



**HAL**  
open science

## Performance analysis of an adaptive algorithm for tracking dominant subspaces

Jean-Pierre Delmas, Jean-François Cardoso

► **To cite this version:**

Jean-Pierre Delmas, Jean-François Cardoso. Performance analysis of an adaptive algorithm for tracking dominant subspaces. IEEE Transactions on Signal Processing, 1998. hal-03435753

**HAL Id: hal-03435753**

**<https://hal.science/hal-03435753>**

Submitted on 18 Nov 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Performance analysis of an adaptive algorithm for tracking dominant subspaces.

Jean-Pierre Delmas \*

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  $\mathbf{W}$  to an orthonormal basis of a dominant invariant subspace of a correlation matrix. We consider the stochastic algorithm governing the evolution of  $\mathbf{W}\mathbf{W}^H$  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 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  $\mathbf{W}$  to have orthonormal columns, there is deviation from orthonormality at first order. We also give a closed form expression of the asymptotic covariance of DOAs estimated by the MUSIC algorithm applied to the adaptive estimate of the projector. It is found that the asymptotic distributions have a structure 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.

---

\*Institut National des Télécommunications. 9 rue Charles Fourier, 91011 Paris Cedex, France. tel: (33)-1-60 76 46

32. fax: (33)-1-60 77 55 74. Email: [delmas@int-evry.fr](mailto:delmas@int-evry.fr)

†Ecole Nationale Supérieure des Télécommunications. 46 rue Barrault, 75634 Paris Cedex, France. tel: (33)-1-45 81

78 59. fax: (33)-1-45 88 79 35. Email: [cardoso@sig.enst.fr](mailto:cardoso@sig.enst.fr)

Keyword: subspace tracking, asymptotic performance analysis, adaptive/batch MUSIC.

## 1 Introduction

**Subspace tracking.** Over 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 in [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, while 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], [8] and [9]. It is the purpose of this paper to provide a thorough study of the behavior of a LMS type approximation algorithm presented by Yang in [2].

**Algorithm under study.** For a given  $n \times n$  covariance matrix  $\mathbf{R}$ , denote  $\lambda_1 \geq \dots \geq \lambda_n$  the eigenvalues of  $\mathbf{R}$  and  $\mathbf{v}_1, \dots, \mathbf{v}_n$  corresponding eigen-vectors. The  $r$ -dimensional dominant invariant subspace of  $\mathbf{R}$  is the span of  $\mathbf{v}_1, \dots, \mathbf{v}_r$  and it is well defined if, as assumed throughout the paper,  $\lambda_r > \lambda_{r+1}$ . Denote  $\mathbf{\Pi}$  the orthogonal projector onto this subspace. One has:

$$\mathbf{R}\mathbf{v}_a = \lambda_a\mathbf{v}_a \quad \mathbf{\Pi}\mathbf{v}_a = \mathbf{v}_a \quad 1 \leq a \leq n. \quad (1)$$

where we have defined  $\pi_1 = \dots = \pi_r = 1$  and  $\pi_{r+1} = \dots = \pi_n = 0$ . Defining the rank-one projection matrices  $\mathbf{\Pi}_a$  onto each eigen-vector, one can also write

$$\mathbf{\Pi}_a \stackrel{\text{def}}{=} \mathbf{v}_a \mathbf{v}_a^H \quad \mathbf{R} = \sum_{a=1,n} \lambda_a \mathbf{\Pi}_a \quad \mathbf{\Pi} = \sum_{a=1,n} \pi_a \mathbf{\Pi}_a = \sum_{1 \leq a \leq r} \mathbf{\Pi}_a \quad (2)$$

Subspace tracking consists in recursively updating at time  $t$  an (approximately) orthonormal basis of this subspace upon reception of sample  $\mathbf{x}_t$  of a stationary process with covariance  $\mathbf{R} = E \mathbf{x}_t \mathbf{x}_t^H$ ,<sup>1</sup>  $\mathbf{x}_t$  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:

$$J_a(\mathbf{W}) \stackrel{\text{def}}{=} E \|\mathbf{x}_t - \mathbf{W} \mathbf{W}^H \mathbf{x}_t\|_{\text{Fro}}^2 \quad (3)$$

with respect to the  $n \times r$  matrix  $\mathbf{W}$ . In this contribution, we consider the stochastic gradient algorithm for the minimization of  $J_a(\mathbf{W})$ . This yields the following algorithm where  $\gamma_t > 0$  is a step size:

$$\mathbf{W}_{t+1} = \mathbf{W}_t + \gamma_t h(\mathbf{W}_t, \mathbf{x}_t) \quad (4)$$

$$h(\mathbf{W}_t, \mathbf{x}_t) \stackrel{\text{def}}{=} (2\mathbf{x}_t \mathbf{x}_t^H - \mathbf{x}_t \mathbf{x}_t^H \mathbf{W}_t \mathbf{W}_t^H - \mathbf{W}_t \mathbf{W}_t^H \mathbf{x}_t \mathbf{x}_t^H) \mathbf{W}_t. \quad (5)$$

Baldi and Hornik [10] have shown (in the real case) that the stationary points of  $J_a(\mathbf{W})$  are  $\mathcal{W} = \{\mathbf{W} | \mathbf{W} = \mathbf{V}_r \mathbf{U}\}$  where  $\mathbf{V}_r$  is any  $r$ -dimensional eigenvector basis of  $\mathbf{R}$  and  $\mathbf{U}$  is an  $r \times r$  arbitrary orthogonal matrix. All these points are saddle points except when  $\mathbf{V}_r = \mathbf{V}_S \stackrel{\text{def}}{=} [\mathbf{v}_1, \dots, \mathbf{v}_r]$ . In this case  $J_a(\mathbf{W})$  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 L. Xu [11]. In stationary situations and if the step size  $\gamma_t$  satisfies the conditions  $\sum_{t=1}^{\infty} \gamma_t = +\infty$  and  $\lim_{t \rightarrow \infty} \gamma_t = 0$ , the study of convergence of algorithm (4) is intimately connected to the associated ordinary differential equation (ODE):

$$\frac{d\mathbf{W}_t}{dt} = E(h(\mathbf{W}_t, \mathbf{x})). \quad (6)$$

So, we can conjecture that the stochastic algorithm (4) converges almost surely to  $\mathcal{W}_S = \{\mathbf{W} | \mathbf{W} = \mathbf{V}_S \mathbf{U}\}$ .

---

<sup>1</sup>Complex valued quantities are implicitly assumed here. Most of the paper deals with the complex case, but the real case is also considered in section 3.1.

**Projector tracking.** A difficulty arises in the study of the behavior of  $\mathbf{W}_t$  because the set  $\mathcal{W}_S$  forms a *continuum* of attractors: the column vectors of  $\mathbf{W}_t$  do not in general tend to the eigenvectors  $\mathbf{v}_1, \dots, \mathbf{v}_r$  and we have no proof of convergence of  $\mathbf{W}_t$  to a particular orthonormal basis of their span.<sup>2</sup> Therefore, the approach followed in this paper is to study the trajectory of matrix  $\mathbf{P}_t$

$$\mathbf{P}_t \stackrel{\text{def}}{=} \mathbf{W}_t \mathbf{W}_t^H \quad (7)$$

whose dynamic is governed by the stochastic equation:

$$\mathbf{P}_{t+1} = \mathbf{P}_t + \gamma_t f(\mathbf{P}_t, \mathbf{x}_t \mathbf{x}_t^H) + \gamma_t^2 g(\mathbf{P}_t, \mathbf{x}_t \mathbf{x}_t^H) \quad (8)$$

$$f(\mathbf{P}, \mathbf{M}) \stackrel{\text{def}}{=} \mathbf{P}(2\mathbf{M} - \mathbf{M}\mathbf{P} - \mathbf{P}\mathbf{M}) + (2\mathbf{M} - \mathbf{M}\mathbf{P} - \mathbf{P}\mathbf{M})\mathbf{P} \quad (9)$$

$$g(\mathbf{P}, \mathbf{M}) \stackrel{\text{def}}{=} (2\mathbf{M} - \mathbf{M}\mathbf{P} - \mathbf{P}\mathbf{M})\mathbf{P}(2\mathbf{M} - \mathbf{M}\mathbf{P} - \mathbf{P}\mathbf{M}) \quad (10)$$

obtained by combining (4), (5) and (7). In the following, we are interested in first order asymptotic effects: we derive the asymptotic variance of  $\mathbf{P}_t$  around  $\mathbf{\Pi}$  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  $O(\gamma_t^2)$  term in (8), so that this term can be neglected. A remarkable feature of (8) is that the field  $f$  actually depends only on  $\mathbf{P}_t$  and *not* on  $\mathbf{W}_t$ . This fortunate circumstance makes it possible to study the evolution of  $\mathbf{P}_t$  without determining the evolution of the underlying matrix  $\mathbf{W}_t$ . The characteristics of  $\mathbf{P}_t$  are indeed the most interesting since they completely characterize the estimated subspace.

**Outline of the paper.** This paper is organized as follows. In section 2, after presenting a brief review of a general Gaussian approximation result, we obtain in closed form the asymptotic covariance of  $\mathbf{P}_t$  for the case where  $\mathbf{x}_t$  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 3, we extend this result to real signals and we compare the asymptotic performance of the algorithm with the performance of the PAST algorithm [2]

---

<sup>2</sup>We note, that in the particular case  $r = 1$  and  $\mathbf{x}_t$  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  $\{-\mathbf{v}_1, +\mathbf{v}_1\}$  and consequently  $\mathbf{W}_t$  converges almost surely to one of these points.

and of batch algorithm for subspace estimation. We also investigate the performance of DOA (direction of arrival) estimation based on the adaptive subspace algorithm. Section 4 presents some simulation results and investigates the validity of the asymptotic approach.

## 2 Asymptotic performance analysis

### 2.1 A short review of a general Gaussian approximation result

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

$$\theta_{t+1} = \theta_t + \gamma\phi(\theta_t, \mathbf{x}_t) \quad (11)$$

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  $\theta_*$  in the corresponding decreasing step size algorithm. Then, in real stationary situation we have when  $\gamma \rightarrow 0$  and  $t \rightarrow \infty$ , (where  $\rightarrow_{\mathcal{L}}$  denotes the convergence in distribution):

$$\frac{1}{\sqrt{\gamma}}(\theta_t - \theta_*) \rightarrow_{\mathcal{L}} \mathcal{N}(0, \mathbf{C}_\theta) \quad (12)$$

where  $\mathbf{C}_\theta$  is the unique symmetric solution of the continuous Lyapunov equation:

$$\mathbf{D}\mathbf{C}_\theta + \mathbf{C}_\theta\mathbf{D}^T + \mathbf{\Gamma} = \mathbf{0} \quad (13)$$

and where  $\mathbf{D}$  and  $\mathbf{\Gamma}$  are respectively the derivative of the mean field and the covariance of the field:

$$\mathbf{D} \stackrel{\text{def}}{=} E\left[\frac{\partial\phi}{\partial\theta}(\theta, \mathbf{x}_t)\right]_{\theta=\theta_*} \quad \left([\mathbf{D}]_{i,j} \stackrel{\text{def}}{=} \frac{\partial\phi_j}{\partial\theta_i}\right) \quad (14)$$

$$\mathbf{\Gamma} \stackrel{\text{def}}{=} \sum_{t=-\infty}^{\infty} \text{Cov}[\phi(\theta_*, \mathbf{x}_t), \phi(\theta_*, \mathbf{x}_0)] \quad (15)$$

Thus  $\theta_t$  behaves asymptotically, for  $t$  “large enough” and  $\gamma$  “small enough” like an unbiased Gaussian estimator of  $\theta_*$  with covariance matrix  $\gamma\mathbf{C}_\theta$ . In cases where a closed form for the EVD of  $\mathbf{D}^T$  is available:

$\mathbf{D}^T \mathbf{m}_i = \mu_i \mathbf{m}_i$ , for  $1 \leq i \leq n$ , we have:

$$\mathbf{m}_i^T \mathbf{C}_\theta \mathbf{m}_j = -\frac{\mathbf{m}_i^T \mathbf{\Gamma} \mathbf{m}_j}{\mu_i + \mu_j} \quad (16)$$

Further, if  $\mathbf{D}$  is symmetric, the eigenvectors  $\mathbf{m}_1, \dots, \mathbf{m}_n$  can be chosen orthonormal and equation (16) yields

$$\mathbf{C}_\theta = - \sum_{1 \leq i, j \leq n} \frac{\mathbf{m}_i \mathbf{m}_i^T \mathbf{\Gamma} \mathbf{m}_j \mathbf{m}_j^T}{\mu_i + \mu_j} \quad (17)$$

Further, if  $\mathbf{\Gamma}$  and  $\mathbf{D}$  share the same set of eigenvectors *i.e.* if we have  $\mathbf{\Gamma} \mathbf{m}_i = \nu_i \mathbf{m}_i$  for  $1 \leq i \leq n$ , the above reduces to

$$\mathbf{C}_\theta = \sum_{i=1}^n \frac{\nu_i}{-2\mu_i} \mathbf{m}_i \mathbf{m}_i^T. \quad (18)$$

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  $\mathbf{\Gamma}$  are both diagonal; therefore  $\mathbf{C}_\theta$  will become available in closed form via equation (18).

## 2.2 A short review of an asymptotic property of M-estimates

In this subsection, we recall the asymptotic distribution of M-estimates ([13, theorem 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 if  $\theta_T$  is an isolated solution of

$$\frac{1}{T} \sum_{t=1}^T \phi(\theta, \mathbf{x}_t) = 0 \quad (19)$$

with  $E(\phi(\theta_*, \mathbf{x}_t)) = 0$ , then, if  $\phi$  is sufficiently smooth, the following convergence in distribution result holds when  $T$  tends to  $\infty$ :

$$\sqrt{T}(\theta_T - \theta_*) \rightarrow_{\mathcal{L}} \mathcal{N}(0, \mathbf{C}_\theta) \quad (20)$$

where  $\mathbf{C}_\theta = \mathbf{D}^{-T} \mathbf{\Gamma} \mathbf{D}^{-1}$ ,  $\mathbf{D}$  is defined in (14) and  $\mathbf{\Gamma} = \text{Cov}[\phi(\theta_*, \mathbf{x}_t), \phi(\theta_*, \mathbf{x}_t)]$ . Further, if  $\mathbf{\Gamma}$  and  $\mathbf{D}$  (symmetric) share the same set of eigenvectors, *i.e.* if we have  $\mathbf{\Gamma} \mathbf{m}_i = \nu_i \mathbf{m}_i$  and  $\mathbf{D} \mathbf{m}_i = \mu_i \mathbf{m}_i$  for  $1 \leq i \leq n$ , the matrix  $\mathbf{C}_\theta$  reduces to

$$\mathbf{C}_\theta = \sum_{i=1}^n \frac{\nu_i}{\mu_i^2} \mathbf{m}_i \mathbf{m}_i^T. \quad (21)$$

### 2.3 Local characterization of the field

According to previous section, one needs 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  $\mathbf{P} = \mathbf{\Pi}$ . To proceed, it will be convenient to define the following set of orthonormal Hermitian matrices, (the inner product under consideration is  $(\mathbf{A}, \mathbf{B}) \stackrel{\text{def}}{=} \text{Tr}(\mathbf{A}^H \mathbf{B})$ ):

$$\mathbf{H}_{ab} = \begin{cases} \mathbf{v}_a \mathbf{v}_a^H & a = b \\ \frac{\mathbf{v}_a \mathbf{v}_b^H + \mathbf{v}_b \mathbf{v}_a^H}{\sqrt{2}} & a < b \\ \frac{\mathbf{v}_a \mathbf{v}_b^H - \mathbf{v}_b \mathbf{v}_a^H}{i\sqrt{2}} & a > b. \end{cases} \quad (22)$$

With this definition, a first order approximation of the mean field in the neighborhood of  $\mathbf{\Pi}$  is given by the following lemma:

**Lemma 1** For  $1 \leq a, b \leq n$ ,

$$Ef(\mathbf{\Pi} + \epsilon \mathbf{H}_{ab}, \mathbf{x}_t \mathbf{x}_t^H) = \epsilon \mu_{ab} \mathbf{H}_{ab} + O(\epsilon^2) \quad (23)$$

with

$$\mu_{ab} \stackrel{\text{def}}{=} 2\lambda_a(1 - \pi_a) + 2\lambda_b(1 - \pi_b) - (\lambda_a + \lambda_b)(\pi_a + \pi_b). \quad (24)$$

To characterize the covariance of the field at point  $\mathbf{P}$ , we use the  $\text{Vec}$  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<sup>3</sup> the block matrix whose  $(i, j)$  block element is  $b_{i,j}^* \mathbf{A}$ . We have the classic properties (see for example in the real case [16, Appendix A]):

$$\text{Vec}(\mathbf{ABC}) = (\mathbf{A} \otimes \mathbf{C}^H) \text{Vec}(\mathbf{B}) \quad (25)$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD} \quad (26)$$

$$(\mathbf{A} \otimes \mathbf{B})^H = \mathbf{A}^H \otimes \mathbf{B}^H. \quad (27)$$

Define then the following two matrices:

$$\mathbf{\Pi}^\perp \stackrel{\text{def}}{=} \mathbf{I}_n - \mathbf{\Pi} \quad \mathbf{Q} \stackrel{\text{def}}{=} \mathbf{\Pi}^\perp \otimes \mathbf{\Pi} + \mathbf{\Pi} \otimes \mathbf{\Pi}^\perp. \quad (28)$$

---

<sup>3</sup>This slightly unusual convention makes it easier to deal with complex matrices.



The covariance of the field at point  $\mathbf{P} = \mathbf{\Pi}$  is given by the following lemma:

**Lemma 2** For a circular complex vector  $\mathbf{x}$ ,

$$\text{Cov}(\text{Vec}(f(\mathbf{\Pi}, \mathbf{x}\mathbf{x}^H))) = \mathbf{Q}(\mathbf{R} \otimes \mathbf{R})\mathbf{Q}^H. \quad (29)$$

The eigen-structure of this covariance matrix is characterized as follows:

**Lemma 3** For  $1 \leq a, b \leq n$ ,  $\text{Cov}(\text{Vec}(f(\mathbf{\Pi}, \mathbf{x}\mathbf{x}^H))) \text{Vec}(\mathbf{H}_{ab}) = \nu_{ab} \text{Vec}(\mathbf{H}_{ab})$  with

$$\nu_{ab} \stackrel{\text{def}}{=} (\pi_a - \pi_b)^2 \lambda_a \lambda_b. \quad (30)$$

## 2.4 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 2.1 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  $\mathbf{\Pi}$  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

$$\theta_{ab}(\mathbf{M}) \stackrel{\text{def}}{=} \text{Tr}\{\mathbf{H}_{ab}(\mathbf{M} - \mathbf{\Pi})\}. \quad (31)$$

These are *real* scalars if  $\mathbf{M}$  is Hermitian. Since  $\{\mathbf{H}_{ab} \mid 1 \leq a, b \leq n\}$  is an orthonormal basis for the linear space of  $n \times n$  matrices, the scalars  $\theta_{ab}(\mathbf{M})$  are the coordinates of  $\mathbf{M} - \mathbf{\Pi}$  on this basis. Thus any  $n \times n$  matrix is parameterized by the values of  $\theta_{ab}(\mathbf{M})$  according to:

$$\mathbf{M} = \mathbf{\Pi} + \sum_{1 \leq a, b \leq n} \theta_{ab}(\mathbf{M}) \mathbf{H}_{ab}. \quad (32)$$

Matrices close to  $\mathbf{\Pi}$  are parameterized by small values of these parameters: by definition  $\theta_{ab}(\mathbf{M}) = O(\|\mathbf{M} - \mathbf{\Pi}\|)$  for any pair  $(a, b)$ . The relevance of these parameters is shown by this lemma:

**Lemma 4** *If  $\mathbf{P}$  is an  $n \times n$  rank- $r$  Hermitian matrix, then*

$$\mathbf{P} = \mathbf{\Pi} + \sum_{(a,b) \in P_h} \theta_{ab}(\mathbf{P}) \mathbf{H}_{ab} + O(\|\mathbf{P} - \mathbf{\Pi}\|^2) \quad (33)$$

where  $P_h$  is the complement of  $\{(a,b) \mid r < a, b \leq n\}$ , i.e.  $P_h \stackrel{\text{def}}{=} \{(a,b) \mid 1 \leq a \leq r \text{ or } 1 \leq b \leq r\}$ .

In other words, a rank- $r$  Hermitian matrix lying less than  $\epsilon$  away from  $\mathbf{\Pi}$  (i.e.  $\|\mathbf{P} - \mathbf{\Pi}\| < \epsilon$ ) have negligible (of order  $\epsilon^2$ ) components in the direction of  $\mathbf{H}_{ab}$  for  $r < a, b \leq n$ . Eq. (33) is more compactly expressed by using an  $n^2 \times r(2n - r)$  matrix  $\mathcal{H}$ :

$$\mathcal{H} \stackrel{\text{def}}{=} [\dots, \text{Vec}(\mathbf{H}_{ab}), \dots], \quad (a,b) \in P_h \quad (34)$$

so that the  $r(2n - r) \times 1$  vector  $\theta(\mathbf{P})$  defined by

$$\theta(\mathbf{P}) \stackrel{\text{def}}{=} \mathcal{H}^H \text{Vec}(\mathbf{P} - \mathbf{\Pi}) \quad (35)$$

contains the values of  $\theta_{ab}(\mathbf{P})$  for  $(a,b) \in P_h$  and equation (33) reads, after vectorization

$$\text{Vec}(\mathbf{P}) = \text{Vec}(\mathbf{\Pi}) + \mathcal{H}\theta(\mathbf{P}) + O(\|\mathbf{P} - \mathbf{\Pi}\|^2). \quad (36)$$

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 eq. (33), shows that the matrix set  $\{\mathbf{H}_{ab} \mid (a,b) \in P_h\}$  in fact is a *basis* of the tangent plane to this manifold at point  $\mathbf{\Pi}$ . It follows that, in a neighborhood of  $\mathbf{\Pi}$ , 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  $\mathcal{P}(\theta)$  the unique (for small enough  $\|\theta\|$ ) Hermitian matrix with rank  $r$  such that  $\mathcal{H}^H \text{Vec}(\mathcal{P}(\theta) - \mathbf{\Pi}) = \theta$ . It is not necessary to express explicitly  $\mathcal{P}(\theta)$ : as will turn out, it is sufficient to use the property derived from (36):

$$\text{Vec}(\mathcal{P}(\theta)) = \text{Vec}(\mathbf{\Pi}) + \mathcal{H}\theta + O(\|\theta\|^2). \quad (37)$$

## 2.5 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  $\mathbf{P}_t \rightarrow \theta_t = \mathcal{H}^H \text{Vec}(\mathbf{P}_t - \mathbf{\Pi})$  to the original equation (8).

$$\theta_{t+1} = \theta_t + \gamma_t \phi(\theta_t, \mathbf{x}_t) + O(\gamma_t^2) \quad (38)$$

where function  $\phi$  appears to be

$$\phi(\theta, \mathbf{x}) \stackrel{\text{def}}{=} \mathcal{H}^H \text{Vec}(f(\mathcal{P}(\theta), \mathbf{x}\mathbf{x}^H)). \quad (39)$$

We need to evaluate the derivative matrix  $\mathbf{D}$  of  $E\phi(\theta, \mathbf{x})$  at point  $\theta = 0$  and, since we first consider only the case of independent observations, the covariance matrix  $\mathbf{\Gamma}$  of  $\phi(0, \mathbf{x})$ . With these notations, the results of section 2.3 are recycled as follows.

$$\begin{aligned} E\phi(\theta, \mathbf{x}) &= \mathcal{H}^H \text{Vec} E f(\mathcal{P}(\theta), \mathbf{x}\mathbf{x}^H) \\ &= \mathcal{H}^H \text{Vec} E f\left(\mathbf{\Pi} + \sum \theta_{ab} \mathbf{H}_{ab} + O(\|\theta\|^2), \mathbf{x}\mathbf{x}^H\right) \\ &= \mathcal{H}^H \text{Vec}\left(\sum \theta_{ab} \mu_{ab} \mathbf{H}_{ab} + O(\|\theta\|^2)\right) \\ &= \mathcal{H}^H (\mathcal{H} \mathbf{\Delta}_\mu \theta) + O(\|\theta\|^2) = \mathbf{\Delta}_\mu \theta + O(\|\theta\|^2), \end{aligned}$$

where the above summations are over  $(a, b) \in P_h$ . First equality uses definition (39) and the linearity of the Vec operation; second equality stems from property (36) of the reparameterization; third equality uses lemma 1 and the differentiability of  $f$ ; fourth equality is by definitions (34) and (40); last equality is due to the orthonormality of the basis  $\{\mathbf{H}_{ab}\}$  and allows to conclude that

$$\mathbf{D} \stackrel{\text{def}}{=} \left. \frac{\partial E\phi(\theta, \mathbf{x})}{\partial \theta} \right|_{\theta=0} = \mathbf{\Delta}_\mu \quad \text{with} \quad \mathbf{\Delta}_\mu \stackrel{\text{def}}{=} \text{Diag}(\dots, \mu_{ab}, \dots) \quad (a, b) \in P_h. \quad (40)$$

We proceed with evaluating the covariance of the field at  $\theta = 0$ :

$$\begin{aligned} \text{Cov}(\phi(0, \mathbf{x})) &= \text{Cov}(\mathcal{H}^H \text{Vec}(f(\mathbf{\Pi}, \mathbf{x}\mathbf{x}^H))) = \mathcal{H}^H \text{Cov}(\text{Vec}(f(\mathbf{\Pi}, \mathbf{x}\mathbf{x}^H))) \mathcal{H} \\ &= \mathcal{H}^H \mathcal{H} \mathbf{\Delta}_\nu = \mathbf{\Delta}_\nu \end{aligned}$$

First equality is by definition of  $\phi$ ; second equality is by bilinearity of the Cov operator; third equality is by noting that lemma 3 also reads  $\text{Cov}(\text{Vec}(f(\mathbf{\Pi}, \mathbf{x}\mathbf{x}^H)))\mathcal{H} = \mathcal{H}\mathbf{\Delta}_\nu$  with  $\mathbf{\Delta}_\nu$  defined by (41); last equality is due to the orthonormality of the basis  $\{\mathbf{H}_{ab}\}$  and allows to conclude that for independent observations

$$\mathbf{\Gamma} \stackrel{\text{def}}{=} \text{Cov}(\phi(0, \mathbf{x})) = \mathbf{\Delta}_\nu \quad \text{with} \quad \mathbf{\Delta}_\nu \stackrel{\text{def}}{=} \text{Diag}(\dots, \nu_{ab}, \dots) \quad (a, b) \in P_h. \quad (41)$$

Thus both  $\mathbf{\Gamma}$  and  $\mathbf{D}$  are diagonal matrices. In this case, the Lyapunov equation (13) reduces to  $r(2n - r)$  uncoupled scalar equations. The solution then trivially is

$$\mathbf{C}_\theta = -\frac{1}{2}\mathbf{\Delta}_\nu\mathbf{\Delta}_\mu^{-1}. \quad (42)$$

According to (12),  $\gamma^{-1/2}\theta_t \rightarrow_{\mathcal{L}} \mathcal{N}(\mathbf{0}, -\frac{1}{2}\mathbf{\Delta}_\nu\mathbf{\Delta}_\mu^{-1})$ . By eq. (36), we have  $\text{Vec}(\mathbf{P}_t) = \text{Vec}(\mathbf{\Pi}) + \mathcal{H}\theta_t + O(\|\theta_t\|^2)$ . We conclude that for  $\gamma \rightarrow 0$  and  $t \rightarrow +\infty$ ,

$$\frac{1}{\sqrt{\gamma}}(\text{Vec}(\mathbf{P}_t) - \text{Vec}(\mathbf{\Pi})) \rightarrow_{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{C}_P) \quad \text{with} \quad \mathbf{C}_P = \mathcal{H}\mathbf{C}_\theta\mathcal{H}^H = -\frac{1}{2}\mathcal{H}\mathbf{\Delta}_\nu\mathbf{\Delta}_\mu^{-1}\mathcal{H}^H. \quad (43)$$

Expression (43) of the covariance matrix  $\mathbf{C}_P$  in the asymptotic distribution of  $\text{Vec}(\mathbf{P}_t)$  may be written as an explicit sum:

$$\mathbf{C}_P = \sum_{(a,b) \in P_h} \frac{\nu_{ab}}{-2\mu_{ab}} \text{Vec}(\mathbf{H}_{ab})\text{Vec}(\mathbf{H}_{ab})^H. \quad (44)$$

Definitions (24) of  $\mu_{ab}$  and (30) of  $\nu_{ab}$  show that these quantities are symmetric and also that  $\nu_{ab} = 0$  for  $1 \leq a, b \leq r$ . Using these facts and the easily established identity:

$$\text{Vec}(\mathbf{H}_{ab})\text{Vec}(\mathbf{H}_{ab})^H + \text{Vec}(\mathbf{H}_{ba})\text{Vec}(\mathbf{H}_{ba})^H = \mathbf{\Pi}_a \otimes \mathbf{\Pi}_b + \mathbf{\Pi}_b \otimes \mathbf{\Pi}_a, \quad (45)$$

the expression (44) of the asymptotic covariance matrix of  $\mathbf{P}_t$  is finally rewritten as

$$\mathbf{C}_P = \sum_{1 \leq a < b \leq n} \frac{\lambda_a \lambda_b}{2(\lambda_a - \lambda_b)} (\mathbf{\Pi}_a \otimes \mathbf{\Pi}_b + \mathbf{\Pi}_b \otimes \mathbf{\Pi}_a). \quad (46)$$

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

## 2.6 Analysis.

Several simple MSE characterizations can be derived from the regular structure of the covariance matrix  $\mathbf{C}_P$  as expressed by (44) or (46). A word of caution is nonetheless necessary because the convergence of  $\frac{1}{\sqrt{\gamma}}\text{Vec}(\mathbf{P}_t - \mathbf{\Pi})$  to a limiting Gaussian distribution with covariance matrix  $\mathbf{C}_P$  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

$$\|E(\mathbf{P}_t) - \mathbf{\Pi}\|_{\text{Fro}}^2 = o(\gamma), \quad \text{Cov}(\text{Vec}(\mathbf{P}_t)) = \gamma\mathbf{C}_P + o(\gamma). \quad (47)$$

The numerical experiments presented in section 4 show an excellent prediction of various MSE quantities, confirming this assumption.

According to equation (47), the MSE between  $\mathbf{P}_t$  and  $\mathbf{\Pi}$  is given by the trace of the covariance matrix in the asymptotic distribution of  $\mathbf{P}_t$ . The trace being invariant under orthonormal changes of basis and  $\{\text{Vec}(\mathbf{H}_{ab})\}_{1 \leq a, b \leq n}$  being an orthonormal basis, we obtain from eq. (44) or eq. (46)

$$E\|\mathbf{P}_t - \mathbf{\Pi}\|_{\text{Fro}}^2 = \gamma \sum_{1 \leq a \leq r < b \leq n} \frac{\lambda_a \lambda_b}{\lambda_a - \lambda_b} + o(\gamma). \quad (48)$$

A finer picture is obtained by decomposing the error  $\mathbf{P}_t - \mathbf{\Pi}$  into three terms

$$\mathbf{P}_1 \stackrel{\text{def}}{=} \mathbf{\Pi}(\mathbf{P} - \mathbf{\Pi})\mathbf{\Pi}, \quad \mathbf{P}_2 \stackrel{\text{def}}{=} \mathbf{\Pi}\mathbf{P}\mathbf{\Pi}^\perp + \mathbf{\Pi}^\perp\mathbf{P}\mathbf{\Pi}, \quad \mathbf{P}_3 \stackrel{\text{def}}{=} \mathbf{\Pi}^\perp\mathbf{P}\mathbf{\Pi}^\perp. \quad (49)$$

Using  $\mathbf{I}_n = \mathbf{\Pi} + \mathbf{\Pi}^\perp$ , this is easily seen to be an orthogonal decomposition:

$$\mathbf{P} = \mathbf{\Pi} + \mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3, \quad \|\mathbf{P} - \mathbf{\Pi}\|^2 = \|\mathbf{P}_1\|^2 + \|\mathbf{P}_2\|^2 + \|\mathbf{P}_3\|^2. \quad (50)$$

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

**Lemma 5** *Let  $\mathbf{P}$  be a rank- $r$  Hermitian matrix, and let  $\hat{\mathbf{\Pi}}$  be the orthogonal projection matrix on the range space of  $\mathbf{P}$ . Then*

$$\mathbf{P}_1 = \mathbf{P}^2 - \mathbf{P} + O(\|\mathbf{P} - \mathbf{\Pi}\|^2), \quad (51)$$

$$\mathbf{P}_2 = \hat{\mathbf{\Pi}} - \mathbf{\Pi} + O(\|\mathbf{P} - \mathbf{\Pi}\|^2), \quad (52)$$

$$\mathbf{P}_3 = O(\|\mathbf{P} - \mathbf{\Pi}\|^2). \quad (53)$$

Thanks to this lemma, each term can be given a simple interpretation. The term  $\mathbf{P}_1$  represents the first order of the deviation of  $\mathbf{P}$  from orthogonality since if  $\mathbf{P}$  was an orthogonal projector, we would have  $\mathbf{P}^2 = \mathbf{P}$ . Maybe more strikingly, recalling that  $\mathbf{P} = \mathbf{W}\mathbf{W}^H$ , one finds that

$$\|\mathbf{P}_1\|^2 = \|\mathbf{W}^H\mathbf{W} - \mathbf{I}_r\|^2 + O(\|\mathbf{P} - \mathbf{\Pi}\|^2),$$

since from eqs (98) and (99):  $\|\mathbf{P}_1\|^2 = \|\mathbf{V}(\mathbf{\Lambda} - \mathbf{I}_r)\mathbf{V}^H\|^2 + O(\|\mathbf{P} - \mathbf{\Pi}\|^2)$  where  $\|\mathbf{V}(\mathbf{\Lambda} - \mathbf{I}_r)\mathbf{V}^H\|^2 = \|\mathbf{\Lambda} - \mathbf{I}_r\|^2 = \|\mathbf{U}(\mathbf{\Lambda} - \mathbf{I}_r)\mathbf{U}^H\|^2$  with  $\mathbf{U}$  is the  $r \times r$  unitary matrix of the singular decomposition of  $\mathbf{W}^H\mathbf{W} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$ . According to this lemma, the term  $\mathbf{P}_2$  represents the deviation between the subspace of interest and the one estimated by  $\mathbf{P}$ . Finally, the last term  $\mathbf{P}_3$  is of order  $O(\|\mathbf{P} - \mathbf{\Pi}\|^2)$  because  $\mathbf{P}$  has rank  $r$  (we already expressed this property in lemma 4).

The above decomposition is purely geometric. Statistical results are obtained by combining it with expressions (44) or (46) of the asymptotic covariance matrix  $\mathbf{C}_P$  of  $\mathbf{P}_t$ . In doing so, massive simplifications occur due to orthogonality. This is summarized by

$$\text{Vec}(\mathbf{v}_a\mathbf{v}_b^H)^H(\mathbf{\Pi}_c \otimes \mathbf{\Pi}_d)\text{Vec}(\mathbf{v}_e\mathbf{v}_f^H) = \delta_{c,a}\delta_{c,e}\delta_{d,b}\delta_{d,f} \quad 1 \leq a, b, c, d, e, f \leq n.$$

where  $\delta_{i,j}$  denotes the Kronecker notation:  $\delta_{i,j} = 1$  if  $i = j$  and  $\delta_{i,j} = 0$  elsewhere. Many terms are canceled by these orthogonality relations. The resulting asymptotic variances are

$$E\|\mathbf{P}_{1,t}\|_{\text{Fro}}^2 = E\|\mathbf{W}_t^H\mathbf{W}_t - \mathbf{I}_r\|_{\text{Fro}}^2 + o(\gamma) = O(\gamma), \quad (54)$$

$$E\|\mathbf{P}_{2,t}\|_{\text{Fro}}^2 = E\|\mathbf{P}_t - \mathbf{\Pi}\|_{\text{Fro}}^2 + o(\gamma) = \gamma \sum_{1 \leq a \leq r < b \leq n} \frac{\lambda_a \lambda_b}{\lambda_a - \lambda_b} + o(\gamma), \quad (55)$$

$$E\|\mathbf{P}_{3,t}\|_{\text{Fro}}^2 = o(\gamma). \quad (56)$$

A very striking result is observed here: the deviation of  $\mathbf{W}_t$  from orthonormality, as quantified by  $\|\mathbf{P}_1\|$ , has a stochastic order lower than  $\gamma^{\frac{1}{2}}$ ! 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 that there seems to be no price

to pay for not constraining matrix  $\mathbf{W}$  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 4 that the MSE of orthonormality is, in first approximation, proportional to  $\gamma^2$ .

A frequently encountered situation is when the observation vector  $\mathbf{x}_t$  is the superposition of a signal with a rank-deficient covariance matrix  $\mathbf{\Sigma}$ , corrupted by a spatially incoherent additive noise with covariance matrix  $\sigma^2\mathbf{I}$ , so that  $\mathbf{R} = \mathbf{\Sigma} + \sigma^2\mathbf{I}$ . The projector of interest is  $\mathbf{\Pi}$ , the projector onto the range of  $\mathbf{\Sigma}$  (this space is usually called the *signal subspace*) or, equivalently, the projector  $\mathbf{\Pi}^\perp$  onto its ortho-complement: the *noise subspace*. In this situation, this corollary follows:

**Corollary 1** *If  $\mathbf{R} = \mathbf{\Sigma} + \sigma^2\mathbf{I}$  with  $\text{Rank}(\mathbf{\Sigma}) = r < n$  and  $\mathbf{\Pi}$  is the projector onto the range of  $\mathbf{\Sigma}$ , then expression (46) of  $\mathbf{C}_P$  becomes,*

$$\mathbf{C}_P = \mathbf{\Gamma} \otimes \mathbf{\Pi}^\perp + \mathbf{\Pi}^\perp \otimes \mathbf{\Gamma} \quad \text{with} \quad \mathbf{\Gamma} = \frac{\sigma^2\mathbf{\Pi} + \sigma^4\mathbf{\Sigma}^\#}{2} = \frac{\sigma^2}{2}\mathbf{\Sigma}^\#\mathbf{R}.$$

In addition,  $\text{Tr}(\mathbf{C}_P) = \sigma^2(n - r)\text{Tr}(\mathbf{\Sigma}^\#\mathbf{R})$ .

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

### 3 Further investigations

#### 3.1 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  $\mathbf{P}_t$ , an appropriate orthonormal basis for the real symmetric matrices is:

$$\mathbf{S}_{ab} = \begin{cases} \mathbf{v}_a \mathbf{v}_a^T & a = b \\ \frac{\mathbf{v}_a \mathbf{v}_b^T + \mathbf{v}_b \mathbf{v}_a^T}{\sqrt{2}} & a < b \end{cases} \quad (57)$$

A lemma similar to lemma 4 holds, replacing  $P_h$  by  $P_s \stackrel{\text{def}}{=} \{(a, b) \mid 1 \leq a \leq b \leq n \text{ and } a \leq r\}$ . Keeping the *same* definitions for  $\mu_{ab}$  and  $\nu_{ab}$ , the results of lemmas 1 and 3 become:

$$Ef(\mathbf{\Pi} + \epsilon \mathbf{S}_{ab}, \mathbf{x}_t \mathbf{x}_t^T) = \epsilon \mu_{ab} \mathbf{S}_{ab} + O(\epsilon^2), \quad 1 \leq a \leq b \leq n \quad (58)$$

$$\text{Cov}(\text{Vec}(f(\mathbf{\Pi}, \mathbf{x} \mathbf{x}^T))) \text{Vec}(\mathbf{S}_{ab}) = 2\nu_{ab} \text{Vec}(\mathbf{S}_{ab}), \quad 1 \leq a \leq b \leq n. \quad (59)$$

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

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

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

$$\mathbf{C}_P = \sum_{(a,b) \in P_s} \frac{\nu_{ab}}{-\mu_{ab}} \text{Vec}(\mathbf{S}_{ab}) \text{Vec}(\mathbf{S}_{ab})^T, \quad (61)$$

which is the real counterpart of eq. (44).  $\mathbf{C}_P$  is finally written as:

$$\mathbf{C}_P = \sum_{1 \leq a \leq r < b \leq n} \frac{\lambda_a \lambda_b}{2(\lambda_a - \lambda_b)} (\mathbf{I} + \mathbf{K}) (\mathbf{\Pi}_a \otimes \mathbf{\Pi}_b + \mathbf{\Pi}_b \otimes \mathbf{\Pi}_a). \quad (62)$$

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

### 3.2 Comparisons with other estimation techniques

**PAST algorithm** The expression (62) can be compared to those derived from the asymptotic distribution of the real PAST estimator given by Yang in [9]. As this PAST algorithm converges to  $\mathbf{W}_* = \mathbf{V}_r \mathbf{U}$  (where  $\mathbf{U}$  is an  $r \times r$  arbitrary orthogonal matrix), Yang considers the covariance  $\mathbf{C}_{\tilde{\mathbf{W}}}$  of the limiting distribution of  $\tilde{\mathbf{W}}_t \stackrel{\text{def}}{=} \mathbf{W}_t \mathbf{U}^T$ . The  $nr \times nr$  block matrix  $\mathbf{C}_{\tilde{\mathbf{W}}}$  reads

$$(\mathbf{C}_{\tilde{\mathbf{W}}})_{i,j} = \begin{cases} \mathbf{O} & 1 \leq i \neq j \leq r \\ \sum_{a>r} \frac{\lambda_a}{2(\lambda_i - \lambda_a)} \mathbf{v}_a \mathbf{v}_a^T & 1 \leq i = j \leq r. \end{cases} \quad (63)$$

Applying a continuity theorem (*e.g.*, theorem 6.2a in [19], p. 386) to the differentiable mapping  $\tilde{\mathbf{W}} = (\tilde{\mathbf{w}}_1, \dots, \tilde{\mathbf{w}}_r) \rightarrow \mathbf{P} = \sum_{k=1}^r \tilde{\mathbf{w}}_k \tilde{\mathbf{w}}_k^T$  gives  $\frac{1}{\sqrt{\gamma}} (\text{Vec}(\mathbf{P}_t) - \text{Vec}(\mathbf{\Pi})) \rightarrow_{\mathcal{L}} \mathcal{N}(0, \mathbf{C}_P)$  with

$$\mathbf{C}_P = \sum_{1 \leq a \leq r < b \leq n} \frac{\lambda_b}{2(\lambda_a - \lambda_b)} (\mathbf{I} + \mathbf{K}) (\mathbf{\Pi}_a \otimes \mathbf{\Pi}_b + \mathbf{\Pi}_b \otimes \mathbf{\Pi}_a), \quad (64)$$



where  $\gamma \stackrel{\text{def}}{=} 1 - \lambda$ , 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  $\mathbf{x}_t$  and in the PAST algorithm,  $\gamma$  has no dimension. algorithm.

**Batch estimation technique** Consider the minimization of the sample version of the Yang/Xu criterion (3)

$$J_b(\mathbf{W}) \stackrel{\text{def}}{=} \frac{1}{T} \sum_{t=1}^T \|\mathbf{x}_t - \mathbf{W}\mathbf{W}^H \mathbf{x}_t\|_{\text{Fro}}^2. \quad (65)$$

Let  $\mathbf{W}_T$  be a minimizer of (65). Because all the properties derived from the minimization of (3) also hold in case (65) is minimized, (just replace  $E(\mathbf{x}_t \mathbf{x}_t^H)$  by  $\frac{1}{T} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t^H$ ), the corresponding estimate of the projector  $\mathbf{\Pi}$  has a common value denoted by  $\mathbf{P}_T \stackrel{\text{def}}{=} \mathbf{W}_T \mathbf{W}_T^H$  and coincides with the standard batch estimator. So, since  $\mathbf{W}_T$  is solution of  $\frac{1}{T} \sum_{t=1}^T h(\mathbf{W}, \mathbf{x}_t) = \mathbf{0}$ , with  $h$  defined in (5),  $\mathbf{P}_T$  is solution of

$$\frac{1}{T} \sum_{t=1}^T f(\mathbf{P}, \mathbf{x}_t \mathbf{x}_t^H) = \mathbf{0}, \quad (66)$$

with  $f$  defined in (9). Applying the results on M-estimates recalled in subsection 2.2 and using the common set of eigenvectors  $\text{Vec}(\mathbf{H}_{ab})_{1 \leq a, b \leq n}$  of  $\mathbf{D}$  and  $\mathbf{\Gamma}$ , and the associated eigenvalues (24), (30), we obtain

$$\sqrt{T}(\text{Vec}(\mathbf{P}_T) - \text{Vec}(\mathbf{\Pi})) \rightarrow_{\mathcal{L}} \mathcal{N}(0, \mathbf{C}_P) \quad (67)$$

with

$$\mathbf{C}_P = \sum_{1 \leq a \leq r < b \leq n} \frac{\lambda_a \lambda_b}{(\lambda_a - \lambda_b)^2} (\mathbf{\Pi}_a \otimes \mathbf{\Pi}_b + \mathbf{\Pi}_b \otimes \mathbf{\Pi}_a). \quad (68)$$

**Comparison** The covariance matrices  $\mathbf{C}_P$  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  $\mathbf{C}_P$  share the same eigenbasis  $\mathbf{H}_{ab}$ , in the complex case and  $\mathbf{S}_{ab}$  in the real case. Only the associated eigenvalues are distinct  $\frac{\lambda_a \lambda_b}{2(\lambda_a - \lambda_b)}$  for our LMS algorithm and  $\frac{\lambda_a \lambda_b}{(\lambda_a - \lambda_b)^2}$  for the standard batch

estimation. The covariance matrix  $\mathbf{C}_P$  associated to the real PAST algorithm shares the same eigenbasis,  $\mathbf{S}_{ab}$  but with eigenvalues  $\frac{\lambda_b}{2(\lambda_a - \lambda_b)}$ .

Algorithm	LMS algorithm	batch estimation	PAST algorithm
Eigenvalues	$\frac{\lambda_a \lambda_b}{2(\lambda_a - \lambda_b)}$	$\frac{\lambda_a \lambda_b}{(\lambda_a - \lambda_b)^2}$	$\frac{\lambda_b}{2(\lambda_a - \lambda_b)}$

### 3.3 Application to DOA Tracking

By continuity, the behavior of any differentiable function of the projection matrix  $\mathbf{P}_t$  can be obtained. In this section, the behavior of the DOAs estimated by the MUSIC algorithm [21] applied from  $\mathbf{P}_t$  is derived. Recall first the standard narrow-band array data model:

$$\mathbf{x}_t = \mathbf{E}(\Theta)\mathbf{s}_t + \mathbf{n}_t \quad (69)$$

where  $\mathbf{s}_t$  and  $\mathbf{n}_t$  are mutually uncorrelated zero-mean circular Gaussian stationary processes. Matrix  $\mathbf{E}(\Theta)$  is partitioned into column vectors as  $\mathbf{E}(\Theta) \stackrel{\text{def}}{=} [\mathbf{e}(\theta_1), \dots, \mathbf{e}(\theta_r)]$ . In addition, we assume a spatially white noise:  $E(\mathbf{n}_t \mathbf{n}_t^H) = \sigma^2 \mathbf{I}_n$ , and not fully correlated sources so that  $\mathbf{R}_s = E(\mathbf{s}_t \mathbf{s}_t^H)$  is full rank. The MUSIC estimates of the DOAs  $\theta_k, k = 1, \dots, r$  are determined as the  $r$  deepest minima of the localization function  $\theta$ :

$$\mathbf{e}^H(\theta)(\mathbf{I}_n - \mathbf{P}_t)\mathbf{e}(\theta) = \text{Tr}((\mathbf{I}_n - \mathbf{P}_t)\mathbf{M}(\theta)). \quad (70)$$

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

$$\text{Tr}((\mathbf{I}_n - \mathbf{P})\mathbf{M}'(\hat{\theta}_k)) = 0 \quad (71)$$

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

This is

$$\hat{\theta}_k = \theta_k + \text{Vec}(\mathbf{D}_k)^H \text{Vec}(\mathbf{P} - \mathbf{\Pi}) + o(\|\mathbf{P} - \mathbf{\Pi}\|) \quad \text{with} \quad \mathbf{D}_k = \frac{\mathbf{M}'(\theta_k)}{\text{Tr}\{\mathbf{\Pi}^\perp \mathbf{M}''(\theta_k)\}}. \quad (72)$$

If the derivative matrix  $\mathbf{D}_k$  is non zero, a continuity theorem (*e.g.* theorem 6.2.a [19], p. 386) gives the convergence in distribution of  $\Theta_t = (\theta_1, \dots, \theta_r)_t$ , namely  $\gamma^{-\frac{1}{2}}(\Theta_t - \Theta) \rightarrow_{\mathcal{L}} \mathcal{N}(0, \mathbf{C}_\theta)$  when  $\gamma \rightarrow 0$  and

$t \rightarrow +\infty$  with and asymptotic covariance matrix  $\mathbf{C}_\theta$  given component-wise by

$$(\mathbf{C}_\theta)_{i,j} = \text{Vec}(\mathbf{D}_i)^H \mathbf{C}_P \text{Vec}(\mathbf{D}_j) = \text{Tr} \left( \mathbf{D}_i \mathbf{\Pi}^\perp \mathbf{D}_j \mathbf{\Gamma} + \mathbf{D}_i \mathbf{\Gamma} \mathbf{D}_j \mathbf{\Pi}^\perp \right). \quad (73)$$

where last equation uses corollary 1. Next using  $\text{Tr} \mathbf{\Pi}^\perp \mathbf{M}''(\theta) = 2\mathbf{e}'^H(\theta) \mathbf{\Pi}^\perp \mathbf{e}'(\theta)$ ,  $\mathbf{M}'(\theta) = \mathbf{e}(\theta) \mathbf{e}'^H(\theta) + \mathbf{e}'(\theta) \mathbf{e}^H(\theta)$  where  $\mathbf{e}'(\theta) \stackrel{\text{def}}{=} \frac{d\mathbf{e}(\theta)}{d\theta}$  and  $\mathbf{\Gamma} = \frac{\sigma^2}{2} \sum_{1 \leq a \leq r} \frac{\lambda_a}{\lambda_a - \sigma^2} \mathbf{\Pi}_a$  from lemma 6,  $(\text{Cov}\theta)_{i,j}$  reads:

$$(\text{Cov}\theta)_{i,j} = \gamma(\mathbf{C}_\theta)_{i,j} + o(\gamma) = \frac{\gamma}{2\alpha_i\alpha_j} \sum_{a=1}^r \frac{\lambda_a\sigma^2}{\lambda_a - \sigma^2} \text{Re}[\alpha_{i,j}^* (\mathbf{e}^H(\theta_i) \mathbf{v}_a) (\mathbf{v}_a^H \mathbf{e}(\theta_j))] + o(\gamma) \quad (74)$$

so that in particular

$$\text{Var}\theta_i = \gamma(\mathbf{C}_\theta)_{i,i} + o(\gamma) = \frac{\gamma}{2\alpha_i} \sum_{a=1}^r \frac{\lambda_a\sigma^2}{\lambda_a - \sigma^2} |\mathbf{e}^H(\theta_i) \mathbf{v}_a|^2 + o(\gamma) \quad (75)$$

where  $\alpha_i$  and  $\alpha_{i,j}$  are given by

$$\alpha_i \stackrel{\text{def}}{=} 2\mathbf{e}'^H(\theta_i) \mathbf{\Pi}^\perp \mathbf{e}'(\theta_i) \quad \text{and} \quad \alpha_{i,j} \stackrel{\text{def}}{=} 2\mathbf{e}'^H(\theta_i) \mathbf{\Pi}^\perp \mathbf{e}'(\theta_j) \quad (76)$$

**Comparison to batch estimation.** We note that the expression (74) is quite close to the expression (3.12 [22]) of the covariance of the DOAs estimated by a batch MUSIC algorithm based on  $T$  independent snapshots  $\mathbf{x}_t$ , as given by

$$(\text{Cov}\theta)_{i,j} \sim \frac{2}{T\alpha_i\alpha_j} \sum_{a=1}^r \frac{\lambda_a\sigma^2}{(\lambda_a - \sigma^2)^2} \text{Re}[\alpha_{i,j}^* (\mathbf{e}^H(\theta_i) \mathbf{v}_a) (\mathbf{v}_a^H \mathbf{e}(\theta_j))] \quad (77)$$

where  $\text{Re}()$  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  $\sigma_1^2$ , the power of the signal source  $\mathbf{s}_1$ , with steering vector normalized by  $n$  ( $|\mathbf{e}(\theta)|^2 = n$ ) (75) gives:

$$\text{Var}\theta_1 \sim \gamma \frac{n\sigma_1^2}{2\alpha_1} \left(1 + \frac{1}{n\rho_1}\right) \frac{1}{\rho_1} \quad (78)$$

that we can compare to the batch MUSIC estimation

$$\text{Var}\theta_1 \sim \frac{1}{T} \frac{1}{\alpha_1} \left(1 + \frac{1}{n\rho_1}\right) \frac{1}{\rho_1} \quad (79)$$

where  $\rho_1$  denotes the SNR  $\frac{\sigma_1^2}{\sigma^2}$ . For example the geometrical factor  $\alpha_1$  is given for a linear uniform array of  $n$  sensors by  $\alpha_1 = \frac{n(n^2-1)}{6}$ . 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. So the ratio  $\frac{\gamma\lambda_a\lambda_b}{2(\lambda_a-\lambda_b)} / \frac{\lambda_a\lambda_b}{T(\lambda_a-\lambda_b)^2} = \frac{\gamma T(\lambda_a-\lambda_b)}{2}$ , obtained for the estimated projection matrices, gives the ratio  $\frac{\gamma T n \sigma_1^2}{2}$  for the variances of the estimated DOA by MUSIC algorithms because  $\lambda_a = n\sigma_1^2 + \sigma^2$  and  $\lambda_b = \sigma^2$ . Regarding the speed of convergence of the adaptive MUSIC algorithm, we note that the product: step size  $\times$  eigenvalues of the derivative of the mean field  $= \gamma(\lambda_a - \lambda_b) = \gamma n \sigma_1^2$  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 4).

**Equipowered sources.** In the case of  $r$  equipowered and uncorrelated sources (with normalized steering vectors), the  $r$  variances  $\text{Var}\theta_i, i = 1, \dots, r$  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  $|\theta_2 - \theta_1|$  in the neighborhood of  $\theta_2 - \theta_1 = 0$  and tend to 0 when  $|\theta_2 - \theta_1|$  tends to 0.

From (75), we have after some algebraic manipulations:

$$\text{Var}\theta_i \sim \frac{\gamma\sigma^2}{2\alpha_i} (|\mathbf{e}(\theta_i)|^2 + \sigma^2 \sum_{a=1}^r \frac{|\mathbf{e}^H(\theta_i)\mathbf{v}_a|^2}{\lambda_a - \sigma^2}) \quad (80)$$

But by symmetry, since  $|\mathbf{e}^H(\theta_i)\mathbf{v}_a|^2$  does not depend on  $i$ , we can deduce from

$$(\sigma_1^2 \sum_{k=1}^r \mathbf{e}(\theta_k)\mathbf{e}^H(\theta_k))\mathbf{v}_a = (\lambda_a - \sigma^2)\mathbf{v}_a, a = 1, \dots, r$$

that  $r\sigma_1^2|\mathbf{e}^H(\theta_i)\mathbf{v}_a|^2 = \lambda_a - \sigma^2, a = 1, \dots, r$ . The relation (78) is readily deduced.

For high SNR,  $\lambda_a - \sigma^2 \simeq \lambda_a$  for  $a = 1, \dots, r$ , so that (74) becomes

$$(\text{Cov}\theta)_{i,j} \simeq \frac{\gamma\sigma^2}{2\alpha_i\alpha_j} \text{Re}[\alpha_{i,j}^* \mathbf{e}^H(\theta_i) \mathbf{e}(\theta_j)] \quad (81)$$

and in batch MUSIC estimation, thanks to  $\frac{\lambda_a\sigma^2}{(\lambda_a - \sigma^2)^2} \simeq \frac{\sigma^2}{\lambda_a - \sigma^2}$  for  $a = 1, \dots, r$ , (77) gives

$$(\text{Cov}\theta)_{i,j} \simeq \frac{2\sigma^2}{T\alpha_i\alpha_j} \text{Re}[\alpha_{i,j}^* (\mathbf{R}_s^{-1})_{i,j}] \quad (82)$$

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

$$\text{Var}\theta_i \simeq \frac{\gamma\sigma^2}{2\alpha_i} |\mathbf{e}(\theta_i)|^2 \quad \text{and} \quad \text{Var}\theta_i \simeq \frac{2\sigma^2}{T\alpha_i} (\mathbf{R}_s^{-1})_{i,i} \quad (83)$$

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 [resp. in batch MUSIC] if and only if  $\text{Re}[\alpha_{i,j}^* \mathbf{e}^H(\theta_i) \mathbf{e}(\theta_j)] = 0$  [resp.  $\text{Re}[\alpha_{i,j}^* (\mathbf{R}_s^{-1})_{i,j}] = 0$ ]. This holds in particular for orthogonal steering vectors [resp. for uncorrelated sources].

## 4 Simulations results

We now examine the accuracy of expressions (48) of the mean square error of the projection matrix and of expressions (75) (78) of the misadjustment of the DOAs estimated by the MUSIC algorithm and ‘small terms’ exhibited in section 2.6.

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  $a_1$ . Fig. 1 shows the learning curve averaged (over 100 independent runs) of the square error  $\|\mathbf{P}_t - \mathbf{\Pi}\|_{\text{Fro}}^2$  for  $a_1 = 0.3$  and  $\gamma = 0.005$ . It tends to a value in perfect agreement with the theoretical values predicted by eq. (48). This figure also shows the evolution of ‘small terms’ *i.e.* terms with scale  $o(\gamma)$  as predicted by eq. (47), (54), (55) and (56). We plot the following quantities  $\|E(\mathbf{P}_t) - \mathbf{\Pi}\|_{\text{Fro}}^2$ ,  $E\|\mathbf{W}_t^H \mathbf{W}_t - \mathbf{I}_r\|_{\text{Fro}}^2$ ,  $E\|\mathbf{P}_{1,t}\|_{\text{Fro}}^2$ ,  $E\|\mathbf{P}_{2,t}\|_{\text{Fro}}^2$  and

$E\|\mathbf{P}_{3,t}\|_{\text{Fro}}^2$  whose significance has been discussed in section 2.6. Fig. 2 shows the ratio of the estimated mean square error  $E\|\mathbf{P}_t - \mathbf{\Pi}\|_{\text{Fro}}^2$  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  $\gamma < 0.065$  for which this ratio stays near 1. Fig. 3 shows that the deviation from orthonormality  $d^2(\gamma) \stackrel{\text{def}}{=} E\|\mathbf{W}_t^H \mathbf{W}_t - \mathbf{I}_r\|_{\text{Fro}}^2$  is proportional to  $\gamma^2$  in the domain of validity of the mean square error (48) because  $\log_{10} d^2(\gamma) = \log_{10} c + 2\log_{10} \gamma$  with  $c \simeq 1.210^{-2}$ . Fig. 4 shows in the same way, that the square norm of the bias  $\|E(\mathbf{P}_t) - \mathbf{\Pi}\|_{\text{Fro}}^2$  is proportional to  $\gamma^2$  in the domain of “stability”. Furthermore, its contribution to the MSE  $E\|\mathbf{P}_t - \mathbf{\Pi}\|_{\text{Fro}}^2$  proves that  $E\|\mathbf{P}_t - \mathbf{\Pi}\|_{\text{Fro}}^2 \sim \gamma\text{Tr}(\mathbf{C}_P)$  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  $\gamma = \frac{\gamma_0}{n\sigma_1^2}$  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  $\text{Var}\theta_1 \sim \frac{3\gamma\sigma^2}{n^2} = \frac{3\gamma_0}{n^3\rho_1}$ . The convergence time ( $T \simeq 1000$  in Fig. 5) can be compared with the observation time  $T$  necessary to get the same batch MSE as the sequential asymptotic MSE: eqs. (78) and (79) give the observation time  $T = \frac{2}{\gamma n \sigma_1^2} = \frac{2}{\gamma_0} \simeq 667$ .

Finally, we present the case of 2 uncorrelated and equipowered sources impinging on a linear uniform array of 4 sensors with  $\rho_i = 10$ ,  $\theta_i = \theta_0 \pm \delta$ . 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  $\gamma = 0.003$ ,  $\delta = 0.15(\frac{2\pi}{4})$  then  $\delta = 0.30(\frac{2\pi}{4})$ . 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): 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.

## 5 Conclusion

We have analyzed the performance of an LMS-type algorithm for tracking dominant invariant subspaces. Because this algorithm and other in its class do not converge to the 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 to 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.

## Acknowledgements

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

## Proofs

**Proof of lemma 1.** The field  $f$  in definition (9) being linear in its second argument, the mean field at any point  $\mathbf{P}$  is:

$$Ef(\mathbf{P}, \mathbf{x}_t \mathbf{x}_t^H) = f(\mathbf{P}, E(\mathbf{x}_t \mathbf{x}_t^H)) = f(\mathbf{P}, \mathbf{R}) = \mathbf{P}(2\mathbf{R} - \mathbf{R}\mathbf{P} - \mathbf{P}\mathbf{R}) + (2\mathbf{R} - \mathbf{R}\mathbf{P} - \mathbf{P}\mathbf{R})\mathbf{P}. \quad (84)$$

Using the identities  $\mathbf{\Pi}^2 = \mathbf{\Pi}$  and  $\mathbf{\Pi}\mathbf{R} = \mathbf{R}\mathbf{\Pi} = \mathbf{\Pi}\mathbf{R}\mathbf{\Pi}$  and

$$\mathbf{\Pi}\mathbf{v}_a \mathbf{v}_b^H = \pi_a \mathbf{v}_a \mathbf{v}_b^H \quad \mathbf{v}_a \mathbf{v}_b^H \mathbf{\Pi} = \pi_b \mathbf{v}_a \mathbf{v}_b^H \quad \mathbf{R}\mathbf{v}_a \mathbf{v}_b^H = \lambda_a \mathbf{v}_a \mathbf{v}_b^H \quad \mathbf{v}_a \mathbf{v}_b^H \mathbf{R} = \lambda_b \mathbf{v}_a \mathbf{v}_b^H,$$

a substitution  $\mathbf{P} = \mathbf{\Pi} + \epsilon \mathbf{v}_a \mathbf{v}_b^H$  in (84) yields after simplification,

$$Ef(\mathbf{\Pi} + \epsilon \mathbf{v}_a \mathbf{v}_b^H, \mathbf{x}_t \mathbf{x}_t^H) = \epsilon \mu_{ab} \mathbf{v}_a \mathbf{v}_b^H + O(\epsilon^2) \quad (85)$$

where  $\mu_{ab}$  is defined in eq. (24). The lemma follows by using the symmetry  $\mu_{ab} = \mu_{ba}$ .

Q.E.D.

**Proof of lemma 2.** At point  $\mathbf{P} = \mathbf{\Pi}$ , using  $\mathbf{\Pi}\mathbf{\Pi} = \mathbf{\Pi}$ , definition (9) of the field reduces to:

$$f(\mathbf{\Pi}, \mathbf{xx}^H) = \mathbf{\Pi}^\perp \mathbf{xx}^H \mathbf{\Pi} + \mathbf{\Pi} \mathbf{xx}^H \mathbf{\Pi}^\perp \quad (86)$$

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

$$\text{Vec}(f(\mathbf{\Pi}, \mathbf{xx}^H)) = \mathbf{Q} \text{Vec}(\mathbf{xx}^H). \quad (87)$$

For a circular complex Gaussian vector  $\mathbf{x}$ , we have ([15], p. 336)

$$\text{Cov}(\text{Vec}(\mathbf{xx}^H)) = \mathbf{R} \otimes \mathbf{R}. \quad (88)$$

Combining eqs (87) and (88) establishes the lemma:  $\text{Cov}(\text{Vec}(f(\mathbf{\Pi}, \mathbf{xx}^H))) = \mathbf{Q} \text{Cov}(\text{Vec}(\mathbf{xx}^H)) \mathbf{Q}^H = \mathbf{Q}(\mathbf{R} \otimes \mathbf{R}) \mathbf{Q}^H$ . Q.E.D.

**Proof of lemma 3.** For any pair  $1 \leq a, b \leq n$ , by simple substitution, we find

$$(\mathbf{R} \otimes \mathbf{R}) \text{Vec}(\mathbf{v}_a \mathbf{v}_b^H) = \lambda_a \lambda_b \text{Vec}(\mathbf{v}_a \mathbf{v}_b^H) \quad (89)$$

$$\mathbf{Q} \text{Vec}(\mathbf{v}_a \mathbf{v}_b^H) = (\pi_a - \pi_b)^2 \text{Vec}(\mathbf{v}_a \mathbf{v}_b^H) \quad (90)$$

by using the properties  $\mathbf{R} \mathbf{v}_a \mathbf{v}_b^H \mathbf{R} = \lambda_a \lambda_b \mathbf{v}_a \mathbf{v}_b^H$  and  $\mathbf{\Pi} \mathbf{v}_a \mathbf{v}_b^H \mathbf{\Pi}^\perp = \pi_a (1 - \pi_b) \mathbf{v}_a \mathbf{v}_b^H$  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  $\pi_a = \pi_a^2$ ). The eigen-vectors of  $\text{Cov}(\text{Vec}(f(\mathbf{\Pi}, \mathbf{xx}^H)))$  follow by recalling that  $\mathbf{Q} = \mathbf{Q}^H$  and combining eqs (29), (89) and (90):

$$\text{Cov}(\text{Vec}(f(\mathbf{\Pi}, \mathbf{xx}^H))) \text{Vec}(\mathbf{v}_a \mathbf{v}_b^H) = \nu_{ab} \text{Vec}(\mathbf{v}_a \mathbf{v}_b^H) \quad (91)$$

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

**Proof of lemma 4.** The proof uses the more general lemma 5. Indeed, we have

$$\sum_{r < a, b \leq n} \theta_{ab}^2(\mathbf{P}) = \|\mathbf{\Pi}^\perp (\mathbf{P} - \mathbf{\Pi}) \mathbf{\Pi}^\perp\|^2$$

and, by eq. (53),  $\mathbf{\Pi}^\perp (\mathbf{P} - \mathbf{\Pi}) \mathbf{\Pi}^\perp = O(\|\mathbf{P} - \mathbf{\Pi}\|^2)$ . Thus  $\theta_{ab}(\mathbf{P}) = O(\|\mathbf{P} - \mathbf{\Pi}\|^2)$  for  $r < a, b \leq n$ . Q.E.D.



**Proof of lemma 5.** Denote  $\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$  the eigenvalue decomposition of  $\mathbf{P}$  and recall that  $\hat{\mathbf{\Pi}}$  denotes the orthogonal projector onto the range of  $\mathbf{P}$ , thus  $\hat{\mathbf{\Pi}} = \mathbf{V}\mathbf{V}^H$  and we can write

$$\mathbf{P} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H = \mathbf{V}\mathbf{V}^H + \mathbf{V}(\mathbf{\Lambda} - \mathbf{I}_r)\mathbf{V}^H = \hat{\mathbf{\Pi}} + \mathbf{V}(\mathbf{\Lambda} - \mathbf{I}_r)\mathbf{V}^H. \quad (92)$$

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

$$\mathbf{V}\mathbf{V}^H = \mathbf{\Pi} + \mathbf{\Pi}(\mathbf{P} - \mathbf{\Pi})\mathbf{\Pi}^\perp + \mathbf{\Pi}^\perp(\mathbf{P} - \mathbf{\Pi})\mathbf{\Pi} + O(\|\mathbf{P} - \mathbf{\Pi}\|^2) \quad (93)$$

$$\mathbf{\Lambda} = \mathbf{I}_r + O(\|\mathbf{P} - \mathbf{\Pi}\|) \quad (94)$$

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

$$\mathbf{\Pi}^\perp\mathbf{V} = O(\|\mathbf{P} - \mathbf{\Pi}\|). \quad (95)$$

Then, eq. (53) is established by the sequence:

$$\mathbf{P}_3 \stackrel{\text{def}}{=} \mathbf{\Pi}^\perp\mathbf{P}\mathbf{\Pi}^\perp = \mathbf{\Pi}^\perp\mathbf{V}\mathbf{\Lambda}\mathbf{V}^H\mathbf{\Pi}^\perp = O(\|\mathbf{P} - \mathbf{\Pi}\|^2) \quad (96)$$

where we have used (95) and (94). Then from eq. (93)

$$\mathbf{P}_2 \stackrel{\text{def}}{=} \mathbf{\Pi}\mathbf{P}\mathbf{\Pi}^\perp + \mathbf{\Pi}^\perp\mathbf{P}\mathbf{\Pi} = \mathbf{\Pi}(\mathbf{P} - \mathbf{\Pi})\mathbf{\Pi}^\perp + \mathbf{\Pi}^\perp(\mathbf{P} - \mathbf{\Pi})\mathbf{\Pi} = \hat{\mathbf{\Pi}} - \mathbf{\Pi} + O(\|\mathbf{P} - \mathbf{\Pi}\|^2) \quad (97)$$

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

$$\begin{aligned} \mathbf{P}^2 - \mathbf{P} &= (\mathbf{V}\mathbf{\Lambda}\mathbf{V}^H)^2 - \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H = \mathbf{V}(\mathbf{\Lambda}^2 - \mathbf{\Lambda})\mathbf{V}^H = \mathbf{V}((\mathbf{\Lambda} - \mathbf{I}_r) + (\mathbf{\Lambda} - \mathbf{I}_r)^2)\mathbf{V}^H \\ &= \mathbf{V}(\mathbf{\Lambda} - \mathbf{I}_r)\mathbf{V}^H + O(\|\mathbf{P} - \mathbf{\Pi}\|^2) = \mathbf{P} - \hat{\mathbf{\Pi}} + O(\|\mathbf{P} - \mathbf{\Pi}\|^2) \end{aligned} \quad (98)$$

$$\begin{aligned} &= (\mathbf{\Pi} + \mathbf{P}_1 + \mathbf{P}_2 + \mathbf{P}_3) - (\mathbf{\Pi} + \mathbf{P}_2 + O(\|\mathbf{P} - \mathbf{\Pi}\|^2)) + O(\|\mathbf{P} - \mathbf{\Pi}\|^2) \\ &= \mathbf{P}_1 + O(\|\mathbf{P} - \mathbf{\Pi}\|^2). \end{aligned} \quad (99)$$

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

## References

- [1] E. Moulines, P. Duhamel, J.F. Cardoso, S. Mayrargue, "Subspace methods for blind identification of multichannel FIR filters," *IEEE Trans. on Signal Processing*, vol. 43, no. 2, pp. 516-525, Feb. 1995.
  - [2] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 97-105, Jan. 1995.
  - [3] R.D. DeGroat, E.M. Dowling *et al.*, "Spherical subspace tracking for efficient, high performance adaptive signal processing applications," *Signal Processing, Special Issue on Subspace Methods, Part I: Array Signal Processing and Subspace Computations*, vol. 50, pp. 101-121, 1996.
  - [4] M.G. Larimore, R.J. Calvert, "Convergence studies of Thompson's unbiased adaptative spectral estimator," in *Proc. 14th Asilomar Conf. Circuits Syst. Comput.* Pacific Grove, CA, 1981.
  - [5] J.F. Yang, M. Kaveh, "Adaptive eigenspace algorithms for direction or frequency estimation and tracking," *IEEE Trans. on ASSP*, vol. 36, no. 2, pp. 241-251, Feb. 1988.
  - [6] 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, pp. 69-84, 1985.
  - [7] J.P. Delmas, "Performance analysis of parametrized adaptive eigensubspace algorithms," in *Proc. ICASSP Detroit*, pp. 2056-2059, May 1995.
  - [8] C. Riou, T. Chonavel, P.Y. Cochet, "Adaptive Subspace Estimation - Application to moving sources localization and blind channel identification," *Proc. ICASSP Atlanta*, pp. 1649-1652, May 1996.
  - [9] B. Yang, F. Gersemsky, "Asymptotic distribution of recursive subspace estimators," *Proc. ICASSP Atlanta*, pp. 1764-1767, May 1996.
- projection approximation subspace tracking algorithm," *SignalProcessing* vol. 50, no. 2, pp. 123-136, April 1996.

- [10] P. Baldi, K. Hornik, “Neural networks and principal component analysis: Learning from examples without local minima,” *Neural Networks*, vol. 2, pp. 53-58, 1989.
- [11] L. Xu, “Least mean square error reconstruction principle for self-organizing neural-net,” *Neural Networks*, vol. 6, pp. 627-648, 1993.
- [12] A. Benveniste, M. Métivier, P. Priouret, *Adaptive algorithms and stochastic approximation*, Springer-Verlag, 1990.
- [13] PJ Bickel, C.A.J. Klaassen, Y.A’ Acov Ritov, J.A. Wellner, *Efficient and adaptive estimation for semiparametric models*, Johns Hopkins series in Math. Sciences, 1993.
- [14] J.C. Fort, G. Pags, “Convergence of stochastic algorithms: from the the Kushner-Clark theorem to the Lyapounov functional method,” *Adv. Appl. Prob.* 28, pp. 1072-1094, 1996.
- [15] D.R. Brillinger, *Times series, data analysis and theory*. Expanded Edition, Holden-Day, Inc., 1980.
- [16] T.W. Anderson, *An introduction to multivariate statistical analysis*. Second Edition, Wiley and Sons, 1984.
- [17] T.Kato, *Perturbation Theory for Linear Operators*. Springer Berlin, 1995.
- [18] J.F. Cardoso, E. Moulines, “Asymptotic performance analysis of direction finding algorithms based on fourth-order cumulants,” *IEEE Trans. on Signal Processing*, vol. 43, no. 1, pp. 214-224, Jan. 1995.
- [19] C.R. Rao, *Linear statistical inference and its applications*. Wiley and Sons, 1973.
- [20] M. Kaveh, A.J. Barabell, “The statistical performance of the MUSIC and the Minimum-Norm algorithms in resolving plane waves in noise,” *IEEE Trans. on ASSP*, vol. 34, no. 2, pp. 331-341, April. 1986.
- [21] R.O. Schmidt, “Multiple emitter location and signal parameter estimation,” in *Proc. RADC Spectrum Estimation Workshop*, pp. 243-258, 1979.

- [22] P. Stoica, A. Nehorai, "MUSIC, maximum likelihood and Cramer-Rao bound," *IEEE Trans. on ASSP*, vol. 37, no. 5, pp. 720-741, May 1989.
- [23] J.E. Hudson, *Adaptive array principles*. Stevenage, UK, Peter Peregrinus, 1982.

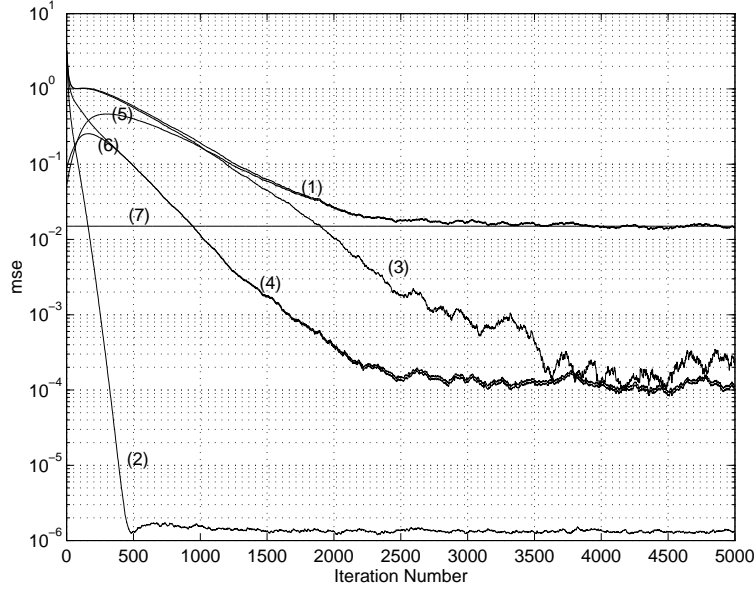


Figure 1: Learning curves of the mean square error  $E\|\mathbf{P}_t - \mathbf{\Pi}\|_{\text{Fro}}^2$  (1), deviation from orthogonality  $E\|\mathbf{W}_t^H \mathbf{W}_t - \mathbf{I}_r\|_{\text{Fro}}^2$  (2), square of bias  $\|E(\mathbf{P}_t) - \mathbf{\Pi}\|_{\text{Fro}}^2$  (3) and variances  $E\|\mathbf{P}_{1,t}\|_{\text{Fro}}^2$  (4),  $E\|\mathbf{P}_{2,t}\|_{\text{Fro}}^2$  (5) and  $E\|\mathbf{P}_{3,t}\|_{\text{Fro}}^2$  (6), averaging 100 independent runs compared to  $\gamma\text{Tr}(\mathbf{C}_P)$  (7) for independent observations  $\mathbf{x}_t$  and  $a_1 = 0.3$  and  $\gamma = 0.005$ .

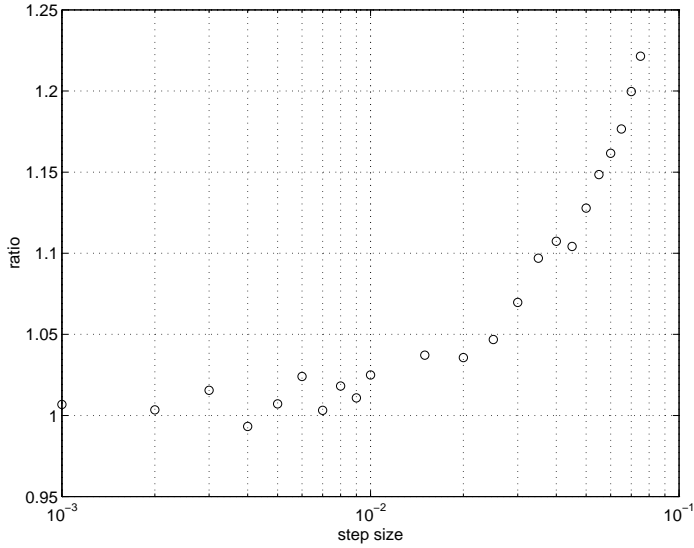


Figure 2: Ratio of the estimated mean square error  $E\|\mathbf{P}_t - \mathbf{\Pi}\|_{\text{Fro}}^2$  by averaging 400 independent runs to the theoretical asymptotic mean square error  $\gamma\text{Tr}(\mathbf{C}_P)$  as a function of  $\gamma$  for independent observations  $\mathbf{x}_t$  and  $a_1 = 0.3$ .

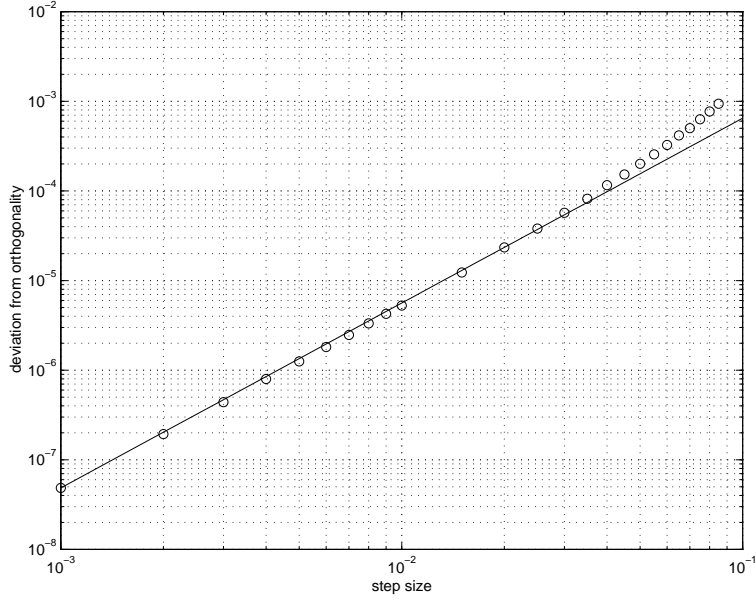


Figure 3: Deviation from orthogonality  $d^2(\gamma) \stackrel{\text{def}}{=} E\|\mathbf{W}_t^H \mathbf{W}_t - \mathbf{I}_r\|_{\text{Fro}}^2$  at “convergence” estimated by averaging 100 independent runs as a function of  $\gamma$  in log-log scales.

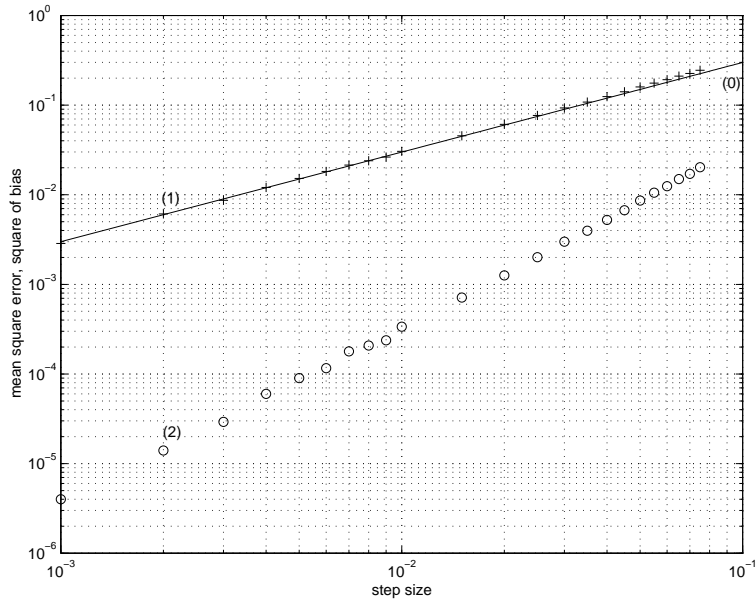


Figure 4: Mean square error  $E\|\mathbf{P}_t - \mathbf{\Pi}\|_{\text{Fro}}^2$  (1), square of bias  $\|E(\mathbf{P}_t) - \mathbf{\Pi}\|_{\text{Fro}}^2$  (2) estimated by averaging 100 independent runs and theoretical asymptotic mean square error  $\gamma \text{Tr}(\mathbf{C}_P)$  (0) as a function of  $\gamma$  in log-log scales.

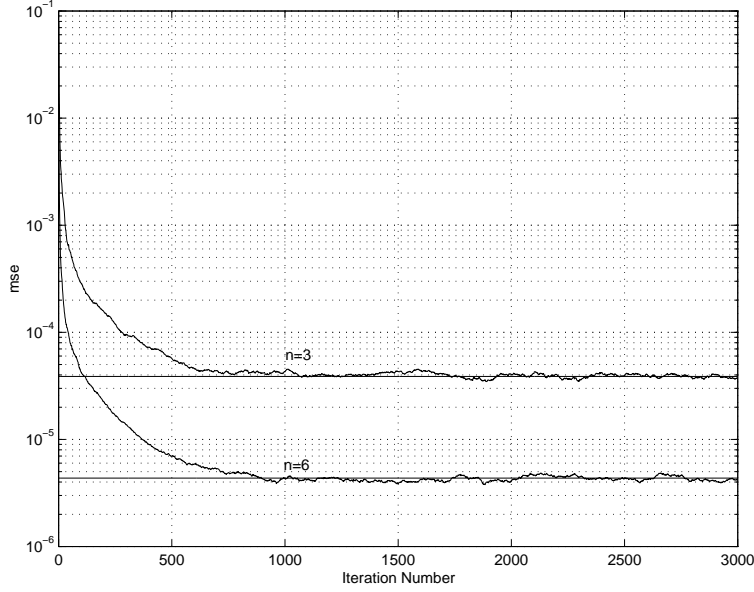


Figure 5: Learning curves of the mean square error of the DOA  $\theta_1$  for one source estimated by MUSIC algorithm where  $\gamma = \frac{\gamma_0}{n\sigma_1^2}$  with  $\gamma_0 = 0.003, \sigma_1^2 = 1, \rho_1 = 10$  for  $n = 3$  and  $n = 6$  compared to the theoretical asymptotic mean square error.

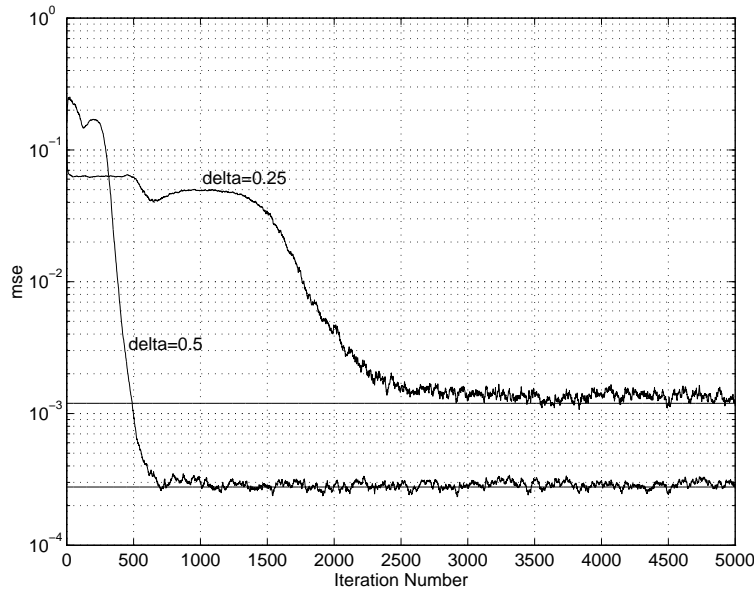


Figure 6: Learning curves of the mean square error of the DOA  $\theta_1$  for two equipowered sources ( $\theta_i = \theta_0 \pm \delta$ ) estimated by MUSIC algorithm where  $\gamma = 0.003, \sigma_1^2 = 1, \rho_i = 10$  for  $\delta = 0.15\text{beamwidth}$  and  $\delta = 0.30\text{beamwidth}$  compared to the theoretical asymptotic mean square error.