$$
$$= 2\left(\mathbf{u}_i^{s^T}\mathbf{R}^s\mathbf{u}_j^s\right)\left(\mathbf{u}_i^{s^T}\mathbf{R}^s\mathbf{u}_j^s\right) = 2\left(\lambda_i^s\right)^2\delta_{i,j}. \tag{A.4}$$

The first and second equalities use, respectively, (A.1) and (A.2), and the third equality uses (2.3).

Appendix B
Proof of the Relations (3.20)

From (3.18), we get

$$\left(\mathbf{D}_{u^s} - \mathbf{I}_{(n^s)^2}\right)\mathbf{C}_{u^s,\lambda^s}^T + \mathbf{C}_{u^s}\mathbf{D}_{u^s,\lambda^s}^T = \mathbf{O}. \tag{B.1}$$

Consider the change of basis stated in [9], which we recall for convenience. Let $\mathbf{U}^s$ be the $(n^s)^2 \times (n^s)^2$ orthonormal matrix $(\mathbf{U}_1^s, \mathbf{U}_2^s)$, where $\mathbf{U}_1^s = \mathrm{Diag}(\mathbf{u}_1^s, \ldots, \mathbf{u}_{n^s}^s)$, and $\mathbf{U}_2^s$ is the $(n^s)^2 \times n^s(n^s-1)$ block matrix made of the $\frac{n^s(n^s-1)}{2}$ matrices $(n^s)^2 \times 2$ $(\mathbf{e}_i \otimes \mathbf{u}_j^s, \mathbf{e}_j \otimes \mathbf{u}_i^s)$ for all pairs (i,j) such that $1 \leq i < j \leq n^s$. We note that the particular ordering of these pairs is irrelevant in what follows. Therefore, from the structures of (3.10), (3.16), and (3.9)

$$\mathbf{D}_{u^s} - \mathbf{I}_{(n^s)^2} = \mathbf{U}^s\mathbf{\Delta}_{D_{u^s}}\mathbf{U}^{s^T}, \quad \mathbf{C}_{u^s} = \mathbf{U}^s\mathbf{\Delta}_{C_{u^s}}\mathbf{U}^{s^T}$$
$$\mathbf{D}_{u^s,\lambda^s}^T = \mathbf{U}^s\mathbf{\Delta}_{D_{u^s,\lambda^s}} \tag{B.2}$$

with $\mathbf{\Delta}_{D_{u^s}} = \begin{bmatrix}\mathbf{\Delta}'_{D_{u^s}} & \mathbf{O} \\ \mathbf{O} & \times\end{bmatrix}$, where $\mathbf{\Delta}'_{D_{u^s}} = -\mathrm{Diag}(2\alpha_1^s\lambda_1^s + 1, \ldots, 2\alpha_{n^s}^s\lambda_{n^s}^s + 1)$, $\mathbf{\Delta}_{C_{u^s}} = \begin{bmatrix}\mathbf{O} & \mathbf{O} \\ \mathbf{O} & \times\end{bmatrix}$, and $\mathbf{\Delta}_{D_{u^s,\lambda^s}} = \begin{bmatrix}\mathbf{\Delta}'_{D_{u^s,\lambda^s}} \\ \mathbf{O}\end{bmatrix}$, where $\mathbf{\Delta}'_{D_{u^s,\lambda^s}} = 2\,\mathrm{Diag}(\lambda_1^s, \ldots, \lambda_{n^s}^s)$. As such, (B.1) becomes

$$\mathbf{\Delta}_{D_{u^s}}\mathbf{U}^{s^T}\mathbf{C}_{u^s,\lambda^s}^T + \mathbf{\Delta}_{C_{u^s}}\mathbf{\Delta}_{D_{u^s,\lambda^s}} = \mathbf{O}. \tag{B.3}$$

Because $\mathbf{\Delta}_{C_{u^s}}\mathbf{\Delta}_{D_{u^s,\lambda^s}} = \mathbf{O}$ and $\mathbf{\Delta}_{D_{u^s}}$ is a (negative) definite matrix, $\mathbf{C}_{u^s,\lambda^s}^T = \mathbf{O}$. Finally, from (3.18), we get $\mathbf{C}_{\lambda^s} = \frac{1}{2}(\mathbf{G}_{\lambda^s} + \mathbf{D}_{u^s,\lambda^s}\mathbf{C}_{u^s,\lambda^s}^T + \mathbf{C}_{u^s,\lambda^s}\mathbf{D}_{u^s,\lambda^s}^T) = \frac{1}{2}\mathbf{G}_{\lambda^s} = \mathrm{Diag}((\lambda_1^s)^2, \ldots, (\lambda_{n^s}^s)^2)$, where the last equality uses (3.15).

References

[1] A. Bunse-Gerstner, R. Byers, and V. Mehrmann, "A chart of numerical methods for structured eigenvalue problems," *SIAM J. Matrix Anal. Appl.*, vol. 13, no. 2, pp. 419–453, Apr. 1992.

[2] A. Cantoni and P. Butler, "Eigenvalues and eigenvectors of symmetric centrosymmetric matrices," *Linear Algebra Appl.*, vol. 13, pp. 275–288, 1976.

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

[4] F. Vanpoucke, S. H. Jensen, and M. Moonen, "A persymmetric adaptive eigenfilter-bank for real data," in *Proc. Asilomar Conf.*, Oct. 1995, pp. 1342–1346.

[5] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, 2nd ed. New York: Wiley, 1984.

[6] P. Delsarte and Y. Genin, "Spectral properties of finite Toeplitz matrices," in *Proc. Int. Symp. Math. Theory Networks Syst.*, 1984, vol. 58, pp. 194–213.

[7] J. Oja, *Subspace Methods of Pattern Recognition.* Letchworth, U.K.: Res. Studies; Wiley, 1983.

[8] A. Benveniste, M. Métivier, and P. Priouret, *Adaptive Algorithms and Stochastic Approximations.* New York: Springer Verlag, 1990.

[9] J. P. Delmas and F. Alberge, "Asymptotic performance analysis of subspace adaptive algorithms introduced in the neural network literature," *IEEE Trans. Signal Processing*, vol. 46, pp. 170–182, Jan. 1998.

[10] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Process.*, no. 45, pp. 59–83, 1995.

[11] K. Hornik and C. M. Kuan, "Convergence analysis of local feature extraction algorithms," *Neural Networks*, vol. 5, pp. 229–240, 1992.

[12] E. Oja, "A simplified neuron model as a principal components analyzer," *J. Math. Biol.*, vol. 15, pp. 267–273, 1982.

[13] E. Oja and J. Karhunen, "On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix," *J. Math. Anal. Appl.*, vol. 106, pp. 69–84, 1985.