Performance Analysis of an Adaptive Algorithm for Tracking Dominant Subspaces

Jean Pierre Delmas and Jean François Cardoso

Abstract—This paper provides a performance analysis of a least mean square (LMS) dominant invariant subspace algorithm. Based on an unconstrained minimization problem, this algorithm is a stochastic gradient algorithm driving the columns of a matrix \( W \) to an orthonormal basis of a dominant invariant subspace of a correlation matrix. We consider the stochastic algorithm governing the evolution of \( WW^H \) to the projection matrix onto this dominant invariant subspace and study its asymptotic distribution. A closed-form expression of its asymptotic covariance is given in the case of independent observations and is further analyzed to provide some insights into the behavior of this LMS type algorithm. In particular, it is shown that even though the algorithm does not constrain \( W \) to have orthonormal columns, there is deviation from orthonormality of the first order. We also give a closed-form expression of the asymptotic covariance of DOA’s estimated by the MUSIC algorithm applied to the adaptive estimate of the projector. It is found that the asymptotic distributions have a structure that is very similar to those describing batch estimation techniques because both algorithms are obtained from the minimization of the same criterion. Finally, the accuracy of the asymptotic analysis is checked by numerical simulations and is found to be valid not only for a “small” step size but in a very large domain.

Index Terms—Adaptive/batch MUSIC, asymptotic performance analysis, subspace tracking.

I. INTRODUCTION

Subspace Tracking

O

VER THE PAST decade, adaptive estimation of subspaces of covariance matrices has been applied successfully to high-resolution spectral analysis in signal processing and principal component analysis in data compression and feature extraction. The interest for these methods (a tool of growing importance in many fields of signal processing) has recently been renewed by the subspace approach used in blind identification of multichannel FIR filters [1]. Numerous solutions have been proposed to recursively updating subspaces of covariance matrices (see, for example, the references in [2] and [3]), but there are relatively few performance analyses concerning stochastic gradient algorithms derived from constrained or unconstrained optimization problems. Among them, Larimore and Calvert [4] presented a convergence study of the Thompson algorithm, whereas Yang and Kaveh [5] made an analysis of convergence rate and stability of their constrained gradient search procedure resorting to the classical independence assumption. Studies of convergence with the help of the associated ordinary differential equation (ODE) were carried out by many authors from Oja and Karhunen [6]. Evaluation of the performance by providing the asymptotic distributions of the estimated eigenvectors were proposed in [7]–[9]. It is the purpose of this paper to provide a thorough study of the behavior of an LMS-type approximation algorithm presented by Yang in [2].

Algorithm Under Study

For a given \( n \times n \) covariance matrix \( R \), denote as \( \lambda_1 \geq \cdots \geq \lambda_n \) the eigenvalues of \( R \) and \( v_1, \ldots, v_n \) corresponding eigenvectors. The \( p \)-dimensional dominant invariant subspace of \( R \) is the span of \( v_1, \ldots, v_p \), and it is well defined if, as assumed throughout the paper, \( \lambda_p > \lambda_{p+1} \). Denote \( \Pi \) as the orthogonal projector onto this subspace. We then have

\[
Rv_\alpha = \lambda_\alpha v_\alpha \quad \Pi v_\alpha = \pi_\alpha v_\alpha \quad 1 \leq \alpha \leq n
\]

where we have defined \( \pi_1 = \cdots = \pi_p = 1 \), and \( \pi_{p+1} = \cdots = \pi_n = 0 \). Defining the rank-one projection matrices \( \Pi_\alpha \) onto each eigenvector, we can also write

\[
\Pi_\alpha \overset{\text{def}}{=} v_\alpha v_\alpha^H \quad R = \sum_{\alpha=1}^{n} \lambda_\alpha \Pi_\alpha
\]

\[
\Pi = \sum_{\alpha=1}^{n} \pi_\alpha \Pi_\alpha = \sum_{1 \leq \alpha \leq p} \Pi_\alpha.
\]

Subspace tracking consists of recursively updating, at time \( t \), an (approximately) orthonormal basis of this subspace on reception of sample \( x_t \) of a stationary process with covariance \( R = E x_t x_t^H \), where \( x_t \) is supposed to be a zero-mean complex circular Gaussian random vector.

There are several interesting algorithms described in Yang’s paper [2] based on the unconstrained minimization of the objective function

\[
J_0(W) \overset{\text{def}}{=} E\|x_t - WW^H x_t\|^2_{\infty}
\]

with respect to the \( n \times \tau \) matrix \( W \). In this contribution, we consider the stochastic gradient algorithm for the minimization

\[
1\text{Complex-valued quantities are implicitly assumed here. Most of the paper}
\]

\[
\text{deals with the complex case, but the real case is also considered in Section III-A.}
\]
of $J_b(W)$. This yields the following algorithm, where $\gamma_t > 0$ is a step size:

$$W_{t+1} = W_t + \gamma_t h(W_t, x_t)$$

(4)

$$h(W_t, x_t) \triangleq (2x_t x_t^H - x_t x_t^H W_t W_t^H - W_t W_t^H x_t x_t^H)W_t.$$  

(5)

Baldi and Hornik [10] have shown (in the real case) that the stationary points of $J_b(W)$ are $W = \{WW = V_r U\}$, where $V_r$ is any $r$-dimensional eigenvector basis of $R$, and $U$ is an $r \times r$ arbitrary orthogonal matrix. All these points are saddle points except when $V_r = V_S \triangleq [v_1, \ldots, v_r]$. In this case, $J_b(W)$ attains the global minimum. These results are restated in [2] in the complex case, and subsequent work in this subject has been carried out by Xu [11]. In stationary situations and if the step size $\gamma_t$ satisfies the conditions $\sum_{t=1}^{\infty} \gamma_t = +\infty$ and $\lim_{t \to \infty} \gamma_t = 0$, the study of the convergence of algorithm (4) is intimately connected to the associated ordinary differential equation (ODE)

$$\frac{dW_t}{dt} = E(h(W_t, x)).$$

(6)

Therefore, we can conjecture that the stochastic algorithm (4) converges almost surely to $W_{\ast S} = \{WW = V_S U\}$.

**Projector Tracking:** A difficulty arises in the study of the behavior of $W_t$ because the set $W_S$ forms a continuum of attractors; the column vectors of $W_t$ do not, in general, tend to the eigenvectors $v_1, \ldots, v_r$, and we have no proof of convergence of $W_t$ to a particular orthonormal basis of their span. Therefore, the approach followed in this paper is to study the trajectory of matrix $P_t$

$$P_t \triangleq W_t W_t^H$$

(7)

whose dynamic is governed by the stochastic equation

$$P_{t+1} = P_t + \gamma_t f(P_t, x_t x_t^H) + \gamma_t^2 g(P_t, x_t x_t^H)$$

(8)

$$f(P, M) \triangleq P(2M - MP - PM)$$

(9)

$$g(P, M) \triangleq (2M - MP - PM)P(2M - MP - PM)$$

(10)

obtained by combining (4), (5), and (7). In the following, we are interested in first-order asymptotic effects. We derive the asymptotic variance of $P_t$, around $P_{\ast S}$ for a small fixed value $\gamma$ of the step size. At first order in $\gamma$, this covariance is proportional to $\gamma$ and is not affected by the $O(\gamma^2)$ term in (8); therefore, this term can be neglected. A remarkable feature of (8) is that the field $f$ actually depends only on $P_t$ and not on $W_t$. This fortunate circumstance makes it possible to study the evolution of $P_t$ without determining the evolution of the underlying matrix $W_t$. The characteristics of $P_t$ are indeed the most interesting since they completely characterize the estimated subspace.

**Outline of the Paper:** This paper is organized as follows. In Section II, after presenting a brief review of a general Gaussian approximation result, we obtain in closed form the asymptotic covariance of $P_t$ for the case where $x_t$ is a white complex circular Gaussian sequence. Several lemmas are included, the proofs of which are reported at the end of the paper. In Section III, we extend this result to real signals, and we compare the asymptotic performance of the algorithm with the performance of the PAST algorithm [2] and of batch algorithm for subspace estimation. We also investigate the performance of direction-of-arrival (DOA) estimation based on the adaptive subspace algorithm. Section IV presents some simulation results and investigates the validity of the asymptotic approach.

**II. ASYMPTOTIC PERFORMANCE ANALYSIS**

**A. A Short Review of a General Gaussian Approximation Result**

In this section, we evaluate the asymptotic performance of algorithm (4). For this purpose, we will use the following result [12, th. 2, p. 108)]. Let

$$\theta_{t+1} = \theta_t + \gamma_t h(\theta_t, x_t)$$

(11)

be a constant step-size recursive stochastic algorithm. Suppose that the $n \times 1$ real parameter $\theta_t$ converges almost surely to the asymptotically stable point $\theta_\ast$ in the corresponding decreasing step-size algorithm. Then, in a real stationary situation, we have when $\gamma \to 0$ and $t \to \infty$ (where $\to \mathcal{L}$ denotes the convergence in distribution)$2$

$$\frac{1}{\sqrt{\gamma}}(\theta_t - \theta_\ast) \to \mathcal{L} N(0, C_\theta)$$

(12)

where $C_\theta$ is the unique symmetric solution of the continuous Lyapunov equation

$$DC_\theta + C_\theta D^T + \Gamma = 0$$

(13)

and where $D$ and $\Gamma$ are, respectively, the derivative of the mean field and the covariance of the field

$$D \triangleq E \left[ \frac{\partial \phi}{\partial \theta} (\theta, x) \right]_{\theta = \theta_\ast} \quad \left[ D \right]_{i,j} \triangleq \frac{\partial^2 \phi}{\partial \theta_i \partial \theta_j}$$

(14)

$$\Gamma \triangleq \sum_{t=-\infty}^{\infty} \text{cov} \left[ \phi(\theta_t, x_t), \phi(\theta_t, x_t) \right]$$

(15)

Thus, $\theta_t$ behaves asymptotically, for $t$ “large enough” and $\gamma$ “small enough,” like an unbiased Gaussian estimator of $\theta_\ast$ with covariance matrix $C_\theta$. In cases where a closed form for the EVD of $D^T$ is available ($D^T m = \mu_i m_i$ for $1 \leq i \leq n$), we have

$$m_i^T C_\theta m_j = \frac{m_i^T \Gamma m_j}{\mu_i + \mu_j}$$

(16)

Further, if $D$ is symmetric, the eigenvectors $m_1, \ldots, m_n$ can be chosen orthonormal, and (16) yields

$$C_\theta = - \sum_{1 \leq i,j \leq n} \frac{m_i m_i^T \Gamma m_j m_j^T}{\mu_i + \mu_j^T}$$

(17)

---

2 We note that in the particular case, $r = 1$ and $x_t$ real, (6) admits only two asymptotically stable points. Under some additional technical assumptions, a result of [14] asserts that each solution of (6) converges to one of the points $\{-x_1, +x_1\}$, and consequently, $W_t$ converges almost surely to one of these points.
Further, if $\mathbf{I}$ and $\mathbf{D}$ share the same set of eigenvectors, i.e., if we have $\mathbf{I} \mathbf{m}_i = \mu_i \mathbf{m}_i$ for $1 \leq i \leq n$, the above reduces to
\[
C_\theta = \sum_{i=1}^{n} \frac{1}{2\mu_i} \mathbf{m}_i \mathbf{m}_i^T. \tag{18}
\]

The main objective of this paper is to give an explicit form of the asymptotic covariance matrix for the algorithm (4). This requires solving the Lyapunov equation (13) in closed form. This turns out to be analytically tractable because we shall exhibit a basis where matrices $\mathbf{D}$ and $\mathbf{I}$ are both diagonal; therefore $C_\theta$ will become available in closed form via (18).

B. A Short Review of an Asymptotic Property of M Estimates

In this subsection, we recall the asymptotic distribution of M estimates [13, Th. 1, p. 312], which we shall use to evaluate the asymptotic distribution in the case of batch estimation. If $\mathbf{x}_t$ is a sequence of independent identically distributed zero mean random vectors and $\theta_T$ is an isolated solution of
\[
\mathbf{I}_T \mathbf{x}_t = \mathbf{0}, \quad \mathbf{x}_t \sim \mathcal{N}(0, \mathbf{C}_\theta)
\]
with $E(\phi(\theta_T, \mathbf{x}_t)) = 0$, and if $\phi$ is sufficiently smooth, the following convergence in distribution result holds when $T$ tends to $\infty$:
\[
\sqrt{T}(\theta_T - \theta_\star) \rightarrow \mathcal{L} \mathcal{N}(0, \mathbf{C}_\theta) \tag{20}
\]
where $\mathbf{C}_\theta = \mathbf{D}^{-T} \mathbf{I} \mathbf{D}^{-1} \mathbf{D}$ is defined in (14), and $\mathbf{I} = \text{cov} [\phi(\theta_T, \mathbf{x}_t), \phi(\theta_\star, \mathbf{x}_t)]$. Further, if $\mathbf{I}$ and $\mathbf{D}$ (symmetric) share the same set of eigenvectors, i.e., if we have $\mathbf{I} \mathbf{m}_i = \mu_i \mathbf{m}_i$ and $\mathbf{D} \mathbf{m}_i = \mu_i \mathbf{m}_i$ for $1 \leq i \leq n$, the matrix $\mathbf{C}_\theta$ reduces to
\[
\mathbf{C}_\theta = \sum_{i=1}^{n} \frac{1}{2\mu_i} \mathbf{m}_i \mathbf{m}_i^T. \tag{21}
\]

C. Local Characterization of the Field

According to previous section, we need to characterize two local properties of the field $f(\mathbf{P}, \mathbf{x}\mathbf{x}^H)$: the mean value of its derivative and its covariance, both evaluated at point $\mathbf{P} = \mathbf{I}$. To proceed, it will be convenient to define the following set of orthonormal Hermitian matrices [the inner product under consideration is $(\mathbf{A}, \mathbf{B}) \triangleq \text{Tr} (\mathbf{A}^H \mathbf{B})$]:
\[
\mathbf{H}_{ab} = \begin{cases} 
\frac{\mathbf{v}_a \mathbf{v}_b^H + \mathbf{v}_b \mathbf{v}_a^H}{\sqrt{2}} & a = b \\
\frac{\mathbf{v}_a \mathbf{v}_b^H - \mathbf{v}_b \mathbf{v}_a^H}{i\sqrt{2}} & a > b
\end{cases} \tag{22}
\]
With this definition, a first-order approximation of the mean field in the neighborhood of $\mathbf{I}$ is given by the following lemma.

Lemma 1: For $1 \leq a, b \leq n$
\[
Ef(\mathbf{I} + \epsilon \mathbf{H}_{ab}, \mathbf{x}\mathbf{x}^H) = \epsilon \mu_{ab} \mathbf{H}_{ab} + O(\epsilon^2) \tag{23}
\]
with
\[
\mu_{ab} \triangleq 2\lambda_a (1 - \pi_a) + 2\lambda_b (1 - \pi_b) - (\lambda_a + \lambda_b) (\pi_a + \pi_b). \tag{24}
\]

To characterize the covariance of the field at point $\mathbf{P}$, we use the Vec operator, which turns a $p \times q$ matrix into a $pq \times 1$ vector by stacking successive columns one below another. It will be used in conjunction with the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ as the block matrix whose $(i, j)$ block element is $b_{ij} \otimes \mathbf{A}$. We have the classic properties (see, for example, in the real case [16, Appendix A])
\[
\text{Vec}(\mathbf{A} \mathbf{B}) = (\mathbf{A} \otimes \mathbf{B}) \text{Vec} (\mathbf{B}) \tag{25}
\]
\[
(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{A} \mathbf{C}) \otimes (\mathbf{B} \mathbf{D}) \tag{26}
\]
\[
(\mathbf{A} \otimes \mathbf{B})^H = (\mathbf{A}^H \otimes \mathbf{B}^H). \tag{27}
\]
Define, then, the following two matrices:
\[
\mathbf{I}^{-1} \triangleq \mathbf{I}_n - \mathbf{I} \quad \mathbf{Q} \triangleq \mathbf{I}^{-1} \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{I}^{-1}. \tag{28}
\]
The covariance of the field at point $\mathbf{P} = \mathbf{I}$ is given by the following lemma.

Lemma 2: For a circular complex vector $\mathbf{x}$
\[
\text{Cov} (\text{Vec}(f(\mathbf{I}, \mathbf{x}\mathbf{x}^H))) = \mathbf{Q} \text{Vec}(\mathbf{R}) \text{Vec}(\mathbf{R})^H. \tag{29}
\]

The eigenstructure of this covariance matrix is characterized as follows.

Lemma 3: For $1 \leq a, b \leq n$
\[
\text{cov}(\text{Vec}(f(\mathbf{I}, \mathbf{x}\mathbf{x}^H))) \text{Vec}(\mathbf{H}_{ab}) = \nu_{ab} \text{Vec}(\mathbf{H}_{ab}) \tag{30}
\]
with
\[
\nu_{ab} \triangleq (\pi_a - \pi_b)^2 \lambda_a \lambda_b, \tag{31}
\]

D. Real Parameterization

The Lyapunov equation (13) deals with a vector $\theta$ of real parameters. Herein, the parameter $\mathbf{P}$ of interest is an $n \times n$ rank-$r$ complex Hermitian matrix. To adapt the results recalled in Section II-A to our needs, matrix $\mathbf{P}$ should be parameterized by a vector $\theta$ of real parameters. Counting degrees of freedom shows that the set of $n \times n$ rank-$r$ complex Hermitian matrices is a $r(2n - r)$-dimensional manifold. This section introduces a parameterization of this manifold in a neighborhood of $\mathbf{I}$ by a $r(2n - r) \times 1$ vector $\theta$ of real parameters.

For an $n \times n$ matrix $\mathbf{M}$ and any pair $1 \leq a, b \leq n$ of indices, define
\[
\theta_{ab}(\mathbf{M}) \triangleq \text{Tr} \{\mathbf{H}_{ab}(\mathbf{M} - \mathbf{I})\}. \tag{31}
\]
These are real scalars if $\mathbf{M}$ is Hermitian. Since $[\mathbf{H}_{ab}]_{1 \leq a, b \leq n}$ is an orthonormal basis for the linear space of $n \times n$ matrices, the scalars $\theta_{ab}(\mathbf{M})$ are the coordinates of $\mathbf{M} - \mathbf{I}$

3 This slightly unusual convention makes it easier to deal with complex matrices.
on this basis. Thus, any $n \times n$ matrix is parameterized by the values of $\theta_{ab}(M)$ according to
\[ M = \Pi + \sum_{1 \leq a, b \leq n} \theta_{ab}(M)H_{ab}. \tag{32} \]
Matrices close to $\Pi$ are parameterized by small values of these parameters. By definition, $\theta_{ab}(M) = O(||M - \Pi||)$ for any pair $(a, b)$. The relevance of these parameters is shown by this lemma.

**Lemma 4:** If $P$ is an $n \times n$ rank-$r$ Hermitian matrix, then
\[ P = \Pi + \sum_{(a, b) \in P_h} \theta_{ab}(P)H_{ab} + O(||P - \Pi||^2) \tag{33} \]
where $P_h$ is the complement of $\{(a,b) | r < a, b \leq n \}$, i.e., $P_h \overset{\text{def}}{=} \{(a,b) | 1 \leq a \leq r, \text{ or } 1 \leq b \leq r \}$.

In other words, a rank-$r$ Hermitian matrix lying less than $\epsilon$ away from $\Pi$ (i.e., $||P - \Pi|| < \epsilon$) has negligible (of order $\epsilon^2$) components in the direction of $H_{ab}$ for $r < a, b \leq n$. Equation (33) is more compactly expressed by using an $n^2 \times 2n^2 - r^2$ matrix $\mathcal{H}$
\[ \mathcal{H} \overset{\text{def}}{=} \ldots, \text{Vec}(H_{ab}), \ldots, \quad (a,b) \in P_h \tag{34} \]
so that the $r(2n - r) \times 1$ vector $\theta(P)$ defined by
\[ \theta(P) \overset{\text{def}}{=} \mathcal{H}^H \text{Vec}(P - \Pi) \tag{35} \]
contains the values of $\theta_{ab}(P)$ for $(a,b) \in P_h$, and (33) reads, after vectorization
\[ \text{Vec}(P) = \text{Vec}(\Pi) + \mathcal{H} \theta(P) + O(||P - \Pi||^2). \tag{36} \]

Note that the particular ordering of the pairs in the set $P_h$ is irrelevant in expressions like $\mathcal{H} \theta$. There are $n^2 - (n - r)^2 = r(2n - r)$ pairs in $P_h$, and this is exactly the dimension of the manifold of $n \times n$ rank-$r$ Hermitian matrices. This point, together with (33), shows that the matrix set $\{H_{ab}, (a,b) \in P_h\}$ in fact is a basis of the tangent plane to this manifold at point $\Pi$. It follows that in a neighborhood of $\Pi$, Hermitian rank-$r$ matrices are uniquely determined by the (real) values of $\theta$. This is the required real (local) reparameterization of a rank-$r$ Hermitian matrix by a $r(2n - r)$-dimensional vector $\theta$. We denote as $\mathcal{P}(\theta)$ the unique (for small enough $||\theta||$) Hermitian matrix with rank $r$ such that $\mathcal{H}^H \text{Vec}(\mathcal{P}(\theta) - \Pi) = \theta$. It is not necessary to express explicitly $\mathcal{P}(\theta)$. As it will occur, it is sufficient to use the property derived from (36)
\[ \text{Vec}(\mathcal{P}(\theta)) = \text{Vec}(\Pi) + \mathcal{H} \theta + O(||\theta||^2). \tag{37} \]

**E. Solution of the Lyapunov Equation**

We are now in position to solve the Lyapunov equation in the new parameter $\theta$ defined in the previous subsection. The stochastic equation governing the evolution of this vector parameter is obtained by applying the transformation $\mathbf{P}_t \rightarrow \theta_t = \mathcal{H}^H \text{vec}(\mathbf{P}_t - \Pi)$ to the original (8).
\[ \theta_{t+1} = \theta_t + \gamma_t \phi(\theta_t, \mathbf{x}_t) + O(\gamma_t^2) \tag{38} \]
where function $\phi$ appears to be
\[ \phi(\theta, \mathbf{x}) \overset{\text{def}}{=} \mathcal{H}^H \text{vec}(f(\mathcal{P}(\theta), \mathbf{x} x^H)). \tag{39} \]
We need to evaluate the derivative matrix $D$ of $E \phi(\theta, \mathbf{x})$ at point $\theta = 0$ since we first consider only the case of independent observations, the covariance matrix $\Gamma$ of $\phi(0, \mathbf{x})$.

With these notations, the results of Section II-C are recycled as
\[ E \phi(\theta, \mathbf{x}) = \mathcal{H}^H \text{vec}(f(\mathcal{P}(\theta), \mathbf{x} x^H)) \]
\[ = \mathcal{H}^H \text{vec}(f(\Pi, \mathbf{x} x^H)) \]
\[ = \mathcal{H}^H \text{vec}(\sum \theta_{ab} H_{ab}) + O(||\theta||^2) \]
\[ = \mathcal{H}^H \text{vec}(\sum \theta_{ab} H_{ab}) + O(||\theta||^2) \]
where the above summations are over $(a,b) \in P_h$. The first equality uses definition (39) and the linearity of the $\mathcal{H}$ operation, the second equality stems from property (36) of the reparameterization, the third equality uses Lemma 1 and the differentiability of $f$, the fourth equality is by definitions (34) and (40), and the last equality is due to the orthonormality of the basis $\{H_{ab}\}$ and allows us to conclude that
\[ D \overset{\text{def}}{=} \left. \frac{\partial E \phi(\theta, \mathbf{x})}{\partial \theta} \right|_{\theta=0} = \Delta_\mu \quad \text{with} \]
\[ \Delta_\mu \overset{\text{def}}{=} \text{Diag}(\ldots, \mu_{ab}, \ldots), \quad (a,b) \in P_h. \tag{40} \]
We proceed with evaluating the covariance of the field at $\theta = 0$
\[ \text{Cov}(\phi(0, \mathbf{x})) = \text{Cov}(\mathcal{H}^H \text{vec}(f(\Pi, \mathbf{x} x^H))) \]
\[ = \mathcal{H}^H \text{Cov}(\text{vec}(f(\Pi, \mathbf{x} x^H))) \mathcal{H} \]
\[ = \mathcal{H}^H \mathcal{H} = \Delta_\theta \]
The first equality is by definition of $\phi$, the second equality is by the bilinearity of the $\text{Cov}$ operator, the third equality is by noting that Lemma 3 also reads $\text{Cov}(\text{vec}(f(\Pi, \mathbf{x} x^H))) \mathcal{H} = \mathcal{H} \Delta_\theta$, with $\Delta_\theta$ defined by (41), and the last equality is due to the orthonormality of the basis $\{H_{ab}\}$ and allows us to conclude that for independent observations
\[ \Gamma \overset{\text{def}}{=} \text{Cov}(\phi(0, \mathbf{x})) = \Delta_\theta \quad \text{with} \]
\[ \Delta_\theta \overset{\text{def}}{=} \text{Diag}(\ldots, \Sigma_{ab}, \ldots), \quad (a,b) \in P_h. \tag{41} \]
Thus, both $\Gamma$ and $D$ are diagonal matrices. In this case, the Lyapunov equation (13) reduces to $r(2n - r)$ uncoupled scalar equations. The solution then, trivially, is
\[ C_\theta = -\frac{1}{2} \Delta_\theta \Delta_\theta^{-1}. \tag{42} \]

According to (12), $\gamma^{-1/2} \theta_t \rightarrow \mathcal{N}(0, -\frac{1}{2} \Delta_\theta \Delta_\theta^{-1})$. By (36), we have $\text{vec}(\mathbf{P}_t) = \text{vec}(\Pi) + \mathcal{H} \theta_t + O(||\theta||^2)$, We conclude that for $\gamma \rightarrow 0$ and $t \rightarrow +\infty$
\[ \frac{1}{\sqrt{\gamma}}(\text{vec}(\mathbf{P}_t) - \text{vec}(\Pi)) \rightarrow \mathcal{N}(0, C_P) \quad \text{with} \]
\[ C_P = \mathcal{H} \Delta_\theta \mathcal{H}^H = -\frac{1}{2} \mathcal{H} \Delta_\theta \Delta_\theta^{-1} \mathcal{H}^H. \tag{43} \]
Expression (43) of the covariance matrix $C_P$ in the asymptotic distribution of $\text{vec}(\mathbf{P}_t)$ may be written as an explicit sum
\[ C_P = \sum_{(a,b) \in P_h} \frac{1}{2} \mu_{ab} \text{vec}(H_{ab}) \text{vec}(H_{ab})^H. \tag{44} \]
Definitions (24) of $\mu_{ab}$ and (30) of $\nu_{ab}$ show that these quantities are symmetric and that $\nu_{ab} = 0$ for $1 \leq a, b \leq r$. Using these facts and the easily established identity
\[
\text{Vec}(H_{ab})\text{Vec}(H_{ab})^H + \text{Vec}(H_{ba})\text{Vec}(H_{ba})^H = \Pi_a \otimes \Pi_b + \Pi_b \otimes \Pi_a,
\]
the expression (44) of the asymptotic covariance matrix of $P_t$ is finally rewritten as
\[
C_P = \sum_{1 \leq a \leq r \text{ and } d \leq n} \frac{\lambda_a \lambda_b}{2(\lambda_a - \lambda_b)}(\Pi_a \otimes \Pi_b + \Pi_b \otimes \Pi_a).
\]

See Corollary 1 for an even simpler expression in the case where $R = \Sigma + \sigma^2 I$, and $\Pi$ is the projector onto the range of a rank-deficient non-negative matrix $\Sigma$.

**F. Analysis**

Several simple MSE characterizations can be derived from the regular structure of the covariance matrix $C_P$ as expressed by (44) or (46). A word of caution is nonetheless necessary because the convergence of $(1/\sqrt{n})\text{Vec}(P_t - \Pi)$ to a limiting Gaussian distribution with covariance matrix $C_P$ does not guarantee the convergence of its moments to those of the limiting Gaussian distribution (or even that these limiting moments exist). In the following sections, we assume the convergence of the first- and second-order moments, allowing us to write
\[
\text{cov}(\text{Vec}(P_t)) = \gamma C_P + o(\gamma),
\]
for the numerical experiments presented in Section IV show an excellent prediction of various MSE quantities, confirming this assumption.

According to (47), the MSE between $P_t$ and $\Pi$ is given by the trace of the covariance matrix in the asymptotic distribution of $P_t$. The trace being invariant under orthonormal changes of basis and $\{\text{Vec}(H_{ab})\}_{1 \leq a, b \leq n}$ being an orthonormal basis, we obtain from (44) or (46)
\[
E||P_t - \Pi||^2_{2,\text{obs}} = \gamma \sum_{1 \leq a \leq r \text{ and } d \leq n} \frac{\lambda_a \lambda_b}{\lambda_a - \lambda_b} + o(\gamma).
\]

A finer picture is obtained by decomposing the error $P_t - \Pi$ into three terms
\[
P_1 := \Pi_0 (P - \Pi) \Pi,
\]
\[
P_2 := \Pi_0 \Pi_1 \Pi^\perp + \Pi^\perp \Pi_0 \Pi,
\]
\[
P_3 := \Pi^\perp \Pi_1 \Pi^\perp.
\]

Using $I_0 = \Pi + \Pi^\perp$, this is easily seen to be an orthogonal decomposition:
\[
P = \Pi + P_1 + P_2 + P_3,
\]
\[
||P - \Pi||^2 = ||P_1||^2 + ||P_2||^2 + ||P_3||^2.
\]

The relevance of this orthogonal decomposition stems from this lemma (see proof in Appendix):

**Lemma 5:** Let $P$ be a rank-$r$ Hermitian matrix, and let $\hat{\Pi}$ be the orthogonal projection matrix on the range space of $P$. Then
\[
P_1 = P^2 - P + O(||P - \Pi||^2)
\]
\[
P_2 = \hat{\Pi} - \Pi + O(||P - \Pi||^2)
\]
\[
P_3 = O(||P - \Pi||^2).
\]

Thanks to this lemma, each term can be given a simple interpretation. The term $P_1$ represents the first order of the deviation of $P$ from orthogonality since if $P$ was an orthogonal projector, we would have $P^2 = P$. Maybe more strikingly, recalling that $P = WW^H$, we find that
\[
||P_1||^2 = ||W^H W - I_r||^2 + O(||P - \Pi||^2)
\]

since from (98) and (99), $||P_1||^2 = ||V(A - I_r)V^H||^2 + O(||P - \Pi||^2)$, where $||V(A - I_r)V^H||^2 = ||A - I_r||^2 = ||U(A - I_r)U^H||^2$ with $U$ is the or $r \times r$ unitary matrix of the singular decomposition of $W^H W = U \Sigma T^H$. According to this lemma, the term $P_2$ represents the deviation between the subspace of interest and the one estimated by $\Pi$. Finally, the last term $P_3$ is of order $O(||P - \Pi||^2)$ because $P$ has rank $r$ (we already expressed this property in Lemma 4).

The above decomposition is purely geometric. Statistical results are obtained by combining them with expressions (44) or (46) of the asymptotic covariance matrix $C_P$ of $P_t$. In doing so, massive simplifications occur due to orthogonality. This is summarized by
\[
\text{Vec}(v_t v_t^H) = \delta_{i,j} \delta_{\ell,m} \delta_{c,d} \delta_{\ell,\ell} = 1, \leq a, b, c, d, \ell < n.
\]

A very striking result is observed here: The deviation of $W_t$ from orthonormality, as quantified by $||P_1||^2$, has a stochastic order lower than $r^{1/2}$. This results from the fact that the summation in (46) only is over pairs such that $a \leq r < b$. This is a remarkable feature of this algorithm in that there seems to be no price to pay for not constraining matrix $W$ to have orthonormal columns, at least in the stationary setting considered herein. What is then the order of the deviation from orthonormality? This question cannot be answered by first-order performance analysis, but the order can be determined experimentally. We show in Section IV that the MSE of orthonormality is, in first approximation, proportional to $r^{1/2}$. A frequently encountered situation is when the observation vector $x_t$ is the superposition of a signal with a rank-deficient
covariance matrix $\Sigma$ corrupted by a spatially incoherent additive noise with covariance matrix $\sigma^2 I$ so that $R = \Sigma + \sigma^2 I$. The projector of interest is $P_f$, which is the projector onto the range of $\Sigma$ (this space is usually called the signal subspace) or, equivalently, the projector $P_f^\perp$ onto its ortho-complement: the noise subspace. In this situation, this corollary follows.

Corollary 1: If $R = \Sigma + \sigma^2 I$ with rank $(\Sigma) = r < n$ and $P_f$ is the projector onto the range of $\Sigma$, then (46) of $C_f$ becomes

$$C_f = \Gamma \otimes P_f^\perp + P_f^\perp \otimes \Gamma$$

$$\Gamma = \frac{\sigma^2 P_f^\perp + \sigma^4 \Sigma^\#}{2} = \frac{\sigma^2}{2} \Sigma^\# R_f.$$

In addition, $\text{Tr}(C_f) = \sigma^2 (n-r) \text{Tr}(\Sigma^\# R_f)$.

The proof uses a decomposition of $\Sigma$ as $\Sigma = \sum_{a \leq r} \rho_a P_a$ giving $\lambda_a = \rho_a \sigma^2 + \sigma^2$ and $\Sigma^\# = \sum_{a \leq r} \rho_a^{-1} P_a$ as well as the property $\text{Tr}(P_f^\perp \otimes \Gamma) = \text{Tr}(P_f^\perp) \text{Tr}(\Gamma)$. The corollary follows by elementary algebra.

III. FURTHER INVESTIGATIONS

A. The Real Case

To address the case of real signals, only slight modifications are needed. We outline below the differences with the complex case. Regarding the parameterization of $P_f$, an appropriate orthonormal basis for the real symmetric matrices is

$$S_{ab} = \begin{cases} v_a v_b^T & a = b \\ v_a v_b^T + \lambda_a v_b v_a^T & a < b \end{cases} \frac{1}{\sqrt{2}}. \quad (57)$$

A lemma similar to Lemma 4 holds, replacing $P_h$ with $P_h \triangleq \{a,b\}_{1 \leq a \leq b \leq n}$ and $a < r$. Keeping the same definitions for $\mu_{ab}$ and $\nu_{ab}$, the results of Lemmas 1 and 3 become

$$E[f(H + cS_{ab}, x, x^T)] = c\mu_{ab}S_{ab} + O(c^2), \quad 1 \leq a \leq b \leq n \quad (58)$$

$$\text{Cov}(\text{Vec}(f(H, x, x^T))) = \text{Vec}(S_{ab}), \quad 1 \leq a \leq b \leq n. \quad (59)$$

Using the same arguments as in the real case, with the only difference ([16], p. 57) that

$$\text{Cov}(\text{Vec}(xx^T)) = R \otimes R + (R \otimes R)K \quad (60)$$

where $K$ is an $n^2 \times n^2$ block matrix, acting as a permutation operator, i.e., $K \text{Vec}(xy^T) = \text{Vec}(yx^T)$ for any vectors $x$ and $y$. The asymptotic covariance matrix for $P_f$ is

$$C_f = \sum_{(a,b) \subset P_h} \frac{1}{f_{ab}} \text{Vec}(S_{ab}) \text{Vec}(S_{ab})^T \quad (61)$$

which is the real counterpart of (44). $C_f$ is finally written as

$$C_f = \sum_{1 \leq a \leq r < b \leq n} \frac{\lambda_a \lambda_b}{2(\lambda_a - \lambda_b)} (I + K) \cdot (P_a \otimes P_b + P_b \otimes P_a). \quad (62)$$

Because of the similarity between the asymptotic covariance matrices for $P_f$ in the real case (62) and in the complex case (46), similar conclusions can be drawn. In particular, (55) holds.

B. Comparisons with Other Estimation Techniques

PAST Algorithm: Expression (62) can be compared with those derived from the asymptotic distribution of the real PAST estimator given by Yang in [9]. As this PAST algorithm converges to $W_* = V_f U$ (where $U$ is an $r \times r$ arbitrary orthogonal matrix), Yang considers the covariance $C_{W_1}$ of the limiting distribution of $W = W_u U^T$. The $nr \times nr$ block matrix $C_{W_1}$ reads

$$C_{W_1} = \left\{ \begin{array}{ll} 0 & 1 \leq i \neq j \leq r \\ \sum_{a > b \geq 1} \frac{\lambda_a}{\lambda_a - \lambda_b} v_a v_a^T & 1 \leq i = j \leq r \end{array} \right. \quad (63)$$

Applying a continuity theorem (e.g., [19, Th. 6.2a, p. 386]) to the differentiable mapping $W = (w_1, \cdots, w_r) \rightarrow P = \sum_{a = 1}^r w_a w_a^T$ gives $(1/\sqrt{r}) (\text{Vec}(P_f) - \text{Vec}(P)) \rightarrow_{d} N(0, C_f)$ with

$$C_f = \sum_{1 \leq a \leq r < b \leq n} \frac{\lambda_a}{\lambda_a - \lambda_b} (I + K) \cdot (P_a \otimes P_b + P_b \otimes P_a). \quad (64)$$

where $\gamma^2 \triangleq 1 - \lambda$, with $\lambda$ the forgetting factor of the RLS-type PAST algorithm. The relations (62) and (64) are very similar, except for the multiplicative term $\lambda_a$. Further comparison of these covariance matrices is questionable because the step size $\gamma$ cannot be compared. In our algorithm, $\gamma$ has the dimension of the inverse of the power of $x_a$, and in the PAST algorithm, $\gamma$ has no dimension.

Batch Estimation Technique: Consider the minimization of the sample version of the Yang/Xu criterion (3)

$$J_b(W) \triangleq \frac{1}{T} \sum_{t=1}^T ||x_t - WW^T x_t||_{Fro}^2. \quad (65)$$

Let $W_T$ be a minimizer of (65). Because all the properties derived from the minimization of (3) also hold in case (65) is minimized (just replace $E(x_t x_t^T)$ by $(1/T) \sum_{t=1}^T x_t x_t^T$), the corresponding estimate of the projector $P$ has a common value denoted by $P_T \triangleq W_T W_T^T$ and coincides with the standard batch estimator. Therefore, since $W_T$ is solution of $(1/T) \sum_{t=1}^T h(W, x_t) = 0$ with $h$ defined in (5), $P_T$ is a solution of

$$\frac{1}{T} \sum_{t=1}^T f(P, x_t x_t^T) = 0 \quad (66)$$

with $f$ defined in (9). Applying the results on M-estimates recalled in Subsection II-B and using the common set of eigenvectors $\text{Vec}(H_{ab})_{1 \leq a \leq r \leq n}$ of $D$ and $\Gamma$ and the associated eigenvalues (24), (30), we obtain

$$\sqrt{T} (\text{Vec}(P_T) - \text{Vec}(P)) \rightarrow_{d} N(0, C_f) \quad (67)$$

with

$$C_f \triangleq \sum_{1 \leq a \leq r < b \leq n} \frac{\lambda_a \lambda_b}{(\lambda_a - \lambda_b)^2} (P_a \otimes P_b + P_b \otimes P_a). \quad (68)$$
Comparison: The covariance matrices $C_P$ of the asymptotic distribution of the projection matrix estimated by our LMS algorithm and by the standard batch estimation technique have the same structure [see (18) and (21)] because these estimations derive from the same criterion [see (3) and (65)]. The matrices $C_P$ share the same eigenbasis $H_{ab}$ in the complex case and $S_{ab}$ in the real case. Only the associated eigenvalues are distinct $\lambda_\alpha\lambda_\beta/2(\lambda_\alpha - \lambda_\beta)$ for our LMS algorithm and $\lambda_\alpha\lambda_\beta/(\lambda_\alpha - \lambda_\beta)^2$ for the standard batch estimation. The covariance matrix $C_P$ associated with the real PAST algorithm shares the same eigenbasis $S_{ab}$ but with eigenvalues $\lambda_\alpha/2(\lambda_\alpha - \lambda_\beta)$, as shown in the table at the bottom of the page.

C. Application to DOA Tracking

By continuity, the behavior of any differentiable function of the projection matrix $P_t$ can be obtained. In this section, the behavior of the DOA’s estimated by the MUSIC algorithm [21] applied from $P_t$ is derived. Recall first the standard narrowband array data model

$$x_t = E(\theta)s_t + n_t$$

where $s_t$ and $n_t$ are mutually uncorrelated zero-mean circular Gaussian stationary processes. Matrix $E(\theta)$ is partitioned into column vectors as $E(\theta) \stackrel{\text{def}}{=} [e(\theta_1), \ldots, e(\theta_r)]$. In addition, we assume a spatially white noise $E(n_tn_t^H) = \sigma^2 I_n$ and not fully correlated sources so that $R_x = E(s_t s_t^H)$ is full rank. The MUSIC estimates of the DOA’s $\theta_{ik}, k = 1, \ldots, r$ are determined as the $r$ deepest minima of the localization function

$$e^H(\theta)(I_n - P_t)e(\theta) = \text{Tr}((I_n - P_t)M(\theta))$$

with $M(\theta) \stackrel{\text{def}}{=} e(\theta)e^H(\theta)$. Therefore, each DOA estimate $\hat{\theta}_k$ of $\theta_k$ based on an estimate $P$ of $I$ is the solution of

$$\text{Tr}((I_n - P)M'(\hat{\theta}_k)) = 0$$

where $M'(\theta) \stackrel{\text{def}}{=} (dM/d\theta)$. The first-order perturbation of $\hat{\theta}_k$ is easily found by a Taylor expansion of (71). This is

$$\hat{\theta}_k = \theta_k + \text{Vec}(D_k)H\text{Vec}(P - I) + \alpha(||P - I||)$$

with

$$D_k = \frac{M'(\theta_k)}{\text{Tr}(I - M'(\theta_k))}.$$

If the derivative matrix $D_k$ is nonzero, a continuity theorem (e.g., [19, Th. 6.2, p. 386]) gives the convergence in distribution of $\Theta_k = (\theta_1, \ldots, \theta_r)$, namely, $\gamma^{-1/2}(\Theta_k - \Theta) \Rightarrow \mathcal{N}(0, C_\theta)$ when $\gamma \to 0$ and $t \to +\infty$ with and asymptotic covariance matrix $C_\theta$ given component-wise by

$$C_{\theta_{ij}} = \text{Vec}(D_i)H C_P \text{Vec}(D_j) = \text{Tr}(D_i I - D_j)$$

where last equation uses Corollary 1. Next, using

$$\text{Tr}(H^{-1}M'(\theta)) = 2\epsilon^H(\theta)H^{-1}\epsilon(\theta), M'(\theta) = e(\theta)e^H(\theta) + \epsilon(\theta)e^H(\theta), \epsilon(\theta) = \text{vec}(\epsilon^H(\theta)/d\theta), \text{ and } \Gamma = (\sigma^2/2)\sum_{\alpha<\beta} \lambda_\alpha(\lambda_\alpha - \lambda_\beta)^2$$

from Lemma 6, (Cov $\theta_{ij})_{i,j}$ reads

$$\text{Cov}(\Theta_{ij}) = \gamma(C_\theta)_{i,j} + o(\gamma)$$

so that in particular

$$\text{Var} \hat{\theta}_1 = \gamma(C_\theta)_{1,1} + o(\gamma)$$

where $C_\theta$, $\alpha$, and $\epsilon_{ij}$ are given by

$$\alpha_i \stackrel{\text{def}}{=} 2\epsilon^H(\theta_i)I - \epsilon(\theta_i)$$

and

$$\epsilon_{ij} \stackrel{\text{def}}{=} 2\epsilon^H(\theta_i)I - \epsilon(\theta_j).$$

Comparison with Batch Estimation: We note that (74) is quite close to the expression (3.12, [22]) of the covariance of the DOA’s estimated by a batch MUSIC algorithm based on $T$ independent snapshots $x_t$ as given by

$$\text{Cov}(\Theta_{ij})_{i,j} \sim \frac{2}{T\alpha_i\alpha_j} \lambda_\alpha \sigma^2 \sum_{\alpha_1} \left[ \frac{\lambda_\alpha}{\lambda_\beta(\lambda_\alpha - \lambda_\beta)^2} \text{Tr}(\epsilon^H(\theta_i)\epsilon(\theta_j)) \right] + o(\gamma)$$

where $\text{Re}(\cdot)$ denotes “the real part of.” To gain a better insight into the result (74), we now specialize to a simple case and present some properties.

Consider the case of a single source impinging on any array. If we denote by $\sigma^2_1$, the power of the signal source $s_t$ with steering vector normalized by $n$ $(|e(\theta)|^2 = n)$, (75) gives

$$\text{Var} \hat{\theta}_1 \sim \frac{\gamma \sigma^2_1}{2\alpha_1} \left(1 + \frac{1}{n\rho_1}\right) \frac{1}{\rho_1}$$

that we can compare with the batch MUSIC estimation

$$\text{Var} \hat{\theta}_1 \sim \frac{1}{T} \frac{1}{\alpha_1} \left(1 + \frac{1}{n\rho_1}\right) \frac{1}{\rho_1}$$

where $\rho_1$ denotes the SNR $\sigma^2_1/\sigma^2$. For example, the geometrical factor $\alpha_1$ is given for a linear uniform array of $n$ sensors by $\alpha_1 = (n(n^2 - 1)/6)$. The results (78) and (79) are directly induced by the expressions of the asymptotic covariance (46) and (68) issued from the minimization of the same criterion. Therefore, the ratio $\gamma \lambda_\alpha \lambda_\beta/(\lambda_\alpha - \lambda_\beta)^2 = (\gamma T(\lambda_\alpha - \lambda_\beta)^2) = (\gamma T\sigma^2_1/2)$ for

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>LMS algorithm</th>
<th>batch estimation</th>
<th>PAST algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eigenvalues</td>
<td>$\frac{\lambda_\alpha \lambda_\beta}{2(\lambda_\alpha - \lambda_\beta)}$</td>
<td>$\frac{\lambda_\alpha \lambda_\beta}{(\lambda_\alpha - \lambda_\beta)^2}$</td>
<td>$\frac{\lambda_\beta}{2(\lambda_\alpha - \lambda_\beta)}$</td>
</tr>
</tbody>
</table>
the variances of the estimated DOA by MUSIC algorithms because \( \lambda_0 = n\sigma_1^2 + \sigma_2^2 \) and \( \lambda_0 = \sigma_2^2 \). Regarding the speed of convergence of the adaptive MUSIC algorithm, we note that the product \([\text{step size} \times \text{eigenvalues of the derivative of the mean field} = \gamma(\lambda_0 - \lambda_0) = \gamma n\sigma_1^2] \) must be fixed to fix the convergence speed. This suggests that the step size \( \gamma \) of the adaptive algorithm must be normalized by \( n\sigma_1^2 \) to keep a fixed convergence speed (see also Section IV).

**Equipowered Sources:** In the case of \( r \) equipowered and uncorrelated sources (with normalized steering vectors), the \( r \) variances \( \text{Var}_{\theta_i} = 1, \ldots, r \) are equal and given by (78) obtained for a single source. The influence of other sources manifests itself only in the geometrical factors \( \alpha_i \). In batch MUSIC estimation, the variances cannot be reduced to the form (79).

For example, for \( r = 2 \) sources impinging on a linear uniform array, the increasing variance of the DOA estimates when the sources get closer is due only to the geometrical factors \( \alpha_i \). From the closed-form expressions of \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) given in [23], it is straightforward to see that the \( \{\alpha_i\} \) are a decreasing function of the DOA separation \( \theta_2 - \theta_1 = 0 \) and tend to 0 when \( \theta_2 - \theta_1 \) tends to 0.

From (75), we have after some algebraic manipulations

\[
\text{Var}_{\theta_i} \sim \frac{\gamma n\sigma_1^2}{2\alpha_i} \left( |e^{i\theta_i}|^2 + \sigma_2^2 \sum_{a=1}^r \frac{|e^{iH(\theta_i)w_a}|^2}{\lambda_0 - \sigma_2^2} \right)
\]

(80)

By symmetry, however, since \( |e^{iH(\theta_i)w_a}|^2 \) does not depend on \( i \), we can deduce from

\[
\left( \sigma_1^2 \sum_{k=1}^r e^{i\theta_k} e^{iH(\theta_k)} \right) w_a = (\lambda_0 - \sigma_2^2) w_a, \quad a = 1, \ldots, r
\]

that \( \sigma_1^2 |e^{iH(\theta_i)w_a}|^2 = \lambda_0 - \sigma_2^2, a = 1, \ldots, r \). Relation (78) is readily deduced.

For high SNR, \( \lambda_0 - \sigma_2^2 \sim \lambda_0 \), for \( a = 1, \ldots, r \) so that (74) becomes

\[
(\text{Cov} \theta)_{i,j} \sim \frac{\gamma n\sigma_1^2}{2\lambda_0 \alpha_i} \text{Re} [e^{i\theta_i} e^{iH(\theta_j)}]
\]

(81)

and in batch MUSIC estimation, thanks to \( (\lambda_0 \sigma_2^2/(\lambda_0 - \sigma_2^2)) \sim (\sigma_2^2/(\lambda_0 - \sigma_2^2)) \) for \( a = 1, \ldots, r \), (77) gives

\[
(\text{Cov} \theta)_{i,j} \sim \frac{2\sigma_2^2}{T \alpha_i \alpha_j} \text{Re} [e^{i\theta_i} (\mathbf{R}_s^{-1})_{i,j}].
\]

(82)

In particular, the diagonal elements of (81) and (82), respectively, read

\[
\text{Var}_{\theta_i} \sim \frac{\gamma n\sigma_1^2}{2\alpha_i} |e^{i\theta_i}|^2 \quad \text{and} \quad \text{Var}_{\theta_j} \sim \frac{2\sigma_2^2}{T \alpha_j} (\mathbf{R}_s^{-1})_{j,j}.
\]

(83)

Thus, in striking difference to batch MUSIC, the variances in adaptive MUSIC are identical to those predicted for a single source in a high SNR situation. Furthermore, the \( i \) and \( j \) DOA estimates are uncorrelated in adaptive MUSIC [respectively, in batch MUSIC] if and only if \( \text{Re} [e^{i\theta_i} e^{iH(\theta_j)} e^{i\theta_j}] = 0 \) [respectively, \( \text{Re} [e^{i\theta_i} (\mathbf{R}_s^{-1})_{i,j}] = 0 \)]. This holds in particular for orthogonal steering vectors [respectively, for uncorrelated sources].

**IV. SIMULATIONS RESULTS**

We now examine the accuracy of (48) of the mean square error of the projection matrix and of (75) and (78) of the misadjustment of the DOA’s estimated by the MUSIC algorithm and “small terms” exhibited in Section II-F.

In the first experiment, we consider the case of the projection matrix \( \mathbf{P}_t \) on the eigenspace spanned by the first two eigenvectors of a \( 3 \times 3 \) covariance matrix \( \mathbf{R} \) derived from independent observations \( \mathbf{x}_t \), whose covariance matrix is that of an AR(1) model of parameter \( \alpha_1 \). Fig. 1 shows the learning curve averaged (over 100 independent runs) of the square error \( \|\mathbf{P}_t - \mathbf{I}\|_{\text{Frob}}^2 \) for \( \alpha_1 = 0.3 \) and \( \gamma = 0.005 \). It tends to a value in perfect agreement with the theoretical values predicted by (48). This figure also shows the evolution of “small terms,” i.e., terms with scale \( \alpha(\gamma) \), as predicted by (47) and (54)–(56). We plot the quantities \( \|\mathbf{E}(\mathbf{P}_t) - \mathbf{I}\|_{\text{Frob}}^2, \|\mathbf{W}_t^H \mathbf{W}_t - \mathbf{I}\|_{\text{Frob}}^2, \|\mathbf{P}_t \|_{\text{Frob}}^2 \) and \( \|\mathbf{P}_t \mathbf{x}_t \|_{\text{Frob}}^2 \), whose significance has been discussed in Section II-F. Fig. 2 shows the ratio of the estimated mean square error \( \mathbf{E}[\|\mathbf{P}_t - \mathbf{I}\|_{\text{Frob}}^2] / \mathbf{E}[\|\mathbf{P}_t \|_{\text{Frob}}^2] \) to the theoretical asymptotic mean square error \( \gamma \text{Tr}(\mathbf{C}_p) \) as a function of \( \gamma \). Our present asymptotic analysis is seen to be valid over a large range of \( \gamma \) (\( \gamma < 0.01 \)), and the domain of “stability” is \( \gamma < 0.005 \)

Fig. 3 shows that the deviation from orthonormality \( \mathbf{P}(\gamma)(\text{def} \mathbf{W}_t^H \mathbf{W}_t - \mathbf{I}) \) is proportional to \( \gamma^2 \) in the domain of validity of the mean square error (48) because \( \log_{10} \mathbf{P}(\gamma) = \log_{10} \mathbf{R} + 2 \log_{10} \mathbf{C}_p \) with \( c \approx 1.210^{-2} \). Fig. 4 shows, in the same way, that the square norm of the bias \( \|\mathbf{E}(\mathbf{P}_t) - \mathbf{I}\|_{\text{Frob}}^2 \) is proportional to \( \gamma^2 \) in the domain of “stability.” Furthermore, its contribution to the MSE \( \mathbf{E}[\|\mathbf{P}_t - \mathbf{I}\|_{\text{Frob}}^2] \) proves that \( \mathbf{E}[\|\mathbf{P}_t - \mathbf{I}\|_{\text{Frob}}^2] \sim \gamma \text{Tr}(\mathbf{C}_p) \) in the domain of “stability.”

The third experiment presents the case of one source impinging on a linear uniform array of sensors. After each subspace update, we apply the MUSIC algorithm. Fig. 5 shows the learning curve of the mean square error of the DOA \( \theta_1 \) averaging 400 independent runs of independent observations \( \mathbf{x}_t \). We used the normalized step size \( \gamma = (\gamma_0/\sigma_1^2) \) with \( \gamma_0 = 0.003, \sigma_1^2 = 1, \rho_1 = 10 \). We see that for \( n = 3 \) and \( n = 6 \), the algorithms have about the same convergence speed. As for the mean square error, it agrees with (78), which gives \( \text{Var}_{\theta_1} \sim (3\gamma^2/\sigma_1^2) \approx (3\gamma_0^2/\rho_1) \).

The convergence time \( T \approx 10000 \) in Fig. 5 can be compared with the observation time \( T \) necessary to get the same batch MSE as the sequential asymptotic MSE; expressions (78) and (79) give the observation time \( T = (2/\gamma^2 \sigma_1^2) \approx (2/\gamma_0) \approx 607 \).

Finally, we present the case of two uncorrelated and equipowered sources impinging on a linear uniform array of four sensors with \( \rho_1 = 10, \theta_1 = \theta_2 = 0 \). After each subspace update, we apply the MUSIC algorithm. Fig. 6 shows the learning curve of the DOA \( \theta_1 \) averaging 400 independent runs of independent observations \( \mathbf{x}_t \) for \( \gamma = 0.003, \delta = 0.15(2\pi/4) \); then, \( \delta = 0.30(2\pi/4) \). The mean square error agrees with (78), and the convergence speed is seen to decrease when the angle separation decreases. This agrees with the eigenvalues of the derivative of the mean field (24);
we have the relationship $E||P_t - I||_F^2 = \gamma \text{Tr}(C_P)$ for independent observations $x_t$. Since $\gamma = 0.005$, the ratio of the estimated mean square error $E||P_t - I||_F^2$ by averaging 400 independent runs to the theoretical asymptotic mean square error $\gamma \text{Tr}(C_P)$ as a function of $\gamma$ for independent observations $x_t$ and $a_1 = 0.3$.

some of them are equal to $\lambda_a - \sigma^2, a = 1, 2$. Since $\lambda_2$ tends to $\sigma^2$ when the separation angle tends to zero, the eigenvalues $\lambda_a - \sigma^2$ also tend to zero, which implies slower convergence near the stationary point.

V. CONCLUSION

We have analyzed the performance of an LMS-type algorithm for tracking dominant invariant subspaces. Because this algorithm and others in its class do not converge to the...
Fig. 3. Deviation from orthogonality $d^2(\gamma) = E[\|W_t^T W_t - I\|_F^2]$ at "convergence" estimated by averaging 100 independent runs as a function of $\gamma$ in log-log scales.

Fig. 4. Mean square error $E[\|P_t - \Pi\|^2]$. (1) Square of bias $E[\|P_t - \Pi\|^2]$, (2) estimated by averaging 100 independent runs and theoretical asymptotic mean square error $\gamma \text{Tr}(C_P)$ as a function of $\gamma$ in log-log scales.

dominant eigenvectors but only to a rotated version, it is necessary to develop a particular methodology. The asymptotic covariance of the estimated associated projection matrix is given in closed form and is further analyzed to provide insights into the behavior of this LMS-type algorithm. In particular, it has been compared with the performance of batch estimation, which is derived from the same criterion and of Yang’s PAST algorithm. The accuracy of the asymptotic analysis appears to
be very good even for large step sizes, as shown by numerical experiments.

VI. PROOFS

Proof of Lemma 1: The field \( f \) in definition (9) being linear in its second argument, the mean field at any point \( P \) is

\[
E\{f(P, x_t x_t^H)\} = f(P, E(x_t x_t^H)) = f(P, R) \\
= P(2R - RP - PR) \\
+ (2R - RP - PR)P.
\]

(84)

Using the identities \( P^2 = P \), \( P R = R P = P R P \), and

\[
\Pi v_a v_b^H = \pi_a v_a v_b^H, \quad v_a v_b^H \Pi = \pi_b v_a v_b^H \\
R v_a v_b^H = \lambda_b v_a v_b^H, \quad v_a v_b^H R = \lambda_a v_a v_b^H
\]

a substitution \( P = \Pi + \alpha_a v_a v_b^H \) in (84) yields after simplification

\[
E\{f(\Pi + \alpha_a v_a v_b^H, x_t x_t^H)\} = \epsilon_{\alpha, a, b} v_a v_b^H + O(\epsilon^2)
\]

(85)

where \( \epsilon_{\alpha, a, b} \) is defined in (24). The lemma follows by using the symmetry \( \epsilon_{\alpha, a, b} = \epsilon_{\lambda, a, b} \). Q.E.D.

Proof of Lemma 2: At point \( P = \Pi \) and using \( \Pi \Pi = \Pi \), definition (9) of the field reduces to

\[
f(\Pi, x_t x_t^H) = \Pi^t x_t x_t^H \Pi + \Pi x_t x_t^H \Pi^t
\]

(86)

which, by vectorization and using definition (28), also reads

\[
\text{Vec}(f(\Pi, x_t x_t^H)) = Q \text{Vec}(x_t x_t^H).
\]

(87)

For a circular complex Gaussian vector \( x_t \), we have ([15], p. 336)

\[
\text{Cov}(\text{Vec}(x_t x_t^H)) = R \otimes R.
\]

(88)

Combining (87) and (88) establishes the lemma

\[
\text{Cov}(\text{Vec}(f(\Pi, x_t x_t^H))) = Q \text{Cov}(\text{Vec}(x_t x_t^H)) Q^H = Q(R \otimes R Q)^H.
\]

Q.E.D.

Proof of Lemma 3: For any pair \( 1 \leq a, b \leq n \), by simple substitution, we find

\[
(R \otimes R) \text{Vec}(v_a v_b^H) = \lambda_a \lambda_b \text{Vec}(v_a v_b^H)
\]

(89)

\[
Q \text{Vec}(v_a v_b^H) = (\pi_a - \pi_b)^2 \text{Vec}(v_a v_b^H)
\]

(90)

by using the properties \( R v_a v_b^H R = \lambda_a \lambda_b v_a v_b^H \) and \( \Pi v_a v_b^H \Pi = \pi_a (1 - \pi_b) v_a v_b^H \) and the identity \( \pi_a (1 - \pi_b) + \pi_b (1 - \pi_a) = \pi_a + \pi_b - 2\pi_a \pi_b = \pi_a^2 + \pi_b^2 - 2\pi_a \pi_b \) (since \( \pi_a = \pi_b^2 \)). The eigenvectors of \( \text{Cov}(\text{Vec}(f(\Pi, x_t x_t^H))) \) follow by recalling that \( Q = Q^H \) and combining (29), (89), and (90)

\[
\text{Cov}(\text{Vec}(f(\Pi, x_t x_t^H))) \text{Vec}(v_a v_b^H) = \nu_{ab} \text{Vec}(v_a v_b^H)
\]

(91)

where the scalars \( \nu_{ab} \) are defined in the lemma. Using \( \nu_{ab} = \nu_{ba} \), symmetrization of (91) finishes the proof. Q.E.D.

Proof of Lemma 4: The proof uses the more general Lemma 5. Indeed, we have

\[
\sum_{r < a, b \leq n} \theta_{ab}(P) = ||\Pi^t(P - \Pi)\Pi^t||^2
\]

which, by (53), \( \Pi^t(P - \Pi)\Pi^t = O(||P - \Pi||^2) \). Thus,

\[
\theta_{ab}(P) = O(||P - \Pi||^2) \text{ for } r < a, b \leq n.
\]

Q.E.D.

Proof of Lemma 5: Denote \( P = V \Lambda V^H \) as the eigenvalue decomposition of \( P \), and recall that \( \Pi \) denotes the orthogonal
Fig. 6. Learning curves of the mean square error of the DOA $\theta_1$ for two equipowered sources ($\theta_1 = \theta_0 \pm \delta$) estimated by MUSIC algorithm, where $\gamma = 0.003$, $\sigma^2 = 1$, $\rho = 10$ for $\delta = 0.15$ beamwidth and $\delta = 0.30$ beamwidth compared with the theoretical asymptotic mean square error.

The main difficulty is that this EVD is not differentiable at point $P = \Pi$ because the eigenvalues of $\Pi$ are degenerate. However, results (see [17, Th. 5.4, p. 111]) are available for the perturbation of the projector $VV^H$ and of the eigenvalues. This is

$$VV^H = \Pi + \Pi(P - \Pi)\Pi + \Pi(P - \Pi)\Pi + O(||P - \Pi||^2)$$

$$\Lambda = I_r + O(||P - \Pi||).$$

Based on this, we derive two preliminary results: $\Pi^\perp V = \Pi^\perp V = \Pi^\perp V = \Pi^\perp V = O(||P - \Pi||^2)$, where the last equality results from (93), and $\Pi^\perp \Pi^\perp = 0$. It follows that

$$P = PVV^H = PVV^H + V(A - I_r)PV^H$$

$$= \hat{\Pi} + V(A - I_r)PV^H.$$  (92)

which establishes (52). Finally, property (51) is established as

$$P^c - P = (VAV^H)^2 - VAV^H = V(A^2 - A)V^H$$

$$= V(A - I_r) + (A - I_r)^2 V^H$$

$$= V(A - I_r)PV^H + O(||P - \Pi||^2)$$

$$= P - \hat{\Pi} + O(||P - \Pi||^2)$$

$$= (P + P_1 + P_2 + P_3)$$

$$= (P + P_2 + O(||P - \Pi||^2)) + O(||P - \Pi||^2)$$

by using (92), (50), (97), and (96). This completes the proof of the lemma.

ACKNOWLEDGMENT

The authors gratefully acknowledge the constructive criticism of several anonymous referees, which helped us in improving the first version of the manuscript.

REFERENCES


**Jean Pierre Delmas** was born in France in 1950. He received the engineering degree from Ecole Centrale de Lyon, Lyon, France, in 1973 and the certificat d’études supérieures from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1982.

Since 1980, he has been with the Institut National des Télécommunications, Evry, France, where he is currently Maitre de conférences. His research interests are in statistical signal processing.

**Jean François Cardoso** was born in 1958 and received the agrégation de physique degree from the Ecole Normale Supérieure de Saint-Cloud, Saint-Cloud, France, in 1981 and the doctorat de physique degree from the University of Paris, Paris, France, in 1984.

He currently is with the Centre National de la Recherche Scientifique (CNRS), Paris, and works in the Signal and Images Department of Télécom Paris (a.k.a. ENST). His research interests are in statistical signal processing, with emphasis on (blind) array processing and performance analysis and connections to information theory.