Signal Processing 78 (1999) 101–116

SIGNAL PROCESSING

www.elsevier.nl/locate/sigpro

On eigenvalue decomposition estimators of centro-symmetric covariance matrices

Jean-Pierre Delmas*

Institut National des Télécommunications, 9 rue Charles Fourier, 91011 Evry Cedex, France

Received 2 July 1998; received in revised form 11 March 1999

Abstract

This paper is focused on estimators, both batch and adaptive, of the eigenvalue decomposition (EVD) of centro-symmetric (CS) covariance matrices. These estimators make use of the property that eigenvectors and eigenvalues of such structured matrices can be estimated via two decoupled eigensystems. As a result, the number of operations is roughly halved, and moreover, the statistical properties of the estimators are improved. After deriving the asymptotic distribution of the EVD estimators, the closed-form expressions of the asymptotic bias and covariance of the EVD estimators are compared to those obtained when the CS structure is not taken into account. As a by-product, we show that the closed-form expressions of the asymptotic bias and covariance of the batch and adaptive EVD estimators are very similar provided that the number of samples is replaced by the inverse of the step size. Finally, the accuracy of our asymptotic analysis is checked by numerical simulations, and it is found that the convergence speed is also improved thanks to the use of the CS structure. © 1999 Elsevier Science B.V. All rights reserved.

Zusammenfassung

Dieser Artikel fokusiert auf Schätzer der Eigenwertzerlegung (EVD) von zentrosymmetrischen (CS) Kovarianzmatrizen, wobei sowohl nicht-adaptive als adaptive Schätzer betrachtet werden. Diese Schätzer nutzen die Eigenschaft aus, dass Eigenvektoren und -werte solcher strukturierten Matrizen durch zwei entkoppelte Eigensysteme geschätzt werden können. Das Verfahren halbiert fast die Anzahl der Operationen und überdies werden die statistischen Eigenschaften der Schätzer verbessert. Zunächst wird die asymptotische Verteilung der EVD-Schätzer hergeleitet und anschließend werden die entsprechenden Ausdrücke des asymptotischen Bias und der asymptotischen Kovarianz in geschlossener Form mit denjenigen Schätzern verglichen, die daraus resultieren, dass die CS-Struktur nicht berücksichtigt wird. Zusätzlich zeigen wir, dass sich die angesprochenen Ausdrücke der nicht-adaptiven und der adaptiven EVD-Schätzer gleichen, falls die Anzahl der Abtastwerte durch die Inverse der Schrittweite ersetzt wird. Schließlich wird die Genauigkeit unserer asymptotischen Analysen durch numerische Simulationen überprüft und es stellt sich heraus, dass durch Ausnutzen der CS-Struktur auch die Konvergenzgeschwindigkeit erhöht wird. © 1999 Elsevier Science B.V. All rights reserved.

Résumé

Cet article est consacré à des estimateurs en block et adaptatifs de décomposition en valeurs/vecteurs propres (EVD) de matrices de covariance de structure centro-symmétrique (CS). Ces estimateurs sont construits à partir de la propriété que l'EVD de telles matrices structurées peut s'obtenir à partir des deux EVD découplées. Il en résulte que la complexité

* Tel.: + 33-1-60-76-46-32; fax: + 33-1-60-76-42-84.
E-mail address: delmas@int-evry.fr (J.-P. Delmas)

PII: S0165-1684(99)00050-X

numérique d'une telle EVD est grossièrement divisée par deux et que les propriétés statistiques des ces estimateurs sont améliorées. Après avoir donnés les distributions asymptotiques de ces estimateurs d'EVD, des expressions analytiques des biais et des covariances asymptotiques sont comparées aux expressions obtenues sans tenir compte de la structure CS. Nous montrons en outre que les expressions des biais et des covariances asymptotiques des estimateurs en block et adaptatifs ont des structures similaires si le nombre d'échantillons de l'estimateur en block est remplacé par l'inverse du pas d'adaptation de l'estimateur adaptatif. Finalement la pertinence de l'analyse asymptotique est confirmée par des simulations et l'on montre que la vitesse de convergence des estimateurs est aussi améliorée lorsque la structure CS est prise en compte.

Keywords: Eigenvalue decomposition; Centro-symmetric covariance matrices; Asymptotic bias and covariance of eigenvectors and eigenvalues estimators; Asymptotic distribution of batch and adaptive estimators

1. Introduction

Signal processing applications often lead to structured matrix problems. Algorithms which take this structure into account usually require fewer computations, and have better numerical properties [4]. An important matrix structure is the symmetric centro-symmetric structure, for which the symmetric Toeplitz structure is a particular case. This structure occurs in engineering problems [6] and in signal processing 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 symmetric centro-symmetric sensor arrays (for example on uniform linear arrays) [21] are centro-symmetric; spatio-temporal covariance matrices used in subspace methods for blind identification of multichannel FIR filters [15] are block-symmetric centro-symmetric. 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-sized symmetric real matrices [5]. This property has already been used in [8,9] and in [19] for, respectively, a parameterized adaptive eigenspace algorithm and an adaptive eigenfilter bank. But no asymptotic performance analysis has been done yet. The purpose of this paper is to give the asymptotic bias, covariance and distribution of batch and adaptive EVD estimators. The estimators we study are derived from the maximum likelihood principle, in the bach case, and from the Stochastic Gradient Ascent algorithm (SGA), in the adaptive case, and take into account the CS structure.

This paper is organized as follows. In Section 2, 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 3 (respectively in Section 4), we study the bias, covariance and asymptotic distribution of batch (respectively of adaptive) estimators of EVD of CS covariance matrices. In particular, two theorems give closed form expressions of the covariance of the limiting distributions of such an estimated eigenvalue decomposition. Finally, in Section 5 we present some simulations with two purposes. On the one hand, we examine the accuracy of the expressions of the mean square error of our estimators and investigate the step size domain for which our asymptotic approach is valid. On the other hand, we examine performance criteria for which no analytic results were obtained in the preceding sections, such as the convergence speed, which happens to be improved when the CS structure is taken into account. The following notations are used throughout the paper. Matrices and vectors are represented by bold upper case and bold lower case characters, respectively. Vectors are by default in column orientation and T stands for transpose. The symbol $\lceil n \rceil$ (respectively $\lfloor n \rfloor$) denotes the first integer larger (respectively smaller) than or equal to n. $\boldsymbol{e}_i$ is the ith unit vector in $\mathscr{R}^{\lceil n/2 \rceil}$ or in $\mathscr{R}^{\lfloor n/2 \rfloor}$. $\mathrm{E}(\cdot)$, $\mathrm{Cov}(\cdot)$, $\mathrm{Tr}(\cdot)$, $\det(\cdot)$ and $\|\cdot\|_{\mathrm{Fro}}$ denote the expectation, the covariance, the trace, the determinant operator and the Frobenius matrix norm, respectively. $\mathrm{Vec}(\cdot)$ is the "vectorization" operator that turns a matrix into a vector consisting of the columns of the matrix stacked one below another. It is used in conjunction with the Kronecker product $A\otimes B$ as the block matrix, the (i,j) block element of which is $a_{i,j}\boldsymbol{B}$. $\delta_{i,j}$ denotes the Kronecker notation: $\delta_{i,j}=1$ if $i=j$ and $\delta_{i,j}=0$ otherwise. $\mathrm{Diag}(a_1,\dots,a_n)$ is a diagonal matrix with diagonal elements a_i and $\mathrm{Diag}(\boldsymbol{A}_1,\dots,\boldsymbol{A}_n)$ is a block diagonal matrix with block-diagonal entries $\boldsymbol{A}_i$.

2. Eigenvalue decomposition structure

We consider an $n\times n$ centro-symmetric covariance matrix $\boldsymbol{R}_x=\mathrm{E}(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}})$ of a Gaussian-distributed, zero mean real random vector $\boldsymbol{x}$, and we denote by $\lambda_1>\cdots>\lambda_n$, the distinct eigenvalues of $\boldsymbol{R}_x$ and by $\boldsymbol{v}_1,\dots,\boldsymbol{v}_n$ corresponding normalized eigenvectors. The EVD estimators that we propose stem from the property that an orthonormal eigenbasis of $\boldsymbol{R}_x$ can be obtained from orthonormal eigenbases of half-sized symmetric matrices [5]. This property is recalled here for convenience of the reader and in order to fix notations. For n, respectively even and odd, $\boldsymbol{R}_x$ can be partitioned as follows:

$$\boldsymbol{R}_x=\begin{bmatrix}\boldsymbol{R}_1 & \boldsymbol{R}_2^{\mathrm{T}}\\ \boldsymbol{R}_2 & \boldsymbol{J}\boldsymbol{R}_1\boldsymbol{J}\end{bmatrix},\qquad \boldsymbol{R}_x=\begin{bmatrix}\boldsymbol{R}_1 & \boldsymbol{r} & \boldsymbol{R}_2^{\mathrm{T}}\\ \boldsymbol{r}^{\mathrm{T}} & r_0 & \boldsymbol{r}^{\mathrm{T}}\boldsymbol{J}\\ \boldsymbol{R}_2 & \boldsymbol{J}\boldsymbol{r} & \boldsymbol{J}\boldsymbol{R}_1\boldsymbol{J}\end{bmatrix}, \quad (1)$$

where $\boldsymbol{J}$ is an $\lfloor n/2\rfloor\times\lfloor n/2\rfloor$ matrix with ones on its anti-diagonal and zeros elsewhere, and $\boldsymbol{R}_1=\boldsymbol{R}_1^{\mathrm{T}}$, $\boldsymbol{J}\boldsymbol{R}_2=\boldsymbol{R}_2^{\mathrm{T}}\boldsymbol{J}$. Furthermore, $\boldsymbol{R}_x$ can be reduced to a block diagonal form by a data independent orthogonal transformation $\boldsymbol{K}$:

$$\boldsymbol{R}_x=\boldsymbol{K}\begin{bmatrix}\boldsymbol{R}^- & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{R}^+\end{bmatrix}\boldsymbol{K}^{\mathrm{T}}, \quad (2)$$

with respectively for n even and n odd:

$$\boldsymbol{R}^-=\boldsymbol{R}_1-\boldsymbol{J}\boldsymbol{R}_2,\qquad \boldsymbol{R}^+=\boldsymbol{R}_1+\boldsymbol{J}\boldsymbol{R}_2,$$

$$\boldsymbol{K}=\frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{I}\\ -\boldsymbol{J} & \boldsymbol{J}\end{bmatrix}, \quad (3)$$

$$\boldsymbol{R}^-=\boldsymbol{R}_1-\boldsymbol{J}\boldsymbol{R}_2,\qquad \boldsymbol{R}^+=\begin{bmatrix}r_0 & \sqrt{2}\boldsymbol{r}^{\mathrm{T}}\\ \sqrt{2}\boldsymbol{r} & \boldsymbol{R}_1+\boldsymbol{J}\boldsymbol{R}_2\end{bmatrix},$$

$$\boldsymbol{K}=\frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{0} & \boldsymbol{I}\\ \boldsymbol{0}^{\mathrm{T}} & \sqrt{2} & \boldsymbol{0}^{\mathrm{T}}\\ -\boldsymbol{J} & \boldsymbol{0} & \boldsymbol{J}\end{bmatrix}, \quad (4)$$

where $\boldsymbol{I}$ is the $\lfloor n/2\rfloor\times\lfloor n/2\rfloor$ identity matrix.

So $\lfloor n/2\rfloor$ skew-symmetric and $\lceil n/2\rceil$ symmetric orthonormal eigenvectors of $\boldsymbol{R}_x$ (denoted[1] respectively by $\boldsymbol{v}_1^-,\dots,\boldsymbol{v}_{\lfloor n/2\rfloor}^-$ and $\boldsymbol{v}_1^+,\dots,\boldsymbol{v}_{\lceil n/2\rceil}^+$), and corresponding eigenvalues (denoted respectively by $\lambda_1^-,\dots,\lambda_{\lfloor n/2\rfloor}^-$ and $\lambda_1^+,\dots,\lambda_{\lceil n/2\rceil}^+$) are determined from the solutions of the equations:

$$\boldsymbol{R}^-\boldsymbol{u}_i^-=\lambda_i^-\boldsymbol{u}_i^-,\quad i=1,\dots,\lfloor n/2\rfloor$$

and (5)

$$\boldsymbol{R}^+\boldsymbol{u}_i^+=\lambda_i^+\boldsymbol{u}_i^+,\quad i=1,\dots,\lceil n/2\rceil,$$

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

$$\boldsymbol{v}_i^s=\boldsymbol{K}_e^s\boldsymbol{u}_i^s,\quad [\text{respectively } \boldsymbol{v}_i^s=\boldsymbol{K}_o^s\boldsymbol{u}_i^s],$$

$$s=-,\quad i=1,\dots,\lfloor n/2\rfloor,$$

$$s=+,\quad i=1,\dots,\lceil n/2\rceil, \quad (6)$$

with

$$\boldsymbol{K}_e^-\stackrel{\mathrm{def}}{=}\frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{I}\\ -\boldsymbol{J}\end{bmatrix},\qquad \boldsymbol{K}_e^+\stackrel{\mathrm{def}}{=}\frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{J}\end{bmatrix}$$

and

$$\boldsymbol{K}_o^-\stackrel{\mathrm{def}}{=}\frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{I}\\ \boldsymbol{0}^{\mathrm{T}}\\ -\boldsymbol{J}\end{bmatrix},\qquad \boldsymbol{K}_o^+\stackrel{\mathrm{def}}{=}\frac{1}{\sqrt{2}}\begin{bmatrix}\boldsymbol{0} & \boldsymbol{I}\\ \sqrt{2} & \boldsymbol{0}^{\mathrm{T}}\\ \boldsymbol{0} & \boldsymbol{J}\end{bmatrix}.$$

Moreover, the set $\{\boldsymbol{v}_1^-,\dots,\boldsymbol{v}_{\lfloor n/2\rfloor}^-,\boldsymbol{v}_1^+,\dots,\boldsymbol{v}_{\lceil n/2\rceil}^+\}$ forms an orthonormal set which therefore spans the eigenspace of $\boldsymbol{R}_x$.

[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.

3. Batch estimator

3.1. Maximum likelihood estimator

We consider in this section, the maximum likelihood (ML) estimation of the EVD of $\boldsymbol{R}_x$ from t samples $\boldsymbol{X}_t \stackrel{\text{def}}{=} (\boldsymbol{x}_1,\dots,\boldsymbol{x}_t)$. Let $\begin{bmatrix}\boldsymbol{y}_t^-\\ \boldsymbol{y}_t^+\end{bmatrix} \stackrel{\text{def}}{=} \boldsymbol{K}^{\mathrm{T}}\boldsymbol{x}_t$, $\boldsymbol{Y}_t^- \stackrel{\text{def}}{=} (\boldsymbol{y}_1^-,\dots,\boldsymbol{y}_t^-)$, $\boldsymbol{Y}_t^+ \stackrel{\text{def}}{=} (\boldsymbol{y}_1^+,\dots,\boldsymbol{y}_t^+)$ and $\boldsymbol{Y}_t \stackrel{\text{def}}{=} \begin{bmatrix}\boldsymbol{Y}_t^-\\ \boldsymbol{Y}_t^+\end{bmatrix}$. The Gaussian probability density of $\boldsymbol{X}_t$, parameterized by $(\boldsymbol{R}^-,\boldsymbol{R}^+)$, is

$$f(\boldsymbol{X}_t;(\boldsymbol{R}^-,\boldsymbol{R}^+)) = |\det \boldsymbol{K}|\, f(\boldsymbol{Y}_t;(\boldsymbol{R}^-,\boldsymbol{R}^+)) = f(\boldsymbol{Y}_t;(\boldsymbol{R}^-,\boldsymbol{R}^+)). \quad (7)$$

Then, since

$$\mathrm{E}\begin{bmatrix}\boldsymbol{y}_t^-\\ \boldsymbol{y}_t^+\end{bmatrix}[\boldsymbol{y}_t^{-\mathrm{T}} \quad \boldsymbol{y}_t^{+\mathrm{T}}] = \begin{bmatrix}\boldsymbol{R}^- & \boldsymbol{O}\\ \boldsymbol{O} & \boldsymbol{R}^+\end{bmatrix}, \quad (8)$$

we have

$$f(\boldsymbol{Y}_t;(\boldsymbol{R}^-,\boldsymbol{R}^+)) = f(\boldsymbol{Y}_t^-;\boldsymbol{R}^-)f(\boldsymbol{Y}_t^+;\boldsymbol{R}^+). \quad (9)$$

So the ML estimators of $(\boldsymbol{R}^-,\boldsymbol{R}^+)$ given $\boldsymbol{X}_t$ are obtained from the *ML* estimators of $\boldsymbol{R}^-$ given $\boldsymbol{Y}_t^-$, and of $\boldsymbol{R}^+$ given $\boldsymbol{Y}_t^+$. According to a classical result (see e.g. [1, Theorem 3.2.1]), these ML estimators are, respectively, the sample covariance matrices $\boldsymbol{R}_t^- \stackrel{\text{def}}{=} \frac{1}{t}\sum_{k=1}^t \boldsymbol{y}_k^-\boldsymbol{y}_k^{-\mathrm{T}}$ and $\boldsymbol{R}_t^+ \stackrel{\text{def}}{=} \frac{1}{t}\sum_{k=1}^t \boldsymbol{y}_k^+\boldsymbol{y}_k^{+\mathrm{T}}$. Thus, according to the invariance property[2] of the ML estimator (see e.g. [1, Corollary 3.2.1]), we have the following theorem.

Theorem 1. *Let $(\boldsymbol{x}_1,\dots,\boldsymbol{x}_t)$ be a sample from a zero mean Gaussian distribution with centro-symmetric covariance matrix $\boldsymbol{R}_x$. Then a set of ML estimators of the EVD of $\boldsymbol{R}_x$ can be computed from the EVD of the sample covariance matrices $\boldsymbol{R}_t^-$ and $\boldsymbol{R}_t^+$. The eigenvalues ML estimator is $\lambda_{1,t}^-,\dots,\lambda_{\lfloor n/2\rfloor,t}^-,\lambda_{1,t}^+,\dots,\lambda_{\lceil n/2\rceil,t}^+$; i.e., the eigenvalues of $\boldsymbol{R}_t^-$ and $\boldsymbol{R}_t^+$. The eigenvectors ML estimator is $\boldsymbol{v}_{1,t}^-,\dots,\boldsymbol{v}_{\lfloor n/2\rfloor,t}^-,\boldsymbol{v}_{1,t}^+,\dots,\boldsymbol{v}_{\lceil n/2\rceil,t}^+$, which is deduced from the eigenvectors $\boldsymbol{u}_{1,t}^-,\dots,\boldsymbol{u}_{\lfloor n/2\rfloor,t}^-$ of $\boldsymbol{R}_t^-$ and $\boldsymbol{u}_{1,t}^+,\dots,\boldsymbol{u}_{\lceil n/2\rceil,t}^+$ of $\boldsymbol{R}_t^+$ by the relation* (6).

We note that this procedure yields the same estimator that the EVD of the symmetric centro-symmetric (often called Forward–Backward) sample covariance matrix:

$$\boldsymbol{R}_t^{\mathrm{FB}} \stackrel{\text{def}}{=} \frac{1}{2}(\boldsymbol{R}_t + \boldsymbol{J}\boldsymbol{R}_t\boldsymbol{J}), \quad \text{with } \boldsymbol{R}_t \stackrel{\text{def}}{=} \frac{1}{t}\sum_{k=1}^t \boldsymbol{x}_k\boldsymbol{x}_k^{\mathrm{T}}. \quad (10)$$

From the complexity and precision point of view, it is well known [12] that the EVD of a covariance matrix should in practice be replaced by the SVD of the associated data matrix. For example, an EVD of an $n\times n$ covariance matrix requires $\mathrm{O}(n^3)$ flops by iteration for the EVD computation and $\mathrm{O}(n^2t)$ flops for the covariance computation. Computing the covariance matrix is therefore the major computational burden of the whole algorithm. On the other hand, the SVD of the associated $n\times t$ data matrix requires only $\mathrm{O}(nt)$ flops by iteration (if for instance one uses the bidiagonalisation Lanczos method [12,20]). So, as the data matrix $\frac{1}{\sqrt{2}}(\boldsymbol{X}_t,\boldsymbol{J}\boldsymbol{X}_t)$ associated to $\boldsymbol{R}_t^{\mathrm{FB}}$ doubles the number of data samples, the equivalent procedure which consists in performing the SVD of the $\lfloor n/2\rfloor\times t$ and $\lceil n/2\rceil\times t$ data matrices $\boldsymbol{Y}_t^-$ and $\boldsymbol{Y}_t^+$, roughly reduces the complexity by half, while avoiding less of accuracy due to squaring up data.

3.2. Asymptotic distribution

Since the Gaussian $2n$-vector $\mathrm{Vec}(\boldsymbol{y}_k^-,\boldsymbol{y}_k^+)$ is composed of two Gaussian uncorrelated vectors (8), $\boldsymbol{y}_k^-$ and $\boldsymbol{y}_k^+$ are independent random variables. This implies that $\boldsymbol{R}_t^-$ and $\boldsymbol{R}_t^+$ are independent. So the random vectors Θ_t^- and Θ_t^+, with $\Theta_t^s \stackrel{\text{def}}{=} \mathrm{Vec}(\boldsymbol{u}_{1,t}^s,\dots,\boldsymbol{u}_{n^s,t}^s,\lambda_{1,t}^s,\dots,\lambda_{n^s,t}^s)$, $s=-,+$ are independent. n^s used throughout the paper denotes $\lfloor n/2\rfloor$ for $s=-$ and $\lceil n/2\rceil$ for $s=+$. Let $\boldsymbol{u}_t^s \stackrel{\text{def}}{=} \mathrm{Vec}(\boldsymbol{u}_{1,t}^s,\dots,\boldsymbol{u}_{n^s,t}^s)$, $\boldsymbol{u}^s \stackrel{\text{def}}{=} \mathrm{Vec}(\boldsymbol{u}_1^s,\dots,\boldsymbol{u}_{n^s}^s)$, $(\Lambda_t^s \stackrel{\text{def}}{=} (\lambda_{1,t}^s,\dots,\lambda_{n^s,t}^s)^{\mathrm{T}}$ and $\Lambda^s \stackrel{\text{def}}{=} (\lambda_1^s,\dots,\lambda_{n^s}^s)^{\mathrm{T}})$. Then, according to a classical result (e.g. in [1, Theorem

[2] Eigenvectors are uniquely defined if we require that their first nonzero component be positive.

13.5.1, p. 541]), $\sqrt{t}\,(\mathrm{Vec}(\boldsymbol{u}^s_t, \Lambda^s_t) - \mathrm{Vec}(\boldsymbol{u}^s, \Lambda^s))$ converges in distribution to the zero mean Gaussian distribution of covariance $\boldsymbol{C}_{u^s,\lambda^s} = \mathrm{Diag}(\boldsymbol{C}_{u^s}, \boldsymbol{C}_{\lambda^s})$, where

$$\boldsymbol{C}_{\lambda^s} = \mathrm{Diag}(2(\lambda^s_1)^2, \ldots, 2(\lambda^s_{n^s})^2), \quad s = -, + \tag{11}$$

and where $\boldsymbol{C}_{u^s}$ are the $n^s \times n^s$ matrix, the block $(\boldsymbol{C}_{u^s})_{i,j}$ of which are for $s = -, +$:

$$(\boldsymbol{C}_{u^s})_{i,j} = \begin{cases} \sum_{1 \leqslant k \neq i \leqslant n^s} \dfrac{\lambda^s_i \lambda^s_k}{(\lambda^s_k - \lambda^s_i)^2} \boldsymbol{u}^s_k \boldsymbol{u}^{s^{\mathrm{T}}}_k, & i = j, \\ -\dfrac{\lambda^s_j \lambda^s_i}{(\lambda^s_i - \lambda^s_j)^2} \boldsymbol{u}^s_j \boldsymbol{u}^{s^{\mathrm{T}}}_i, & i \neq j. \end{cases} \tag{12}$$

Lastly, we apply the linear mapping deduced from Eq. (6), in which $\boldsymbol{v}^s \overset{\mathrm{def}}{=} \mathrm{Vec}(\boldsymbol{v}^s_1, \ldots, \boldsymbol{v}^s_{n^s})$ is equal for $s = -, +$:

$$\boldsymbol{v}^s = \begin{cases} \mathrm{Diag}(\boldsymbol{K}^s_e, \ldots, \boldsymbol{K}^s_e)\boldsymbol{u}^s, & \text{for } n \text{ even}, \\ \mathrm{Diag}(\boldsymbol{K}^s_o, \ldots, \boldsymbol{K}^s_o)\boldsymbol{u}^s, & \text{for } n \text{ odd}, \end{cases} \tag{13}$$

Theorem 2 follows immediately.

Theorem 2. $\sqrt{t}\,(\mathrm{Vec}(\boldsymbol{v}^-_t, \Lambda^-_t, \boldsymbol{v}^+_t, \Lambda^+_t) - \mathrm{Vec}(\boldsymbol{v}^-, \Lambda^-, \boldsymbol{v}^+, \Lambda^+))$ *converges in distribution to the zero mean Gaussian distribution of covariance* $\boldsymbol{C}_{v,\lambda}$ *with* $\boldsymbol{C}_{v,\lambda} = \mathrm{Diag}(\boldsymbol{C}_{v^-}, \boldsymbol{C}_{\lambda^-}, \boldsymbol{C}_{v^+}, \boldsymbol{C}_{\lambda^+})$,

$$\boldsymbol{C}_{\lambda^s} = \mathrm{Diag}(2(\lambda^s_1)^2, \ldots, 2(\lambda^s_{n^s})^2), \quad s = -, +, \tag{14}$$

$$\boldsymbol{C}_{v^s} = \sum_{1 \leqslant i \neq j \leqslant n^s} \frac{\lambda^s_i \lambda^s_j}{(\lambda^s_i - \lambda^s_j)^2} \boldsymbol{e}_i \boldsymbol{e}^{\mathrm{T}}_i \otimes \boldsymbol{v}^s_j \boldsymbol{v}^{s^{\mathrm{T}}}_j - \sum_{1 \leqslant i \neq j \leqslant n^s} \frac{\lambda^s_i \lambda^s_j}{(\lambda^s_i - \lambda^s_j)^2} \boldsymbol{e}_i \boldsymbol{e}^{\mathrm{T}}_j \otimes \boldsymbol{v}^s_j \boldsymbol{v}^{s^{\mathrm{T}}}_i, \quad s = -, +. \tag{15}$$

This result is quite comparable to the classical results obtained when the centro-symmetric structure is not taken into account ([1, Theorem 13.5.1, p. 541]). The only difference lies in the uncorrelation between the skew-symmetric and the symmetric estimated eigenvectors, and in the summations (15), which are only done over the skew-symmetric (respectively the symmetric) eigenvectors.

3.3. Asymptotic bias and asymptotic MSE

Asymptotic bias. By performing a Taylor expansion of λ^-_i and $\boldsymbol{u}^-_i$ (respectively λ^+_i and $\boldsymbol{u}^+_i$) in the neighborhood of $\boldsymbol{R}^-$ (respectively $\boldsymbol{R}^+$), Kaveh et al. [14] showed for complex data a result that can be directly carried over to the real data case:

$$\mathrm{E}(\lambda^s_{k,t}) = \lambda^s_k + \mathrm{o}\left(\frac{1}{t}\right), \quad s = -, +, \tag{16}$$

$$\mathrm{E}(\boldsymbol{u}^s_{k,t}) = \boldsymbol{u}^s_k - \frac{\lambda^s_k}{2t}\left(\sum_{1 \leqslant j \neq k \leqslant n^s} \frac{\lambda^s_j}{(\lambda^s_j - \lambda^s_k)^2}\right)\boldsymbol{u}^s_k + \mathrm{o}\left(\frac{1}{t}\right), \quad s = -, +. \tag{17}$$

So, thanks to the linear mapping $\boldsymbol{u}^s_k \to \boldsymbol{v}^s_k$ (formulas (6)), Eq. (17) holds for the bias of $\boldsymbol{v}^s_{k,t}$ by replacing $\boldsymbol{u}^s_k$ by $\boldsymbol{v}^s_k$:

$$\mathrm{E}(\boldsymbol{v}^s_{k,t}) = \boldsymbol{v}^s_k - \frac{\lambda^s_k}{2t}\left(\sum_{1 \leqslant j \neq k \leqslant n^s} \frac{\lambda^s_j}{(\lambda^s_j - \lambda^s_k)^2}\right)\boldsymbol{v}^s_k + \mathrm{o}\left(\frac{1}{t}\right), \quad s = -, +. \tag{18}$$

Asymptotic MSE. A simple global measure of performance of our batch estimator is the MSE between $\boldsymbol{v}_t \overset{\mathrm{def}}{=} \mathrm{Vec}(\boldsymbol{v}_{1,t}, \ldots, \boldsymbol{v}_{n,t})$ and $\boldsymbol{v} \overset{\mathrm{def}}{=} \mathrm{Vec}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$, and between $\Lambda_t \overset{\mathrm{def}}{=} (\lambda_{1,t}, \ldots, \lambda_{n,t})^{\mathrm{T}}$ and $\Lambda \overset{\mathrm{def}}{=} (\lambda_1, \ldots, \lambda_n)^{\mathrm{T}}$. These MSE are obtained from the asymptotic bias (16), (18) and from the asymptotic covariances $\mathrm{cov}(\Lambda^s_t)$ and $\mathrm{cov}(\boldsymbol{u}^s_t)$. It has been shown [3, Theorem 9.22, p. 340] that

$$\mathrm{Cov}(\Lambda^s_t) = \frac{1}{t}\boldsymbol{C}_{\lambda^s} + \mathrm{O}\left(\frac{1}{t^2}\right) \quad \text{and} \quad \mathrm{Cov}(\boldsymbol{u}^s_t) = \frac{1}{t}\boldsymbol{C}_{u^s} + \mathrm{O}\left(\frac{1}{t^2}\right). \tag{19}$$

So, from Eqs. (16), (17) and (19), we get according to Theorem 2

$$\mathrm{E}\|\Lambda_t - \Lambda\|^2_{\mathrm{Fro}} = \frac{2}{t}\sum_{k=1}^{n} \lambda^2_k + \mathrm{O}\left(\frac{1}{t^2}\right), \tag{20}$$

$$\mathrm{E}\|\boldsymbol{v}_t - \boldsymbol{v}\|^2_{\mathrm{Fro}} = \frac{1}{t}\left(\sum_{1 \leqslant j \neq k \leqslant \lfloor n/2 \rfloor} \frac{\lambda^-_j \lambda^-_k}{(\lambda^-_j - \lambda^-_k)^2} + \sum_{1 \leqslant j \neq k \leqslant \lceil n/2 \rceil} \frac{\lambda^+_j \lambda^+_k}{(\lambda^+_j - \lambda^+_k)^2}\right) + \mathrm{O}\left(\frac{1}{t^2}\right). \tag{21}$$

3.4. Analysis of the results

Let us now compare the results to the case where the CS structure is not taken into account ([14] and [1, Theorem 13.5.1, p. 541]), which we now recall for convenience of the reader:

$$\mathrm{E}(\lambda_{k,t}) = \lambda_k + \mathrm{o}\left(\frac{1}{t}\right), \tag{22}$$

$$\mathrm{E}(\boldsymbol{v}_{k,t}) = \boldsymbol{v}_k - \frac{\lambda_k}{2t}\left(\sum_{1 \leqslant j \neq k \leqslant n} \frac{\lambda_j}{(\lambda_j - \lambda_k)^2}\right)\boldsymbol{v}_k + \mathrm{o}\left(\frac{1}{t}\right), \tag{23}$$

$$\mathrm{E}\|\Lambda_t - \Lambda\|^2_{\mathrm{Fro}} = \frac{2}{t}\sum_{k=1}^{n} \lambda_k^2 + \mathrm{O}\left(\frac{1}{t^2}\right),$$
$$\mathrm{E}\|\boldsymbol{v}_t - \boldsymbol{v}\|^2_{\mathrm{Fro}} = \frac{1}{t}\left(\sum_{1 \leqslant j \neq k \leqslant n} \frac{\lambda_j\lambda_k}{(\lambda_j - \lambda_k)^2}\right) + \mathrm{O}\left(\frac{1}{t^2}\right). \tag{24}$$

We note that if the asymptotic bias and the asymptotic MSE of the estimated eigenvalues are unchanged, the asymptotic bias and the asymptotic MSE of the estimated eigenvectors are reduced when the CS structure is taken into account. The number of terms in the summations (18) and (21) 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 bias and the asymptotic MSE can be considerably reduced. Necessary conditions for this interlaced distribution are given in [5] for the general CS structure and in [11] for the Toeplitz structure.

4. Adaptive estimator

4.1. Adaptive algorithm

For batch estimation, the estimator we proposed in Section 3 is the ML estimator. For adaptive estimation, many estimators are available. Any gradient-type algorithm or RLS-type algorithm (i.e. [22]) built upon $\boldsymbol{x}_t$ can be split into two decoupled algorithms in which one is built upon $\boldsymbol{y}_t^-$ and the other upon $\boldsymbol{y}_t^+$. We propose to use, as 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 [10]. Its convergence is studied in [17] and, as was shown in [10], it achieves a good convergence speed/misadjustment tradeoff among a family of numerically simple algorithms. Of course, the study set out in this section could be immediately extended to other gradient-type algorithm. For exemple, the generalized Hebbian algorithm (GHA), the weighted subspace algorithm (WSA) and the optimal fitting analyser (OFA), the distribution of which are derived in [10], can be studied along the same lines. Adapted to our structured situation, the SGA algorithm splits into two decoupled algorithms:

$$\boldsymbol{u}^s_{k,t+1} = \boldsymbol{u}^s_{k,t} + \alpha^s_k\gamma\left[\boldsymbol{I}_{n^s} - \boldsymbol{u}^s_{k,t}\boldsymbol{u}^{s^{\mathrm{T}}}_{k,t} - \sum_{i=1}^{k-1}\left(1 + \frac{\alpha^s_i}{\alpha^s_k}\right)\boldsymbol{u}^s_{i,t}\boldsymbol{u}^{s^{\mathrm{T}}}_{i,t}\right]\boldsymbol{y}^s_t\boldsymbol{y}^{s^{\mathrm{T}}}_t\boldsymbol{u}^s_{k,t}, \tag{25}$$

$$\lambda^s_{k,t+1} = \lambda^s_{k,t} + \gamma[\boldsymbol{u}^{s^{\mathrm{T}}}_{k,t}\boldsymbol{y}^s_t\boldsymbol{y}^{s^{\mathrm{T}}}_t\boldsymbol{u}^s_{k,t} - \lambda^s_{k,t}], \tag{26}$$

for $s = -, +$ and $k = 1,\dots,n^s$ ($n^- \overset{\mathrm{def}}{=} \lfloor n/2 \rfloor$ and $n^+ \overset{\mathrm{def}}{=} \lceil n/2 \rceil$). $\boldsymbol{u}^-_{k,t}$ (respectively $\boldsymbol{u}^+_{k,t}$) is associated to the $\lfloor n/2 \rfloor$ skew-symmetric eigenvectors $\boldsymbol{v}_i$, (respectively the $\lceil n/2 \rceil$ symmetric eigenvectors $\boldsymbol{v}_i$). The parameters α^s_k ($\alpha^s_1 = 1$ and $\alpha^s_k > 0$, $k = 1,\dots,n^s$) affords a better tradeoff between the convergence speed and misadjustment [10], and γ is the step size. As the computational cost of the SGA algorithm is $\mathrm{O}(n^2)$ flops by iteration, the number of operations of our split procedure is roughly halved. To evaluate the asymptotic distribution of this EVD estimator, we shall use a general approximation result [2, Theorem 2, p. 108] which we now recall for convenience of the reader.

4.2. Asymptotic distribution

4.2.1. A short review of a general Gaussian approximation result

Consider a constant step size recursive stochastic algorithm (we write Θ^γ_t for the sequence of estimates to emphasize the dependence on γ):

$$\Theta^{\gamma}_{t+1} = \Theta^{\gamma}_t + \gamma g(\Theta^{\gamma}_t, \boldsymbol{x}_t), \tag{27}$$

with $\boldsymbol{x}_t = h(\xi_t)$, where ξ_t is a Markov chain independent of Θ_t. The field $g(\Theta, \boldsymbol{x})$ is the function which essentially defines how the parameter Θ^{γ}_t is updated as a function of new observation $\boldsymbol{x}_t$. Suppose that the parameter vector Θ^{γ}_t converges almost surely to the unique asymptotically stable point Θ_* in the corresponding decreasing step-size algorithm. Consider the continuous Lyapunov equation:

$$\boldsymbol{D}\boldsymbol{C}_{\Theta} + \boldsymbol{C}_{\Theta}\boldsymbol{D}^{\mathrm{T}} + \boldsymbol{G} = \boldsymbol{O}, \tag{28}$$

where $\boldsymbol{D}$ and $\boldsymbol{G}$ are, respectively, the derivative of the mean field and the covariance of the field of the algorithm (27):

$$\boldsymbol{D} \overset{\text{def}}{=} \mathrm{E}\left[\frac{\partial g}{\partial \Theta}(\Theta, \boldsymbol{x}_t)\right]_{\Theta = \Theta_*}, \tag{29}$$

$$\boldsymbol{G} \overset{\text{def}}{=} \sum_{t=-\infty}^{\infty} \mathrm{cov}[g(\Theta_*, \boldsymbol{x}_t), g(\Theta_*, \boldsymbol{x}_0)]. \tag{30}$$

If all the eigenvalues of the derivative $\boldsymbol{D}$ of the mean field have strictly negative real parts, then, in a stationary situation, for γ arbitrarily small, we have as $t \to \infty$:

$$\frac{1}{\sqrt{\gamma}}(\Theta^{\gamma}_t - \Theta_*) \overset{\mathscr{L}}{\to} \mathscr{N}(\boldsymbol{0}, \boldsymbol{C}_{\Theta}), \tag{31}$$

where $\boldsymbol{C}_{\Theta}$ is the unique symmetric solution of the Lyapunov equation (28).

4.2.2. Local caracterisation of the field

With $\Theta_t = \mathrm{Vec}(\Theta^-_t, \Theta^+_t)$ with Θ^s_t, $s = -, +$ defined in Section 3.2, the SGA algorithms (25) and (26) can be globally written in a form similar to that of Eq. (27). According to the previous section, one needs to characterize two local properties of the field $g(\Theta_t, \boldsymbol{x}_t)$: the mean value of its derivative and its covariance, both evaluated at the point $\Theta_t = \Theta_*$.

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

$$\boldsymbol{D} = \begin{bmatrix} \boldsymbol{D}_{u^-} & \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{D}_{u^-,\lambda^-} & -\boldsymbol{I}_{n^-} & \boldsymbol{O} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{D}_{u^+} & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{O} & \boldsymbol{D}_{u^+,\lambda^+} & -\boldsymbol{I}_{n^+} \end{bmatrix}, \tag{32}$$

with

$$\boldsymbol{D}_{u^s,\lambda^s} = 2\mathrm{Diag}(\lambda^s_1 \boldsymbol{u}^{s^{\mathrm{T}}}_1, \dots, \lambda^s_{n^s}\boldsymbol{u}^{s^{\mathrm{T}}}_{n^s}), \quad s = -, +. \tag{33}$$

$\boldsymbol{D}_{u^s}$ are $n^s \times n^s$ block matrices, the block $(\boldsymbol{D}_{u^s})_{i,j}$ of which is given in [10] by

$$(\boldsymbol{D}_{u^s})_{i,j} = \begin{cases} -\alpha^s_i[\sum_{k=1}^{i-1}(\lambda^s_i + \frac{\alpha^s_k}{\alpha^s_i}\lambda^s_k)\boldsymbol{u}^s_k\boldsymbol{u}^{s^{\mathrm{T}}}_k + 2\lambda^s_i\boldsymbol{u}^s_i\boldsymbol{u}^{s^{\mathrm{T}}}_i \\ \quad + \sum_{k=i+1}^{n^s}(\lambda^s_i - \lambda^s_k)\boldsymbol{u}^s_k\boldsymbol{u}^{s^{\mathrm{T}}}_k], & i = j, \\ \boldsymbol{O}, & i < j, \\ -\alpha^s_i(1 + \frac{\alpha^s_j}{\alpha^s_i})\lambda^s_i\boldsymbol{u}^s_j\boldsymbol{u}^{s^{\mathrm{T}}}_i, & i > j. \end{cases} \tag{34}$$

Covariance of the field. The field of the algorithms (25) and (26) can be globally written in the form:

$$g(\Theta_t, \boldsymbol{x}_t) = \begin{bmatrix} g^-(\Theta^-_t, \boldsymbol{y}^-_t\boldsymbol{y}^{-^{\mathrm{T}}}_t) \\ g^+(\Theta^+_t, \boldsymbol{y}^+_t\boldsymbol{y}^{+^{\mathrm{T}}}_t) \end{bmatrix} = \begin{bmatrix} g_{u^-}(\Theta^-_t, \boldsymbol{y}^-_t\boldsymbol{y}^{-^{\mathrm{T}}}_t) \\ g_{\lambda^-}(\Theta^-_t, \boldsymbol{y}^-_t\boldsymbol{y}^{-^{\mathrm{T}}}_t) \\ g_{u^+}(\Theta^+_t, \boldsymbol{y}^+_t\boldsymbol{y}^{+^{\mathrm{T}}}_t) \\ g_{\lambda^+}(\Theta^+_t, \boldsymbol{y}^+_t\boldsymbol{y}^{+^{\mathrm{T}}}_t) \end{bmatrix}$$

$$= \begin{bmatrix} \boldsymbol{A}^- & \boldsymbol{O} \\ \boldsymbol{B}^- & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{A}^+ \\ \boldsymbol{O} & \boldsymbol{B}^+ \end{bmatrix} \begin{bmatrix} \mathrm{Vec}(\boldsymbol{y}^-_t\boldsymbol{y}^{-\mathrm{T}}_t) \\ \mathrm{Vec}(\boldsymbol{y}^+_t\boldsymbol{y}^{+\mathrm{T}}_t) \end{bmatrix} - \begin{bmatrix} \boldsymbol{O} \\ \Lambda^- \\ \boldsymbol{O} \\ \Lambda^+ \end{bmatrix}, \tag{35}$$

with

$$\boldsymbol{A}^s \overset{\text{def}}{=} \begin{bmatrix} \boldsymbol{u}^{s^{\mathrm{T}}}_1 \otimes \boldsymbol{A}^s_1 \\ \vdots \\ \boldsymbol{u}^{s^{\mathrm{T}}}_{n^s} \otimes \boldsymbol{A}^s_{n^s} \end{bmatrix} \quad \text{and} \quad \boldsymbol{B}^s \overset{\text{def}}{=} \begin{bmatrix} \boldsymbol{u}^{s^{\mathrm{T}}}_1 \otimes \boldsymbol{u}^{s^{\mathrm{T}}}_1 \\ \vdots \\ \boldsymbol{u}^{s^{\mathrm{T}}}_{n^s} \otimes \boldsymbol{u}^{s^{\mathrm{T}}}_{n^s} \end{bmatrix},$$

$$s = -, +, \tag{36}$$

and

$$A_k^s = \alpha_k^s\left[I_{n^s} - u_{k,t}^s u_{k,t}^{s^{\mathrm{T}}} - \sum_{i=1}^{k-1}\left(1+\frac{\alpha_i^s}{\alpha_k^s}\right)u_{i,t}^s u_{i,t}^{s^{\mathrm{T}}}\right],$$

$$k = 1,\dots,n^s,\ s = -,+. \tag{37}$$

The covariance G of the field evaluated at $\Theta = \Theta_* = \mathrm{Vec}(v^-, \Lambda^-, v^+, \Lambda^+)$, in the case where the observations x_t are independent, can be partitioned as follows:

$$G = \begin{bmatrix} G_{u^-} & G_{u^-,\lambda^-}^{\mathrm{T}} & O & O \\ G_{u^-,\lambda^-} & G_{\lambda^-} & O & O \\ O & O & G_{u^+} & G_{u^+,\lambda^+}^{\mathrm{T}} \\ O & O & G_{u^+,\lambda^+} & G_{\lambda^+} \end{bmatrix}. \tag{38}$$

Using the expression of $\mathrm{Cov}(\mathrm{Vec}(y_t^s y_t^{s^{\mathrm{T}}}))$ given in [10] and after some manipulations, the following relations hold:

$$G_{u^s,\lambda^s} = O \quad \text{and} \quad G_{\lambda^s} = 2\mathrm{Diag}((\lambda_1^s)^2,\dots,(\lambda_{n^s}^s)^2), \tag{39}$$

for $s = -,+$. Here G_{u^s} are the $n^s \times n^s$ matrices, the block $(G_{u^s})_{i,j}$ of which is given in [10] by

$$(G_{u^s})_{i,j} = \begin{cases} \sum_{k=1}^{i-1}(\alpha_k^s)^2\lambda_i^s\lambda_k^s u_k^s u_k^{s^{\mathrm{T}}} + \sum_{k=i+1}^{n^s}(\alpha_i^s)^2\lambda_i^s\lambda_k^s u_k^s u_k^{s^{\mathrm{T}}}], & i = j \\ -(\alpha_{\min(i,j)}^s)^2\lambda_i^s\lambda_j^s u_j^s u_i^{s^{\mathrm{T}}}, & i \neq j. \end{cases} \tag{40}$$

4.2.3. Solution of the Lyapunov equation

For independent observations x_t and for the investigated algorithm, which can be written in a form similar to Eq. (27) with $\xi_t = x_t$ for which the eigenvalues of the derivative (32) of the mean field have strictly negative real parts (see [10] for the eigenvalues of D_{u^s}), the hypotheses of the model of Benveniste et al. ([2, Theorem 2, p. 108]) are fulfilled. But, 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 situation is usually justified (for example in [7]) by using formally a general approximation result ([2, Theorem 1, p. 107]). Furthermore, the almost sure 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. But, as discussed in [13], this condition was proved for only the Oja learning rule [16] designed for extracting the most dominant eigenvector by means of a single linear unit neuron network, where Oja et al. [18] showed that if this algorithm is used with uniformly bounded inputs x_t, then $v_{1,t}$ remains inside some bounded subset. If we allow ourselves the Benveniste results in our situation, the Lyapunov continuous equations can be solved exactly. Since the matrices D and G are 2×2 block diagonal, the Lyapunov equation (28) can be reduced to two decoupled equations. Thus

$$C_\Theta = \mathrm{Diag}(C_{\Theta^-}, C_{\Theta^+}), \tag{41}$$

where $C_{\Theta^s} = \begin{bmatrix} C_{u^s} & C_{u^s,\lambda^s}^{\mathrm{T}} \\ C_{u^s,\lambda^s} & C_{\lambda^s} \end{bmatrix}$ are solutions of the Lyapunov equation:

$$\begin{bmatrix} D_{u^s} & O \\ D_{u^s,\lambda^s} & -I_{n^s} \end{bmatrix}\begin{bmatrix} C_{u^s} & C_{u^s,\lambda^s}^{\mathrm{T}} \\ C_{u^s,\lambda^s} & C_{\lambda^s} \end{bmatrix} + \begin{bmatrix} C_{u^s} & C_{u^s,\lambda^s}^{\mathrm{T}} \\ C_{u^s,\lambda^s} & C_{\lambda^s} \end{bmatrix}\begin{bmatrix} D_{u^s}^{\mathrm{T}} & D_{u^s,\lambda^s}^{\mathrm{T}} \\ O & -I_{n^s} \end{bmatrix} = -\begin{bmatrix} G_{u^s} & O \\ O & G_{\lambda^s} \end{bmatrix}. \tag{42}$$

So C_{u^s} are solutions of the Lyapunov equation: $D_{u^s}C_{u^s} + C_{u^s}D_{u^s}^{\mathrm{T}} + G_{u^s} = O$, the block $(C_{u^s})_{i,j}$ of which is given in [10] by

$$(C_{u^s})_{i,j} = \begin{cases} \sum_{1 \leqslant k \neq i \leqslant n^s} \dfrac{\alpha_{\min(i,k)}^s\lambda_i^s\lambda_k^s}{2|\lambda_i^s - \lambda_k^s|} u_k^s u_k^{s^{\mathrm{T}}}, & i = j, \\ -\dfrac{\alpha_{\min(i,j)}^s\lambda_i^s\lambda_j^s}{2|\lambda_i^s - \lambda_j^s|} u_j^s u_i^{s^{\mathrm{T}}}, & i \neq j, \end{cases} \tag{43}$$

for $s=-,+$. Considering the change of basis stated in [10], Eq. (42) gives the following expressions after some manipulations:

$$C_{u^s,\lambda^s}=O \quad \text{and} \quad C_{\lambda^s}=\mathrm{Diag}((\lambda_1^s)^2,\ldots,(\lambda_{n^s}^s)^2), \tag{44}$$

for $s=-,+$. Along the same steps as for the batch estimation, the following theorem is proved.

Theorem 3. $1/\sqrt{\gamma}(\mathrm{Vec}(v_t^-,\Lambda_t^-,v_t^+,\Lambda_t)-\mathrm{Vec}(v^-,\Lambda^-,v^+,\Lambda^+))$ *converges in distribution* ($t\to\infty$ *and* $\gamma\to 0$) *to the zero mean Gaussian distribution of covariance* $C_{v,\lambda}$ *with* $C_{v,\lambda}=\mathrm{Diag}(C_{v^-},C_{\lambda^-},C_{v^+},C_{\lambda^+})$,

$$C_{\lambda^s}=\mathrm{Diag}((\lambda_1^s)^2,\ldots,\lambda_{n^s}^s)^2), \quad s=-,+, \tag{45}$$

$$C_{v^s}=\sum_{1\leqslant i\neq j\leqslant n^s}\frac{\alpha^s_{\min(i,j)}\lambda_i^s\lambda_j^s}{2|\lambda_i^s-\lambda_j^s|}e_ie_i^{\mathrm T}\otimes v_j^s v_j^{s^{\mathrm T}} - \sum_{1\leqslant i\neq j\leqslant n^s}\frac{\alpha^s_{\min(i,j)}\lambda_i^s\lambda_j^s}{2|\lambda_i^s-\lambda_j^s|}e_ie_j^{\mathrm T}\otimes v_j^s v_i^{s^{\mathrm T}},$$
$$s=-,+. \tag{46}$$

As far as the asymptotic distribution is concerned, similar results could be derived from other gradient-like algorithms such as the generalized Hebbian algorithm (GHA), the weighted subspace algorithm (WSA) and the optimal fitting analyzer (OFA). It would be sufficient to use the asymptotic distributions of their unstructured eigenvectors estimators given in [10].

4.3. Asymptotic bias and asymptotic MSE

Asymptotic bias. Several simple bias and MSE characterizations can be derived from the regular structure of the covariance matrix C_Θ as expressed by Eqs. (45) and (46). A word of caution is nonetheless necessary because the convergence of $1/\sqrt{\gamma}\,(\Theta_t^\gamma-\Theta_*)$ to a limiting Gaussian distribution with covariance matrix C_Θ does not guarantee the convergence of its moments to those of the limiting Gaussian distribution. In batch estimation, both the first and the second moments of the limiting distribution of $\sqrt{t}\,(\Theta_t-\Theta_*)$ are equal to the corresponding asymptotic moments (Section 3.3). In the following sections, we *assume* the convergence of the second-order moments allowing us to write:

$$\mathrm{Cov}\Theta_t=\gamma C_\Theta+\mathrm{o}(\gamma). \tag{47}$$

Let $\Theta_t=\Theta_*+\delta\Theta_t$. Provided Θ_t is stationary, taking the expectation of both sides of Eqs. (25) and (26) gives

$$\mathbf{0}=\mathrm{E}(g^s(\Theta_*^s+\delta\Theta_t^s,y_t^s y_t^{s^{\mathrm T}})), \quad s=-,+. \tag{48}$$

As the field g^s is linear in its second argument and Θ_t^s and $y_t^s y_t^{s^{\mathrm T}}$ are independent (for independent observations x_t), the mean field at point $\Theta_*+\delta\Theta_t$ is

$$\mathrm{E}(g^s(\Theta_*^s+\delta\Theta_t^s,y_t^s y_t^{s^{\mathrm T}}))=\mathrm{E}(g^s(\Theta_*^s+\delta\Theta_t^s,R^s)),$$
$$s=-,+. \tag{49}$$

Then, $\mathrm{E}(g^s(\Theta_*^s+\delta\Theta_t^s,R^s))$, $s=-,+$ can be expanded in the neighborhood of Θ_*^s as

$$\mathrm{E}(g^s(\Theta_*^s+\delta\Theta_t^s,R^s)) = \mathbf{0}+\frac{\partial g^s}{\partial\Theta^s}_{\Theta_t^s=\Theta_*^s}\mathrm{E}(\delta\Theta_t^s) + \frac{1}{2}\frac{\partial^2 g^s}{\partial\Theta^{s^2}}_{\Theta_t^s=\Theta_*^s}\mathrm{E}(\mathrm{Vec}(\delta\Theta_t^s\delta\Theta_t^{s^{\mathrm T}})) + \mathrm{E}(O\|\delta\Theta_t^s\|^3), \tag{50}$$

in which, thanks to Eq. (32), the first-order term is given by

$$\begin{bmatrix} D_{u^s} & O \\ D_{u^s,\lambda^s} & -I_{n^s}\end{bmatrix}\begin{bmatrix}\mathrm{E}(\delta u_t^s)\\ \mathrm{E}(\delta\Lambda_t^s)\end{bmatrix}. \tag{51}$$

From Eq. (5) and from the expression of $\mathrm{E}(\delta u_{i,t}^s\delta u_{j,t}^s)$ deduced from Eqs. (43) and (47), we get after some algebric manipulations the second-order term:

$$\frac{1}{2}\frac{\partial^2 g^s}{\partial \Theta^{s^2}}_{\Theta^s=\Theta^s_*} \mathrm{E}(\mathrm{Vec}(\delta\Theta^s_t \delta\Theta^{s^{\mathrm{T}}}_t)) = \gamma \begin{bmatrix} -\alpha^s_1 \sum_{1<i\leqslant n^s} \frac{\alpha^s_1 \lambda^{s^2}_i \lambda^s_1}{2|\lambda^s_i - \lambda^s_1|} \\ -\alpha^s_2 \sum_{1\leqslant i\neq 2\leqslant n^s} \frac{\alpha^s_{\min(i,2)} \lambda^{s^2}_i \lambda^s_2}{2|\lambda^s_i - \lambda^s_2|} + \sum_{i=1}^{2-1} \frac{(\alpha^s_i + \alpha^s_2)\alpha^s_i \lambda^s_i \lambda^s_2}{2} \\ \vdots \\ -\alpha^s_{n^s} \sum_{1\leqslant i<n^s} \frac{\alpha^s_i \lambda^{s^2}_i \lambda^s_{n^s}}{2|\lambda^s_i - \lambda^s_{n^s}|} + \sum_{i=1}^{n^s-1} \frac{(\alpha^s_i + \alpha^s_{n^s})\alpha^s_i \lambda^s_i \lambda^s_{n^s}}{2} \\ \sum_{1<i\leqslant n^s} \frac{\alpha^s_1 \lambda^{s^2}_i \lambda^s_1}{2|\lambda^s_i - \lambda^s_1|} \\ \sum_{1\leqslant i\neq 2\leqslant n^s} \frac{\alpha^s_{\min(i,2)} \lambda^{s^2}_i \lambda^s_2}{2|\lambda^s_i - \lambda^s_2|} \\ \vdots \\ \sum_{1\leqslant i<n^s} \frac{\alpha^s_i \lambda^{s^2}_i \lambda^s_{n^s}}{2|\lambda^s_i - \lambda^s_{n^s}|} \end{bmatrix} + \mathrm{o}(\gamma). \quad (52)$$

Then, the resolution of the block triangular equation (50) in which $\boldsymbol{D}_{u^s}$ is a block triangular matrix too, is straightforward. We successively get

$$\mathrm{E}(\delta \boldsymbol{u}^s_{k,t}) = -\frac{\gamma}{4}\left(\sum_{1\leqslant j\neq k\leqslant n^s} \frac{\alpha^s_{\min(j,k)} \lambda^{s^2}_j}{|\lambda^s_j - \lambda^s_k|} - \sum_{j=1}^{k-1} \left(1 + \frac{\alpha^s_j}{\alpha^s_k}\right) \alpha^s_j \lambda^s_j \right) \boldsymbol{u}^s_k + \mathrm{o}(\gamma),$$

$$s = -, +, \quad (53)$$

$$\mathrm{E}(\delta\lambda^s_{1,t}) = \mathrm{o}(\gamma), \quad s = -, +, \quad (54)$$

$$\mathrm{E}(\delta\lambda^s_{k,t}) = \frac{\gamma}{2}\sum_{j=1}^{k-1} \left(1 + \frac{\alpha^s_j}{\alpha^s_k}\right) \alpha^s_j \lambda^s_j \lambda^s_k + \mathrm{o}(\gamma),$$

$$s = -, +, \ k = 2, \ldots, n^s. \quad (55)$$

So, thanks to the linear mapping $\boldsymbol{u}^s_k \to \boldsymbol{v}^s_k$ (6), Eq. (53) holds for the bias of $\boldsymbol{v}^s_{k,t}$ by replacing $\boldsymbol{u}^s_k$ by $\boldsymbol{v}^s_k$:

$$\mathrm{E}(\boldsymbol{v}^s_{k,t}) = \boldsymbol{v}^s_k - \frac{\gamma}{4}\left(\sum_{1\leqslant j\neq k\leqslant n^s} \frac{\alpha^s_{\min(j,k)} \lambda^{s^2}_j}{|\lambda^s_j - \lambda^s_k|} - \sum_{j=1}^{k-1} \left(1 + \frac{\alpha^s_j}{\alpha^s_k}\right) \alpha^s_j \lambda^s_j \right) \boldsymbol{v}^s_k + \mathrm{o}(\gamma),$$

$$s = -, +, \quad (56)$$

$$\mathrm{E}(\lambda^s_{1,t}) = \lambda^s_1 + \mathrm{o}(\gamma), \quad s = -, +, \quad (57)$$

$$\mathrm{E}(\lambda^s_{k,t}) = \lambda^s_k + \frac{\gamma}{2}\sum_{j=1}^{k-1} \left(1 + \frac{\alpha^s_j}{\alpha^s_k}\right) \alpha^s_j \lambda^s_j \lambda^s_k + \mathrm{o}(\gamma),$$

$$s = -, +, \quad k = 2, \ldots, n^s. \quad (58)$$

Asymptotic MSE. The MSE between $\boldsymbol{v}_t \stackrel{\text{def}}{=} \mathrm{Vec}(\boldsymbol{v}_{1,t}, \ldots, \boldsymbol{v}_{n,t})$ and $\boldsymbol{v} \stackrel{\text{def}}{=} \mathrm{Vec}(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n)$, and between $\Lambda_t \stackrel{\text{def}}{=} (\lambda_{1,t}, \ldots, \lambda_{n,t})^{\mathrm{T}}$ and $\Lambda \stackrel{\text{def}}{=} (\lambda_1, \ldots, \lambda_n)^{\mathrm{T}}$, is obtained from the asymptotic bias (56), (57), (58) and from the asymptotic covariance (47). And according to Theorem 3:

$$\mathrm{E}\|\Lambda_t - \Lambda\|^2_{\mathrm{Fro}} = \gamma \sum_{k=1}^{n} \lambda^2_k + \mathrm{o}(\gamma), \quad (59)$$

$$\mathrm{E}\|\boldsymbol{v}_t - \boldsymbol{v}\|^2_{\mathrm{Fro}} = \gamma\left(\sum_{1\leqslant j\neq k\leqslant \lfloor n/2 \rfloor} \frac{\alpha^-_{\min(j,k)} \lambda^-_j \lambda^-_k}{2|\lambda^-_j - \lambda^-_k|} + \sum_{1\leqslant j\neq k\leqslant \lceil n/2 \rceil} \frac{\alpha^+_{\min(j,k)} \lambda^+_j \lambda^+_k}{2|\lambda^+_j - \lambda^+_k|} \right) + \mathrm{o}(\gamma). \quad (60)$$

4.4. Analysis of the results

Let us now compare the results to the case where the CS structure is not taken into account [10], which we now recall for convenience of the reader:

$$\mathrm{E}(\boldsymbol{v}_{k,t}) = \boldsymbol{v}_k - \frac{\gamma}{4}\left(\sum_{1 \leqslant j \neq k \leqslant n} \frac{\alpha_{\min(j,k)}\lambda_j^2}{|\lambda_j - \lambda_k|} - \sum_{j=1}^{k-1} \left(1 + \frac{\alpha_j}{\alpha_k}\right)\alpha_j\lambda_j \right)\boldsymbol{v}_k + \mathrm{o}(\gamma), \quad (61)$$

$$\mathrm{E}(\lambda_{1,t}) = \lambda_1 + \mathrm{o}(\gamma), \quad (62)$$

$$\mathrm{E}(\lambda_{k,t}) = \lambda_k + \frac{\gamma}{2}\sum_{j=1}^{k-1} \left(1 + \frac{\alpha_j}{\alpha_k}\right)\alpha_j\lambda_j\lambda_k + \mathrm{o}(\gamma), \quad k = 2, \dots ,n, \quad (63)$$

$$\mathrm{E}\|\Lambda_t - \Lambda\|^2_{\mathrm{Fro}} = \gamma \sum_{k=1}^{n} \lambda_k^2 + \mathrm{o}(\gamma),$$
$$\mathrm{E}\|\boldsymbol{v}_t - \boldsymbol{v}\|^2_{\mathrm{Fro}} = \gamma\left(\sum_{1 \leqslant j \neq k \leqslant n} \frac{\alpha_{\min(j,k)}\lambda_j\lambda_k}{2|\lambda_j - \lambda_k|} \right) + \mathrm{o}(\gamma). \quad (64)$$

We note that if the asymptotic bias of the largest estimated eigenvalue and the asymptotic MSE of the estimated eigenvalues are unchanged, the asymptotic bias and the asymptotic MSE of the EVD are generally reduced when the CS structure is taken into account.

4.5. Comparison between batch and adaptive EVD estimators

It is worth noticing, from Theorems 2 and 3, that both estimators have very similar asymptotic distributions. With $\alpha_i = 1$, $i = 1, \dots ,n$, these distributions are equivalent if we substitute $2/t$ by γ and the differences $(\lambda_j - \lambda_k)^2$ by $|\lambda_j - \lambda_k|$. Furthermore, it is the same for the MSE of the estimated eigenvalues and eigenvectors. As for the bias, we note that $\lambda^-_{1,t}$, $\lambda^+_{1,t}\boldsymbol{v}^-_{1,t}\boldsymbol{v}^+_{1,t}$ have similar bias for both estimations, but $\lambda^s_{k,t}$, $s = -, +$, $k = 2, \dots ,n^s$ have a bias in $\mathrm{o}(1/t)$ for batch estimation and in $\mathrm{O}(\gamma)$ for adaptive estimation; lastly, the bias of $\boldsymbol{v}^s_{k,t}$, $s = -, +$, $k = 2, \dots ,n^s$ has an extra term in the case of adaptive estimation. However, in this latter case the bias is always directed along $\boldsymbol{v}^s_k$.

We note 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, $\boldsymbol{U}^s_k \overset{\mathrm{def}}{=} (\boldsymbol{u}^s_1, \dots ,\boldsymbol{u}^s_{n^s})$ derives from the successive constrained minimizations of $\mathrm{Tr}(\boldsymbol{U}^{s^{\mathrm{T}}}_k \boldsymbol{R}^s_t \boldsymbol{U}^s_k)$, $k = 1, \dots ,n^s$ with respect to $\boldsymbol{u}^s_k$ under the constraint that $\boldsymbol{U}^{s^{\mathrm{T}}}_k \boldsymbol{U}^s_k = \boldsymbol{I}_k$ in ML batch estimation. On the other hand, $\boldsymbol{U}^s_{n^s}$ derives from the minimization of $\mathrm{Tr}(\boldsymbol{U}^{s^{\mathrm{T}}}_{n^s}\boldsymbol{R}^s_t\boldsymbol{U}^s_{n^s})$, from a projected gradient-like procedure in SGA instantaneous adaptive estimation. The projection on the constraint $\boldsymbol{U}^{s^{\mathrm{T}}}_{n^s}\boldsymbol{U}^s_{n^s} = \boldsymbol{I}_{n^s}$ is realized thanks to an expansion of a Gram–Schmidt orthogonalization [10].

Concerning the deviation from orthonormality, we note that the ML batch estimation gives a canonic orthonormal estimated eigenbasis whereas the SGA adaptive estimation gives an approximately orthonormal estimated eigenbasis only. Since $\boldsymbol{K}^{s^{\mathrm{T}}}_e\boldsymbol{K}^s_e = \boldsymbol{I}$, $\boldsymbol{K}^{s^{\mathrm{T}}}_o\boldsymbol{K}^s_o = \boldsymbol{I}$, $\boldsymbol{K}^{s^{\mathrm{T}}}_e\boldsymbol{K}^{-s}_e = \boldsymbol{O}$ and $\boldsymbol{K}^{s^{\mathrm{T}}}_o\boldsymbol{K}^{-s}_o = \boldsymbol{O}$, it is straightforward to see from Eq. (6) that

$$\mathrm{E}\|\boldsymbol{V}^{\mathrm{T}}_{n,t}\boldsymbol{V}_{n,t} - \boldsymbol{I}_n\|^2_{\mathrm{Fro}} = \mathrm{E}\|\boldsymbol{U}^{-^{\mathrm{T}}}_{n^-,t}\boldsymbol{U}^-_{n^-,t} - \boldsymbol{I}_{n^-}\|^2_{\mathrm{Fro}} + \mathrm{E}\|\boldsymbol{U}^{+^{\mathrm{T}}}_{n^+,t}\boldsymbol{U}^+_{n^+,t} - \boldsymbol{I}_{n^+}\|^2_{\mathrm{Fro}}, \quad (65)$$

with $\boldsymbol{V}_{n,t} \overset{\mathrm{def}}{=} (\boldsymbol{v}_{1,t}, \dots ,\boldsymbol{v}_{n,t})$ and $\boldsymbol{U}^s_{n^s,t} \overset{\mathrm{def}}{=} (\boldsymbol{u}^s_{1,t}, \dots ,\boldsymbol{u}^s_{n^s,t})$. In [10], it is shown by simulation that these deviations from orthonormality $\mathrm{E}\|\boldsymbol{U}^{-^{\mathrm{T}}}_{n^-,t}\boldsymbol{U}^-_{n^-,t} - \boldsymbol{I}_{n^-}\|^2_{\mathrm{Fro}}$ and $\mathrm{E}\|\boldsymbol{U}^{+^{\mathrm{T}}}_{n^+,t}\boldsymbol{U}^+_{n^+,t} - \boldsymbol{I}_{n^+}\|^2_{\mathrm{Fro}}$ are both, up to the first order, proportional to γ^2. In this paper, $\mathrm{E}\|\boldsymbol{V}^{\mathrm{T}}_{n,t}\boldsymbol{V}_{n,t} - \boldsymbol{I}_n\|^2_{\mathrm{Fro}}$ is proportional to γ^2 as well in the domain of validity of the MSE (60).

5. Simulations

We consider throughout this section independent observations $\boldsymbol{x}_t$ in $\mathscr{R}^4$ associated to the symmetric Toeplitz matrix $\boldsymbol{R}_x = \mathrm{Toeplitz}(1, -0.3633, 0.0209, -0.0043)$ 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. $\boldsymbol{R}_x$ has the following eigenvalues:

$$\lambda_1 = \lambda_1^- = 1.6079, \quad \lambda_2 = \lambda_1^+ = 1.2028,$$
$$\lambda_3 = \lambda_2^- = 0.7597, \quad \lambda_4 = \lambda_2^+ = 0.4296.$$

The first experiment presents the case of ML batch estimation. Fig. 1 shows the eigenvalues

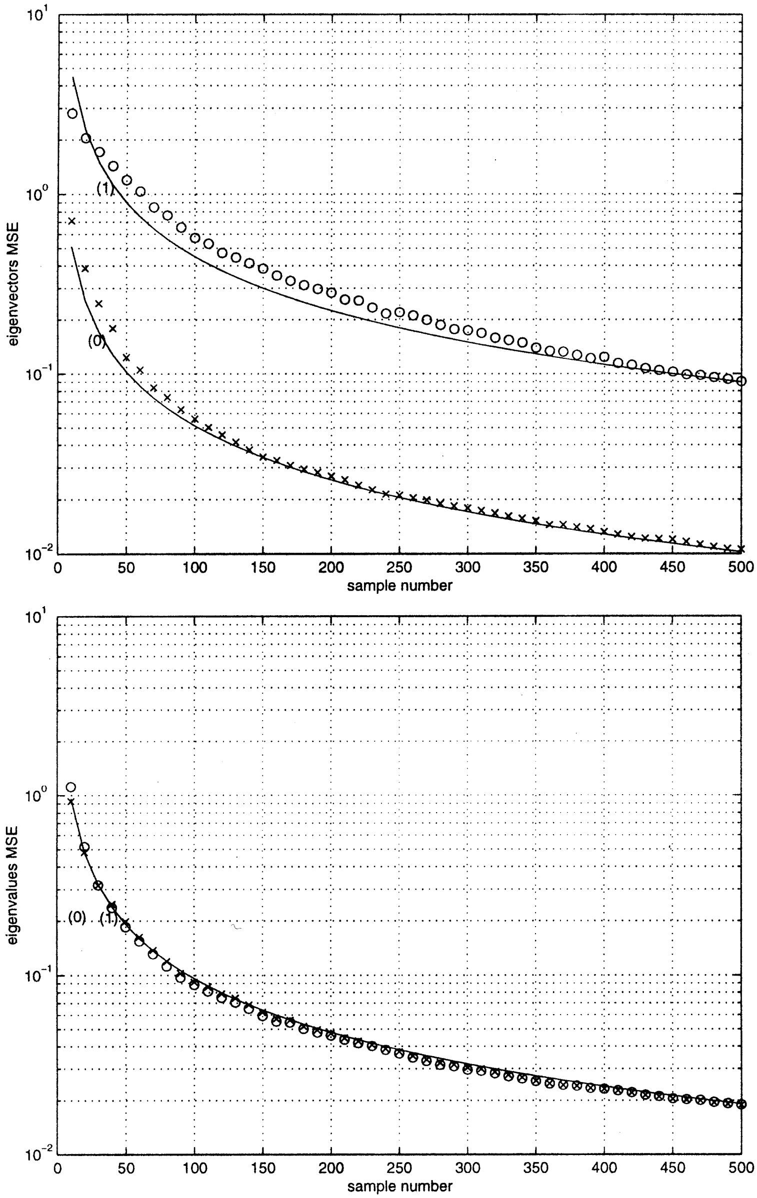


Fig. 1. MSE of the estimated eigenvectors and eigenvalues $\mathrm{E}\|\boldsymbol{v}_t - \boldsymbol{v}_*\|^2_{\mathrm{Fro}}$ and $\mathrm{E}\|\Lambda_t - \Lambda_*\|^2_{\mathrm{Fro}}$ averaging 100 independent runs for batch estimation when the CS structure is taken into account (×) or not (○) as a function of the sample number t, compared to the theoretical asymptotic values $1/t\,\mathrm{Tr}(\boldsymbol{C}_v)$ and $1/t\,\mathrm{Tr}(\boldsymbol{C}_\lambda)$ when the CS structure is taken into account (0) or not (1).

MSE and eigenvectors MSE (averaged over 100 independent runs), when the CS structure is taken into account or not, as a function of the sample number. We observe a reduction of the eigenvector MSE of 9 dB when the CS structure is taken into account, and these MSE tend to values in excellent agreement with the theoretical asymptotic values predicted by Eqs. (20), (21) and (24).

The second experiment presents the case of SGA adaptive estimation ($\alpha_i^s = 1$, $s = -, +$, $i = 1, \ldots, n^s$). Fig. 2 shows the learning curves (averaged over 200 independent runs) of the

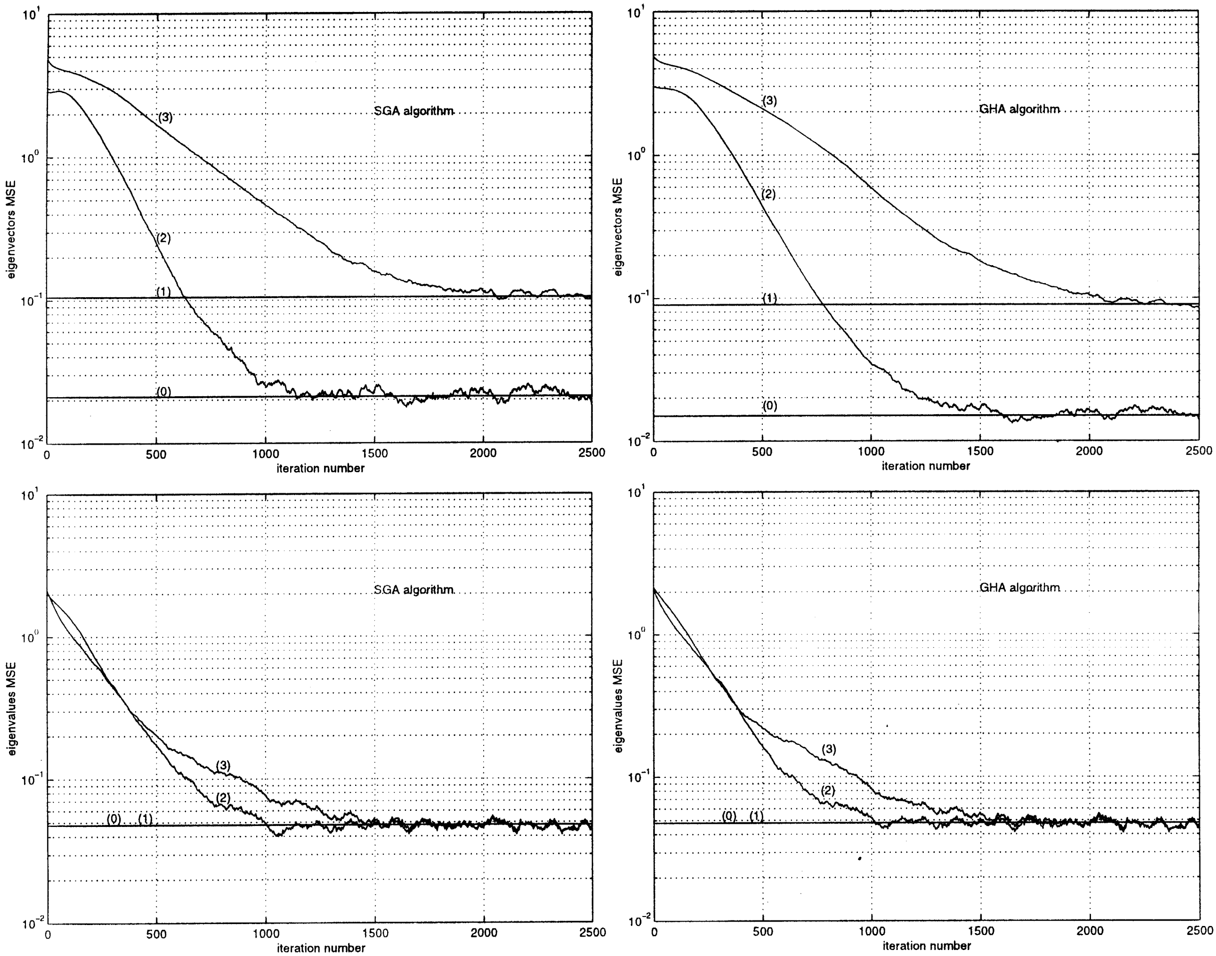


Fig. 2. Learning curves of the MSE $\mathrm{E}\|\boldsymbol{v}_t - \boldsymbol{v}_*\|^2_{\mathrm{Fro}}$ and $\mathrm{E}\|\Lambda_t - \Lambda_*\|^2_{\mathrm{Fro}}$ averaging 200 independent runs for respectively the SGA and the GHA algorithms when the CS structure is taken into account (2) or not (3), compared to the theoretical asymptotic values $\gamma\,\mathrm{Tr}(\boldsymbol{C}_v)$ and $\gamma\,\mathrm{Tr}(\boldsymbol{C}_\lambda)$ when the CS structure is taken into account (0) or not (1).

eigenvalue MSE, and eigenvector MSE when the CS structure is taken into account or not, with the common step size $\gamma = 0.01$. These MSEs tend to values in excellent agreement with the theoretical values predicted by Eqs. (59), (60) and (64). 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. We note that the same analysis carried out for the GHA algorithm gives the same conclusion.

Fig. 3 shows the theoretical asymptotic and the estimated eigenvalue and eigenvector MSEs as a function of γ. Our present asymptotic analysis is seen to be valid over a large range of γ ($\gamma < 0.03$), and the domain of "stability" is $\gamma < 0.07$, for which we observe good agreement between the theoretical and estimated MSEs. This result supports our conjecture that the asymptotic covariance matrice of our adaptive EVD estimator is identical to the covariance matrix in the limiting distribution.

Fig. 4 reveals something which could not be determined from our first-order analysis: the true order of deviation from orthonormality. In this figure, we plot on a log-log scale $\mathrm{E}\|\boldsymbol{V}_t^{\mathrm{T}}\boldsymbol{V}_t - \boldsymbol{I}_n\|^2_{\mathrm{Fro}}$ as a function of γ. We find a slope equal to 2

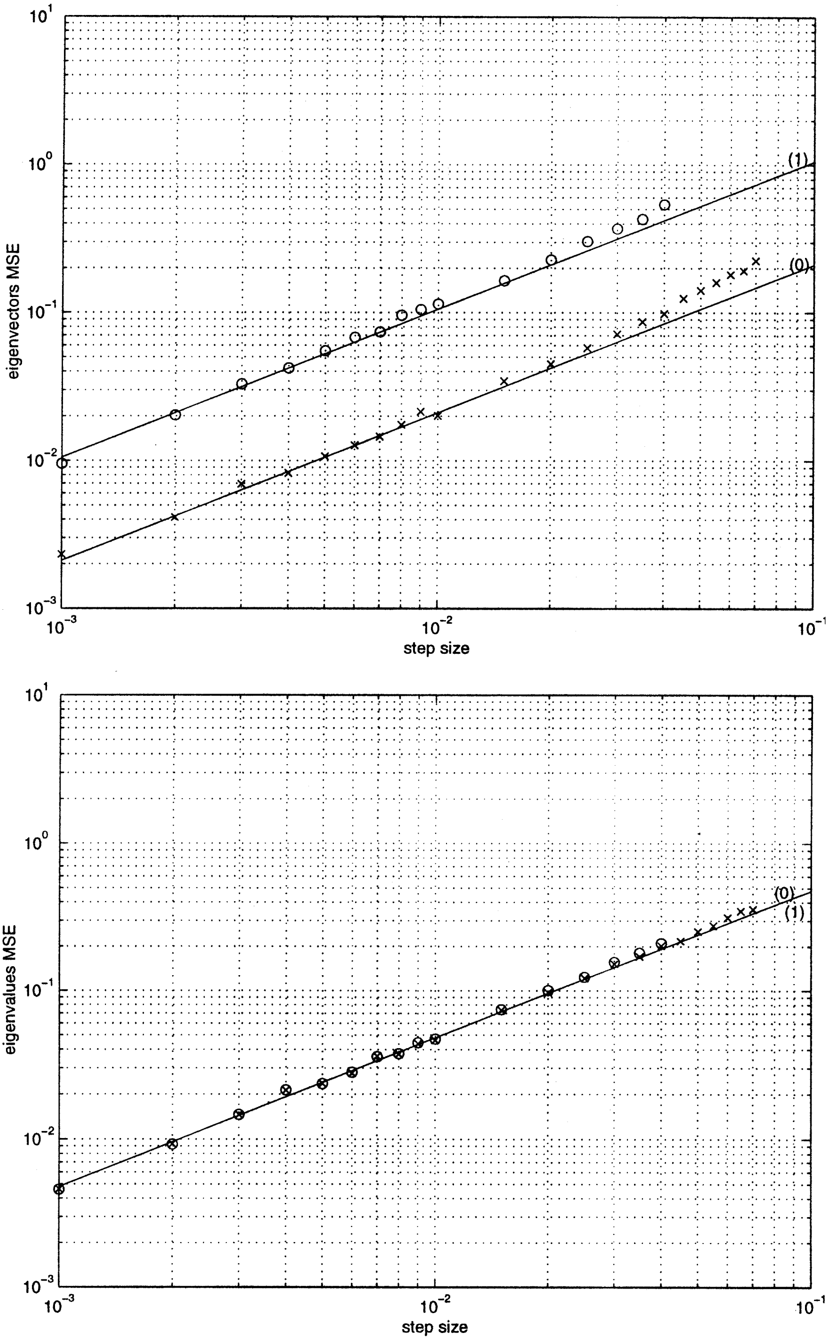
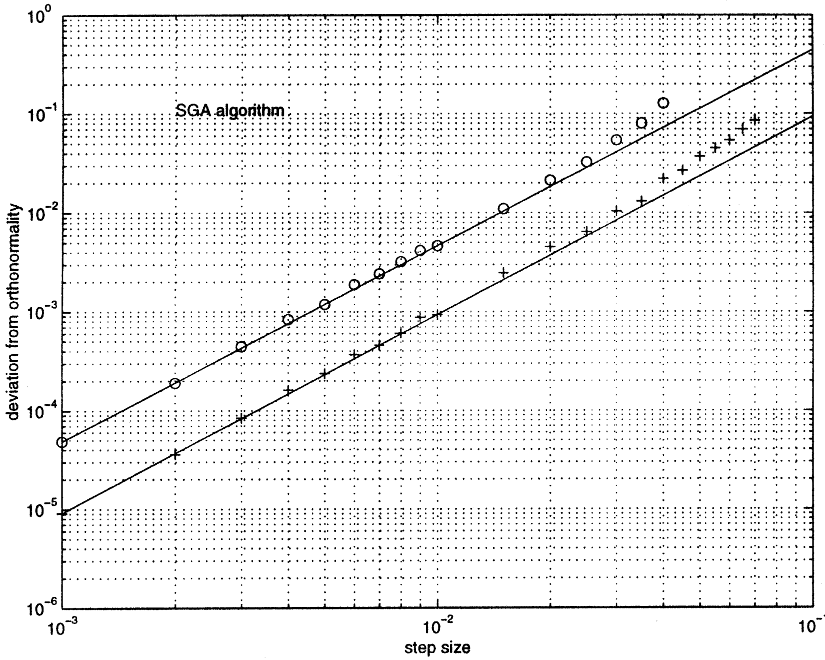


Fig. 3. Estimated (respectively, theoretical asymptotic) eigenvectors and eigenvalues MSE as a function of the step size γ when the CS structure is taken into account ($\times$) (respectively, (0)) or not ($\bigcirc$) (respectively, (1)).



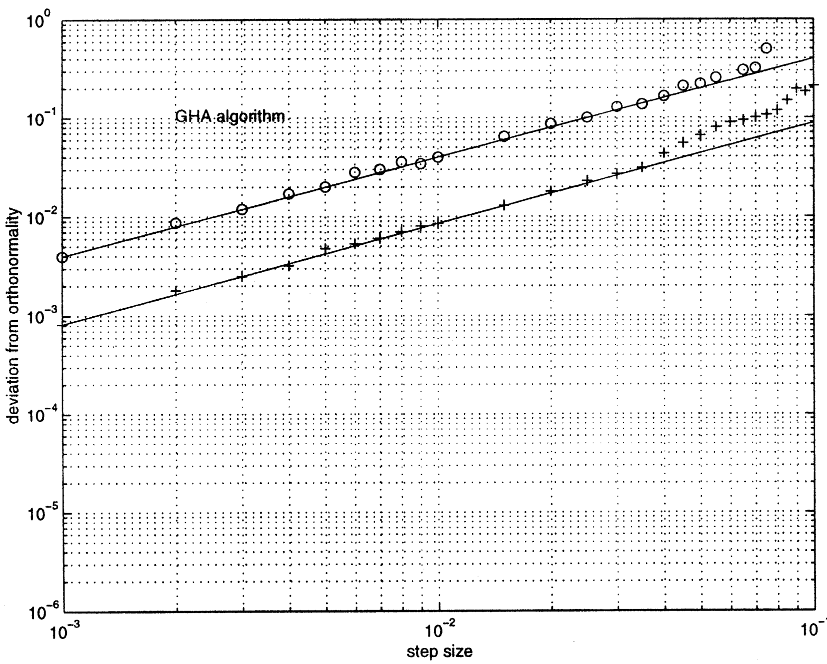


Fig. 4. Deviation from orthonormality $\mathrm{E}\|V_t^T V_t - I_n\|_{\mathrm{Fro}}^2$ at "convergence" estimated by averaging 100 independent runs as a function of γ in log–log scales for the SGA and GHA algorithms when the CS structure is taken into account (+) or not (○).

[respectively 1] which means that, experimentally, $\mathrm{E}\|\boldsymbol{V}_t^{\mathrm{T}}\boldsymbol{V}_t-\boldsymbol{I}_n\|^2_{\mathrm{Fro}}\propto\gamma^2$ [respectively $\propto\gamma$] for the SGA [respectively GHA] algorithm. And furthermore, this deviation from orthonormality is reduced when the CS strucuture is taken into account

6. Conclusion

In this paper, we have presented a performance analysis of ML batch estimation and adaptive estimation of the EVD of CS covariance matrices. When the CS structure is taken into account, it is shown that the EVD estimation can be split into two independent EVD estimations. The asymptotic distributions are derived, and closed-form expressions are given for the asymptotic covariance, bias and MSE for ML batch estimation and SGA adaptive estimation. As a result of the exploitation of the CS structure, it is shown that the complexity of the EVD is roughly halved, and that the covariance, bias and MSE are reduced. Of course, these results extend to any gradient-type or RLS-type algorithm since all these EVD algorithms can be split into two independent EVD algorithms, the conditioning of which is improved because the difference between two successive eigenvalues increases in general. Finally, numerical simulations confirm the accuracy of our asymptotic analysis and show that for the adaptive estimation, the deviation from orthonormality is reduced and the convergence speed is improved yielding a better tradeoff between convergence speed and misadjustment.

References

[1] T.W. Anderson, An Introduction to Multivariate Statistical Analysis, Second Edition, Wiley and Sons, 1984.

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

[3] D.R. Brillinger, Times Series, Data Analysis and Theory, Expanded Edition, Holden-Day, Inc., 1980.

[4] A. Bunse-Gerstner, R. Byers, V. Mehrmann, A chart of numerical methods for structured eigenvalue problems, SIAM J. Matrix Anal. Appl. 13 (2) (April 1992) 419–453.

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

[6] L. Datta, S.D. Morgera, On the reducibility of centrosymmetric matrices: Applications in engineering problems, Circuits Systems Signal Process. 8 (1) (1989) 71–96.

[7] N. Delfosse, P. Loubaton, Adaptive blind separation of independent sources: A deflation approach, Signal Processing 45 (1995) 59–83.

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

[9] J.-P. Delmas, Performance analysis of a Givens parametrized adaptive eigenspace algorithms, Signal Processing 68 (1998) 87–105.

[10] J.-P. Delmas, F. Alberge, Asymptotic performance analysis of subspace adaptive algorithms introduced in the neural network literature, IEEE Trans. Signal Process. 46 (1) (January 1988) 170–182.

[11] P. Delsarte, Y, Genin, Spectral properties of finite Toeplitz matrices, in: Proc. Internat. Symp. on the Math. Theory of Networks and Systems, Vol. 58, Springer, New York, 1984, pp. 194–213.

[12] G.H. Golub, C.F. Van Loan, Matrix Computations, Johns Hopkins University Press, Baltimore, MD, 1990.

[13] K. Hornik, C.M. Kuan, Convergence analysis of local feature extraction algorithms, Neural Networks 5 (1992) 229–240.

[14] M. Kaveh, A.J. Barabell, The statistical performance of the MUSIC and the Minimum-Norm algorithms in resolving plane waves in noise, IEEE Trans. ASSP 34 (2) (April 1986) 331–341.

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

[16] E. Oja, A simplified neuron model as a principal components analyzer, Journal of Math. Biology 15 (1982) 267–273.

[17] J. Oja, Subspace Methods of Pattern Recognition, Research Studies Press and John Wiley and Sons, Letchworth, England, 1983.

[18] 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.

[19] F. Vanpoucke, S.H. Jensen, M. Moonen, A Persymmetric adaptive eigenfilter-bank for real data, in: Proc. of Asilomar-29, October 1995, pp. 1342–1346.

[20] G. Xu, Y. Cho, T. Kailath, Application of fast subspace decomposition to signal processing and communication problems, IEEE Trans. on Signal Process. 42 (6) (June 1994) 1543–1461.

[21] 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.

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