



**HAL**  
open science

# Prior Knowledge Optimum Understanding by Means of Oblique Projectors and Their First Order Derivatives

Guillaume Bouleux

► **To cite this version:**

Guillaume Bouleux. Prior Knowledge Optimum Understanding by Means of Oblique Projectors and Their First Order Derivatives. IEEE Signal Processing Letters, 2013, 20, pp.205 - 208. 10.1109/LSP.2013.2238928 . hal-01508680

**HAL Id: hal-01508680**

**<https://hal.science/hal-01508680>**

Submitted on 14 Apr 2017

**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.

# Prior Knowledge Optimum Understanding by means of Oblique Projectors and Their First Order Derivatives

Guillaume Bouleux

## Abstract

Recently, an optimal Prior-knowledge method for DOA estimation has been proposed. This method solely estimates a subset of DOA's accounting known ones. The global idea is to maximize the orthogonality between an estimated signal subspace and noise subspace by constraining the orthogonal noise-made projector to only project onto the desired unknown signal subspace. As it could be surprising, no deflation process is used for. Understanding how it is made possible needs to derive the variance for the DOA estimates. During the derivation, oblique projection operators and their first order derivatives appear and are needed. Those operators show in consequence how the optimal Prior-knowledge criterion can focus only on DOA's of interest and how the optimality is reached.

## I. INTRODUCTION

Trying to sort maximal information from an acquired signal, whatever the quality of sensors, their robustness or their bad (or good) calibration, is the very aim of signal processing. Depending on the point of view, we are forced to accept that prior knowledge is essential for treating those signals as best as possible. The knowledge in question, albeit unconscious knowledge, is at the heart of the modeling process; it could be the statistics of the measured information, the knowledge of a particular model of observations or even sometimes the knowledge of some parameters such as Direction Of Arrival (DOA) [1], frequencies [2] or polarization state of plane waves [3], [4]. When some parameters are assumed to be perfectly known, recent works have been proposed; some of them tackle the underlying parameter estimation problem for DOA by using either subspace deflation [5] or by constraining the roots of polynomial to be fixed, being then referenced as MODE-like (PLEDGE) [6], [7] or Weighted Subspace Fitting based criterion. Even if the derivation of an optimal matrix, which gives the minimum variance estimate, could be more or less intuitive or obvious for these criteria, it appears that a particular manner

Copyright (c) 2012 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to [pubs-permissions@ieee.org](mailto:pubs-permissions@ieee.org).

University of Lyon, University of Saint Etienne, LASPI, F-42334, IUT de Roanne, France.

for reaching it is to use oblique projection operators and their first order derivatives. The contribution of this correspondence is then twofold, an algebraic proof for obtaining the optimal matrix which gives the minimum variance DOA estimate with the prior knowledge assumption and the first order expansion of oblique projection operators, which is considered as the main effort.

## II. MODEL AND ASYMPTOTIC PLEDGE VARIANCE

Consider  $n$  narrowband and far-field plane waves impinging on an Uniform and Linear Array (ULA) composed of  $L$  sensors separated by a half wavelength. Let  $t$  be the index of a sample (“snapshot”) and assume that the total number of available samples is  $N$ , then  $t = 1, \dots, N$ . The one sample response, or equivalently the single-experiment time series model can be parametrized in the following manner

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{n}(t) \quad (1)$$

where  $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1) \dots \mathbf{a}(\theta_n)]$  and where  $\mathbf{a}(\theta_i)$  is the  $i$ -th steering vector defined by

$$\mathbf{a}(\theta_i) = [1 \ e^{j\theta_i} \dots e^{j(L-1)\theta_i}]^T \quad (2)$$

with  $\theta_i$  the spatial pulsation (phase) lying to the set  $\boldsymbol{\theta} = \{\theta_i\}_{i=1}^n$ . Amplitude wave  $\mathbf{x}(t)$  and noise signal  $\mathbf{n}(t)$  are assumed to be stochastic processes, jointly Gaussian with zero-mean, stationary and circular of second order moments

$$\mathbb{E}[\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{P} \quad \text{and} \quad \mathbb{E}[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma^2\mathbf{I} \quad (3)$$

where  $\mathbb{E}$  stands for the mathematical expectation and  $\cdot^H$  for the transpose and conjugate. For the latter use we introduce  $\cdot^T$  as the transpose and  $\cdot^*$  as the conjugate. At least, all the notified hat ( $\hat{\cdot}$ ) quantities correspond to estimated ones. The sample covariance matrix of the data observation is classically given by  $\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^N \mathbf{y}(t)\mathbf{y}^H(t)$  and its eigendecomposition admits the form  $\hat{\mathbf{R}} = \hat{\mathbf{E}}_S \hat{\boldsymbol{\Sigma}}_S \hat{\mathbf{E}}_S^H + \hat{\mathbf{E}}_N \hat{\boldsymbol{\Sigma}}_N \hat{\mathbf{E}}_N^H$  in which  $\hat{\mathbf{E}}_S$  and  $\hat{\mathbf{E}}_N$  are respectively the estimated signal subspace and noise subspace eigenvectors. Finally, the diagonal matrices  $\hat{\boldsymbol{\Sigma}}_S$  and  $\hat{\boldsymbol{\Sigma}}_N$  hold for the corresponding eigenvalues. Suppose now that only  $n_u$  DOA are unknown and then  $n_k$  DOA are known; the set  $\boldsymbol{\theta}$  is therefore composed of disjoint sets  $\boldsymbol{\theta}_u = \{\theta_i\}_{i=1}^{n_u}$  and  $\boldsymbol{\theta}_k = \{\theta_j\}_{j=1}^{n_k}$  DOA. The consequence of this assumption is to disjointly separate all the matrices into two subsets which depend directly or indirectly on parameter  $\boldsymbol{\theta}$ . Then and without loss of generality we have

- $\mathbf{A} = [\mathbf{A}_u \ \mathbf{A}_k]$ , and  $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta}_k, \boldsymbol{\theta}_u)$ ,
- $\hat{\boldsymbol{\Sigma}}_S = \begin{bmatrix} \hat{\boldsymbol{\Sigma}}_{S_u} & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\Sigma}}_{S_k} \end{bmatrix}$ ,

- $\mathbf{P}_A(\boldsymbol{\theta}) = \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H = \mathbf{P}_A(\boldsymbol{\theta}_u, \boldsymbol{\theta}_k)$ ,
- $\mathbf{P}_A^\perp(\boldsymbol{\theta}_u, \boldsymbol{\theta}_k) = \mathbf{I} - \mathbf{P}_A(\boldsymbol{\theta}_u, \boldsymbol{\theta}_k)$ ,
- $\mathbf{P}_A^\perp(\boldsymbol{\theta}_u, \boldsymbol{\theta}_k) = \mathbf{P}_B(\boldsymbol{\theta}_u, \boldsymbol{\theta}_k)$

where  $\mathbf{B}$  is the matrix whose columns span the null space of  $\mathbf{A}$ . The estimation problem is therefore reduced at estimating only the set  $\boldsymbol{\theta}_u$  of unknown DOA. This problem has been already approached and it is optimally solved by the PLEDGE minimizer [6],  $\arg \min_{\boldsymbol{\theta}_u} \text{Trace} \left[ \mathbf{P}_B(\boldsymbol{\theta}_u, \boldsymbol{\theta}_k) \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \right]$ , where  $\mathbf{W}$  is a positive definite weighting matrix. To show now that PLEDGE is optimum, i.e. it reaches the Stochastic Cramèr-Rao Bound (CRB) [6]

$$\mathbf{C}(\boldsymbol{\theta}_u) = \frac{\sigma^2}{2N} \left[ \text{Re} \left\{ \left( \mathbf{D}_u^H \boldsymbol{\Pi}_A^\perp \mathbf{D}_u \right) \odot \left( \mathbf{P}_u^H \mathbf{A}^H \mathbf{R}^{-1} \mathbf{A} \mathbf{P}_u \right)^T \right\} \right]^{-1}$$

where  $\odot$  is the classical Hadamard-shur product; we need the following

**Theorem 1.** *The asymptotic variance for the  $n_u$  unknown estimates admits the expression*

$$\bar{\mathbf{C}} = \frac{\sigma^2}{2N} \left[ \text{Re}(\mathbf{M} \odot (\mathbf{W}_T)^T) \right]^{-1} \left\{ \text{Re} \left[ \mathbf{M} \odot \left( \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \mathbf{E}_S \mathbf{W} \hat{\boldsymbol{\Lambda}} \mathbf{W} \mathbf{E}_S^H \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}^H \mathbf{A}_u^{\dagger H} \right)^T \right] \right\} \left[ \text{Re}(\mathbf{M} \odot (\mathbf{W}_T)^T) \right]^{-1} \quad (4)$$

where  $\mathbf{W}_T = \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \mathbf{E}_S \mathbf{W} \mathbf{E}_S^H \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}^H \mathbf{A}_u^{\dagger H}$ . Here,  $\mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}$  is defined as an oblique projector. This projector projects onto the subspace spanned by the columns of  $\mathbf{A}_u$  along with the subspace spanned by the columns of  $\mathbf{A}_k$  [8]. It has the expression

$$\mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} = \mathbf{A}_u (\mathbf{A}_u^H \mathbf{P}_{\mathbf{A}_k}^\perp \mathbf{A}_u)^{-1} \mathbf{A}_u^H \mathbf{P}_{\mathbf{A}_k}^\perp \quad (5)$$

and the properties

- $\mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} + \mathbf{P}_{\mathbf{A}_k | \mathbf{A}_u} = \mathbf{P}_A$ .
- $\mathbf{P}_{\mathbf{A}_k | \mathbf{A}_u} = \mathbf{P}_{\mathbf{A}_k} \left( \mathbf{I} - \mathbf{A}_u (\mathbf{A}_u^H \mathbf{P}_{\mathbf{A}_k}^\perp \mathbf{A}_u)^{-1} \mathbf{A}_u^H \mathbf{P}_{\mathbf{A}_k}^\perp \right)$ ,

with  $\mathbf{P}_{\mathbf{A}_u}^\perp = \mathbf{I} - \mathbf{A}_u (\mathbf{A}_u^H \mathbf{A}_u)^{-1} \mathbf{A}_u^H$  and identically  $\mathbf{P}_{\mathbf{A}_k} = \mathbf{A}_k (\mathbf{A}_k^H \mathbf{A}_k)^{-1} \mathbf{A}_k^H$ .

We also define  $\mathbf{M} = (\mathbf{D}_u^H \mathbf{P}_A^\perp \mathbf{D}_u)$  with  $\mathbf{D}_u = [\mathbf{d}_1 \dots \mathbf{d}_{n_u}]$ ,  $\mathbf{d}_i = (d\mathbf{a}(\theta_i)/d\theta_i)$  and  $\hat{\boldsymbol{\Lambda}} = \boldsymbol{\Sigma}_S (\boldsymbol{\Sigma}_S - \hat{\sigma}^2 \mathbf{I})^{-2}$  the covariance matrix of the eigenvectors estimation errors  $(\hat{\mathbf{E}}_S - \mathbf{E}_S)$ .

*Proof:* By denoting the function  $\mathbf{V}(\boldsymbol{\theta}_u) = \arg \max_{\boldsymbol{\theta}_u} \text{Trace} \left[ \mathbf{P}_A(\boldsymbol{\theta}_u, \boldsymbol{\theta}_k) \hat{\mathbf{E}}_s \mathbf{W} \hat{\mathbf{E}}_s^H \right]$ , it yields that  $\hat{\boldsymbol{\theta}}_u$  maximizes  $\mathbf{V}(\boldsymbol{\theta}_u)$ . Under good relation of regularity for  $\mathbf{V}(\boldsymbol{\theta}_u)$  and since  $\hat{\boldsymbol{\theta}}_u$  converges to  $\boldsymbol{\theta}_u$  as  $N \rightarrow \infty$  [9], we can give the first order Taylor series expansion of  $\mathbf{V}'$  around the true value  $\boldsymbol{\theta}_u$  by  $\mathbf{0} = \mathbf{V}'(\boldsymbol{\theta}_u) + \mathbf{V}''(\boldsymbol{\theta}_u)(\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u)$ , and deduce obviously the estimation error

$$(\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u) = -\{\mathbf{V}''(\boldsymbol{\theta}_u)\}^{-1} \mathbf{V}'(\boldsymbol{\theta}_u) + o(\mathbf{V}'(\boldsymbol{\theta}_u)) \quad (6)$$

which holds for large sample. Let  $\mathbf{V}'_i$  and  $\mathbf{V}''_{ij}$  stands for the  $i$ -th component of the gradient  $\mathbf{V}'(\boldsymbol{\theta}_u)$  and the  $i, j$ -th component of the hessian  $\mathbf{V}''(\boldsymbol{\theta}_u)$  respectively and let for ease of notation the projection operator  $\mathbf{P}_A(\boldsymbol{\theta}_u, \boldsymbol{\theta}_k)$  correspond to  $\mathbf{P}_A$ . With this, both derivatives have the following expressions  $\mathbf{V}'_i = \text{Trace} \left[ \partial \mathbf{P}_{A_i} \hat{\mathbf{E}}_S \mathbf{W} \hat{\mathbf{E}}_S^H \right]$  and  $\mathbf{V}''_{ij} = \text{Trace} \left[ \partial \mathbf{P}_{A_{ij}} \hat{\mathbf{E}}_S \mathbf{W} \hat{\mathbf{E}}_S^H \right]$  where  $\partial \mathbf{P}_{A_i}$  and  $\partial \mathbf{P}_{A_{ij}}$  are respectively the first and second order derivatives of  $\mathbf{P}_A$ . At least, by invoking the convergence in distribution for the estimates  $\hat{\mathbf{E}}_S$  towards their true value, we obtain that  $\mathbf{V}'_i = \text{Trace} \left[ \partial \mathbf{P}_{A_i} \mathbf{E}_S \mathbf{W} \mathbf{E}_S^H \right] + o(N^{-1/2})$  and  $\mathbf{V}''_{ij} = \text{Trace} \left[ \partial \mathbf{P}_{A_{ij}} \mathbf{E}_S \mathbf{W} \mathbf{E}_S^H \right]$  for  $N \rightarrow \infty$ ; result used in [9] (see the references therein for originate work on).

The key point here is the derivatives of  $\mathbf{P}_A$  with respect to the unknown parameters and only those ones. If giving a theoretical expression for the first order derivative is somehow more or less straight, the second order derivative, is a long way.

#### A. First order derivative of $\mathbf{P}_A$ w.r.t the unknown parameters

Let us start by recalling the projection operator gradient expression of  $\mathbf{P}_A$  stated in [10]. For  $n$  directions of interest, in other words for the set  $\boldsymbol{\theta}$ , the trilinear gradient tensor has the form

$$\frac{\partial \mathbf{P}_A}{\partial \boldsymbol{\theta}} = \mathbf{P}_A^\perp \frac{\partial \mathbf{A}}{\partial \boldsymbol{\theta}} \mathbf{A}^\dagger + \mathbf{A}^{\dagger H} \frac{\partial \mathbf{A}^H}{\partial \boldsymbol{\theta}} \mathbf{P}_A^\perp \quad (7)$$

where each slice of the tensor is associated with a  $\theta_i, i = 1, \dots, n$  direction, see Fig1-(a) for visual illustration. Since only  $n_u$  directions are of interest, (7) can be rewritten  $\frac{\partial \mathbf{P}_A}{\partial \boldsymbol{\theta}_u} = \mathbf{P}_A^\perp \frac{\partial \mathbf{A}}{\partial \boldsymbol{\theta}_u} \mathbf{A}^\dagger + \mathbf{A}^{\dagger H} \frac{\partial \mathbf{A}^H}{\partial \boldsymbol{\theta}_u} \mathbf{P}_A^\perp$ . Owing to  $\mathbf{A} = [\mathbf{A}_u \ \mathbf{A}_k]$  and making use of the block inversion matrix lemma, the pseudo-inverse of  $\mathbf{A}$  admits the expression

$$\begin{aligned} [\mathbf{A}_u \ \mathbf{A}_k]^\dagger &= \begin{bmatrix} \mathbf{A}_u^H \mathbf{A}_u & \mathbf{A}_u^H \mathbf{A}_k \\ \mathbf{A}_k^H \mathbf{A}_u & \mathbf{A}_k^H \mathbf{A}_k \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{A}_u^H \\ \mathbf{A}_k^H \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \\ \mathbf{A}_k^\dagger \mathbf{P}_{\mathbf{A}_k | \mathbf{A}_u} \end{bmatrix}. \end{aligned} \quad (8)$$

The first order derivative of  $\mathbf{P}_A$  in the  $i$ -th direction (the  $i$ -th component of  $\partial \mathbf{P}_A$ ) is next obtained by remarking that only  $n_u$  directions are of interest among  $n$ . Consequently, each slice of the tensor  $\frac{\partial \mathbf{A}}{\partial \boldsymbol{\theta}}$  has the partition which correspond to the known DOA null, i.e. the sub-tensor  $\frac{\partial \mathbf{A}_k}{\partial \boldsymbol{\theta}_u} = \mathbf{0}$ ; this is easily drawn in Fig1-(b). The bottom partition  $\mathbf{A}_k^\dagger \mathbf{P}_{\mathbf{A}_k | \mathbf{A}_u}$  of the pseudo-inverse  $[\mathbf{A}_u \ \mathbf{A}_k]^\dagger$  is in consequence useless

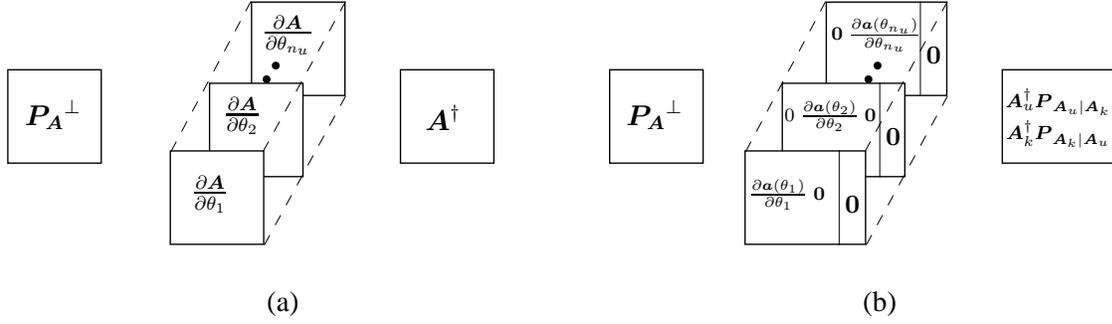


Fig. 1. (a) third order tensor representation for  $\frac{\partial \mathbf{A}}{\partial \theta_{n_u}}$ , (b) third order tensor representation for  $\frac{\partial \mathbf{A}}{\partial \theta_{n_u}}$  with the explicit drawing of  $\frac{\partial \mathbf{A}_k}{\partial \theta_u} = \mathbf{0}$

and we straightforwardly obtain for  $i = 1, \dots, n_u$

$$\begin{aligned} \partial \mathbf{P}_{\mathbf{A}_i} &= \mathbf{P}_{\mathbf{A}}^\perp \frac{\partial \mathbf{A}_u}{\partial \theta_i} \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} + \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}^H \mathbf{A}_u^{\dagger H} \frac{\partial \mathbf{A}_u^H}{\partial \theta_i} \mathbf{P}_{\mathbf{A}}^\perp \\ &\triangleq \underline{\partial \mathbf{P}_{\mathbf{A}_i}} + \left( \underline{\partial \mathbf{P}_{\mathbf{A}_i}} \right)^H. \end{aligned} \quad (9)$$

If we look over the composition of (9), we firstly see that we have two terms transpose and conjugate each other and secondly that thanks to the oblique projector  $\mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}$  action, the subspace associated with the first derivative  $\partial \mathbf{P}_{\mathbf{A}_i}$  only lives onto the subspace associated with the unknown parameters. So, We can here already extrapolate that all the derivatives of  $\mathbf{P}_{\mathbf{A}}$  will lie solely onto the unknown parameters subspace which gives already an idea on the way the PLEDGE criterion acts. But let us pursue the efforts to finish the proof.

### B. Second order derivative of $\mathbf{P}_{\mathbf{A}}$ w.r.t. the unknown parameters

Based on (9),  $\partial \mathbf{P}_{\mathbf{A}_{ij}}$  is basically obtained by deriving the term  $\underline{\partial \mathbf{P}_{\mathbf{A}_i}}$  and by summing next the result to its transpose and conjugate term. The mathematical interest is therefore condensed to  $\underline{\partial \mathbf{P}_{\mathbf{A}_{ij}}}$  the first right hand side term of  $\partial \mathbf{P}_{\mathbf{A}_{ij}} \triangleq \underline{\partial \mathbf{P}_{\mathbf{A}_{ij}}} + \left( \underline{\partial \mathbf{P}_{\mathbf{A}_{ij}}} \right)^H$  with

$$\begin{aligned} \underline{\partial \mathbf{P}_{\mathbf{A}_{ij}}} &= -\mathbf{P}_{\mathbf{A}}^\perp \frac{\partial \mathbf{A}_u}{\partial \theta_j} \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \frac{\partial \mathbf{A}_u}{\partial \theta_i} \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} - \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}^H \mathbf{A}_u^{\dagger H} \frac{\partial \mathbf{A}_u}{\partial \theta_j} \mathbf{P}_{\mathbf{A}}^\perp \frac{\partial \mathbf{A}_u}{\partial \theta_i} \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \\ &+ \mathbf{P}_{\mathbf{A}}^\perp \frac{\partial \mathbf{A}_u}{\partial \theta_i \theta_j} \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} - \mathbf{P}_{\mathbf{A}}^\perp \frac{\partial \mathbf{A}_u}{\partial \theta_i} \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \frac{\partial \mathbf{A}_u}{\partial \theta_j} \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} + \mathbf{P}_{\mathbf{A}}^\perp \frac{\partial \mathbf{A}_u}{\partial \theta_j} \mathbf{A}_u^\dagger \frac{\partial \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}}{\partial \theta_j} \end{aligned} \quad (10)$$

for which the very last element  $\frac{\partial \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}}{\partial \theta_j}$  is so far unknown. Although, to our knowledge there is no theoretical expression for it. The next efforts concern then this element.

The oblique projector is by property idempotent and so is its derivative; as a consequence we have

$$\frac{\partial P_{A_u|A_k}}{\partial \theta_j} = \frac{\partial P_{A_u|A_k}^2}{\partial \theta_j} = \underbrace{\frac{\partial P_{A_u|A_k}}{\partial \theta_j} P_{A_u|A_k}}_I + \underbrace{P_{A_u|A_k} \frac{\partial P_{A_u|A_k}}{\partial \theta_j}}_{II}. \quad (11)$$

Providing an analytical expression for (11) needs obviously to solve (11)-I and (11)-II.

1) *Solving (11)-I*: We make use of

$$\begin{aligned} \frac{\partial (P_{A_u|A_k} A_u)}{\partial \theta_j} &= \frac{\partial P_{A_u|A_k}}{\partial \theta_j} A_u + P_{A_u|A_k} \frac{\partial A_u}{\partial \theta_j} \\ &= \frac{\partial A_u}{\partial \theta_j} \end{aligned} \quad (12)$$

to deduce that

$$\frac{\partial P_{A_u|A_k}}{\partial \theta_j} A_u = (I - P_{A_u|A_k}) \frac{\partial A_u}{\partial \theta_j} \quad (13)$$

and obtain the relation which governs  $\frac{\partial P_{A_u|A_k}}{\partial \theta_j} P_{A_u|A_k}$  by right multiplying (13) by  $(A_u^H P_{A_k}^\perp A_u)^{-1} A_u^H P_{A_k}^\perp$ .

Finally it yields,

$$\frac{\partial P_{A_u|A_k}}{\partial \theta_j} P_{A_u|A_k} = (I - P_{A_u|A_k}) \frac{\partial A_u}{\partial \theta_j} A_u^\dagger P_{A_u|A_k} \quad (14)$$

and (11)-I is fulfilled.

2) *Solving (11)-II*: We utilize here the decomposition of orthogonal projector  $P_A$ , namely  $P_{A_u|A_k} + P_{A_k|A_u} = P_A$ , for having

$$\partial P_{A_j} = \frac{\partial P_{A_u|A_k}}{\partial \theta_j} + \frac{\partial P_{A_k|A_u}}{\partial \theta_j}. \quad (15)$$

We obtain (11)-II plus an annoying term by left multiplying (15) by  $P_{A_u|A_k}$ , the good point now would be that the annoying term vanishes. this is what we are about to prove.

If we consider that the parameters of interest are the unknown ones, they belong so to the set  $\theta_u$ , then the first order derivative  $\frac{\partial P_{A_k|A_u}}{\partial \theta_j}$  can be expressed to as

$$\frac{\partial P_{A_k|A_u}}{\partial \theta_j} = A_k \frac{\partial (A_k^H P_{A_u}^\perp A_k)^{-1}}{\partial \theta_j} A_k^H P_{A_u}^\perp + A_k (A_k^H P_{A_u}^\perp A_k)^{-1} A_k^H \frac{\partial P_{A_u}^\perp}{\partial \theta_j}. \quad (16)$$

Then straightforwardly left multiplying (16) by  $P_{A_u|A_k}$  gives

$$P_{A_u|A_k} \frac{\partial P_{A_k|A_u}}{\partial \theta_j} = 0$$

since  $P_{A_u|A_k} P_A^\perp = 0$  and  $P_{A_u|A_k} \partial P_{A_j}$  clearly reduces to

$$P_{A_u|A_k} \partial P_{A_j} = P_{A_u|A_k} \frac{\partial P_{A_u|A_k}}{\partial \theta_j}. \quad (17)$$

We replace consequently the expression given in (9) inside (17) to obtain finally

$$\begin{aligned}
P_{A_u|A_k} \frac{\partial P_{A_u|A_k}}{\partial \theta_j} &= P_{A_u|A_k} P_A^\perp \frac{\partial A_u}{\partial \theta_j} A_u^\dagger P_{A_u|A_k} \\
&+ P_{A_u|A_k} P_{A_u|A_k}^H A_u^{\dagger H} \frac{\partial A_u^H}{\partial \theta_j} P_A^\perp \\
&= P_{A_u|A_k} P_{A_u|A_k}^H A_u^{\dagger H} \frac{\partial A_u^H}{\partial \theta_j} P_A^\perp
\end{aligned} \tag{18}$$

and we have an expression for (11)-II.

In accordance with (11), the derivative of  $P_{A_u|A_k}$  for the  $j$ -th,  $j = 1, \dots, n_u$  direction is in conclusion given by

$$\frac{\partial P_{A_u|A_k}}{\partial \theta_j} = (I - P_{A_u|A_k}) \frac{\partial A_u}{\partial \theta_j} A_u^\dagger P_{A_u|A_k} + P_{A_u|A_k} P_{A_u|A_k}^H A_u^{\dagger H} \frac{\partial A_u^H}{\partial \theta_j} P_A^\perp.$$

At this stage, it worth nothing to give the expression of  $\frac{\partial P_{A_k|A_u}}{\partial \theta_j}$  obtained by subtracting  $\frac{\partial P_{A_u|A_k}}{\partial \theta_j}$  to  $\partial P_{A_j}$  hence

$$\frac{\partial P_{A_k|A_u}}{\partial \theta_j} = -P_{A_k|A_u} \frac{\partial A_u}{\partial \theta_j} A_u^\dagger P_{A_u|A_k} + (I - P_{A_u|A_k}) P_{A_u|A_k}^H A_u^{\dagger H} \frac{\partial A_u^H}{\partial \theta_j} P_A^\perp.$$

In passing, we note that the conjugate symmetry is lost and that interpreting the hard projector mixing is a real challenge.

### C. Distribution error for the estimates

We embed the previous results inside  $V'(\theta_u)$  and  $V''(\theta_u)$  to have for sufficiently large enough  $N$  and after basic algebraic relations

$$\begin{aligned}
V'_i &= 2 \operatorname{Re} \left\{ \operatorname{Trace} \left[ P_{A_u|A_k}^H A_u^{\dagger H} \frac{\partial A_u^H}{\partial \theta_i} P_A^\perp \hat{E}_S W E_S^H \right] \right\} + o(N^{-1/2}) \\
V''_{ij} &= -2 \operatorname{Re} \left\{ \operatorname{Trace} \left[ \frac{\partial A_u^H}{\partial \theta_j} P_A^\perp \frac{\partial A_u^H}{\partial \theta_i} A_u^\dagger P_{A_u|A_k} E_S W E_S^H P_{A_u|A_k}^H A_u^{\dagger H} \right] \right\}.
\end{aligned}$$

Remind that matrix  $\hat{E}_S^H$  is estimated by the eigen-decomposition of the sample covariance  $\hat{R}$ . The statistics of each column vectors, let say  $\hat{e}_{S_k}$ ,  $k = 1, \dots, n$  depend consequently on those of  $\hat{R}$ . It has been proven, few decades ago, that these statistics are asymptotically jointly Gaussian distributed with

zero means. This result is for example announced in [9, lemma 4] or [11, Lemma 2.1]. As a consequence, the gradient  $\mathbf{V}'(\boldsymbol{\theta}_u)$  is also asymptotically Gaussian, with zero mean and covariance given by

$$\mathbf{H}_{ij} = \frac{2\sigma^2}{N} \text{Re} \left\{ \frac{\partial \mathbf{A}_u^H}{\partial \theta_i} \mathbf{P}_{\mathbf{A}} \perp \frac{\partial \mathbf{A}_u}{\partial \theta_j} \right. \\ \left. \mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \mathbf{E}_S \mathbf{W} \hat{\boldsymbol{\Lambda}} \mathbf{W} \mathbf{E}_S^H \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}^H \mathbf{A}_u^{\dagger H} \right\}$$

with  $\hat{\boldsymbol{\Lambda}} = \boldsymbol{\Sigma}_S (\boldsymbol{\Sigma}_S - \hat{\sigma}^2 \mathbf{I})^{-2}$  the covariance matrix of the eigenvectors estimation errors ( $\hat{\mathbf{E}}_S - \mathbf{E}_S$ ). The asymptotic results of [12] implies that the asymptotic estimation error distribution is therefore

$$(\hat{\boldsymbol{\theta}}_u - \boldsymbol{\theta}_u) \in \mathcal{N}(\mathbf{0}, \bar{\mathbf{C}}), \quad (19)$$

where

$$\bar{\mathbf{C}} = (\mathbf{V}'')^{-1} \mathbf{H} (\mathbf{V}'')^{-1}. \quad (20)$$

The proof is complete by replacing in matrix form  $\mathbf{H}$  and  $\mathbf{V}''$  into (20). ■

### III. THE OPTIMAL WEIGHTED MATRIX

In this section we prove that  $\mathbf{W}_{opt} = \boldsymbol{\Sigma}_S^{-1} (\boldsymbol{\Sigma}_S - \hat{\sigma}^2 \mathbf{I})^2$  yields the asymptotic estimation error distribution to have the minimum variance, i.e. it reaches the Stochastic Cramèr-Rao Bound (CRB).

From  $\mathbf{A} \mathbf{P} \mathbf{A}^H = \mathbf{E}_S (\boldsymbol{\Sigma}_S - \hat{\sigma}^2 \mathbf{I}) \mathbf{E}_S^H$  we deduce that

$$\boldsymbol{\Sigma}_S = \mathbf{E}_S^H \mathbf{A} \mathbf{P} \mathbf{A}^H \mathbf{E}_S + \hat{\sigma}^2 \mathbf{I} \quad (21)$$

and easily write that

$$\mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \mathbf{E}_S (\boldsymbol{\Sigma}_S - \hat{\sigma}^2 \mathbf{I}) \mathbf{E}_S^H \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}^H \mathbf{A}_u^{\dagger H} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{P} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}. \quad (22)$$

We next make use of the relation

$$\boldsymbol{\Sigma}_S^{-1} (\boldsymbol{\Sigma}_S - \hat{\sigma}^2 \mathbf{I})^2 = (\boldsymbol{\Sigma}_S - \hat{\sigma}^2) + \sigma^4 \boldsymbol{\Sigma}_S^{-1} - \hat{\sigma}^2 \mathbf{I} \quad (23)$$

and focus on  $\sigma^4 \boldsymbol{\Sigma}_S^{-1} - \hat{\sigma}^2 \mathbf{I}$  which is equivalent to

$$\mathbf{E}_S (\sigma^4 \boldsymbol{\Sigma}_S^{-1} - \hat{\sigma}^2 \mathbf{I}) \mathbf{E}_S^H = -\mathbf{A} (\hat{\sigma}^{-2} \mathbf{P} \mathbf{A}^H \mathbf{A} + \mathbf{I})^{-1} \mathbf{P} \mathbf{A}^H. \quad (24)$$

Going further, we obtain that

$$\begin{aligned}
\mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \mathbf{E}_S (\sigma^4 \boldsymbol{\Sigma}_S^{-1} - \hat{\sigma}^2 \mathbf{I}) \mathbf{E}_S^H \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}^H \mathbf{A}_k^{\dagger H} &= \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} (\hat{\sigma}^{-2} \mathbf{P} \mathbf{A}^H \mathbf{A} + \mathbf{I})^{-1} \mathbf{P} \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} (\mathbf{I} - \mathbf{P} \mathbf{A}^H (\mathbf{A} \mathbf{P} \mathbf{A}^H + \hat{\sigma}^2 \mathbf{I})^{-1} \mathbf{A}) \mathbf{P} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{P} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{I} & \mathbf{0} \end{bmatrix} \mathbf{P} \mathbf{A}^H \mathbf{R}^{-1} \mathbf{A} \mathbf{P} \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}. \quad (25)
\end{aligned}$$

Then, subtracting (25) to (22) gives the following equality

$$\mathbf{A}_u^\dagger \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k} \mathbf{E}_S \mathbf{W}_{opt} \mathbf{E}_S^H \mathbf{P}_{\mathbf{A}_u | \mathbf{A}_k}^H \mathbf{A}_u^{\dagger H} = \mathbf{P}_u^H \mathbf{A}^H \mathbf{R}^{-1} \mathbf{A} \mathbf{P}_u.$$

Replacing this expression inside (4) shows at least the optimality of the PLEDGE approach with an optimal weighted matrix  $\mathbf{W}_{opt} = \boldsymbol{\Sigma}_S^{-1} (\boldsymbol{\Sigma}_S - \hat{\sigma}^2 \mathbf{I})^2$ .

The reader can notice that the optimal weighted matrix is expressed with respect to  $\boldsymbol{\Sigma}_S$  and not  $\boldsymbol{\Sigma}_{S_u}$  which could be surprising. Although the PLEDGE minimizer focuses only on the unknown DOA, the optimal weighted matrix utilizes the whole signal subspace information. One could think this is obvious since no deflation is employed but things are not that simple. Actually, the sole contribution of the unknown parameters is used but this is hidden behind the way the criterion is constrained by the known parameters. This is the force of the PLEDGE approach - it gives an optimal solution without explicitly deflating. Inspecting the variance for the estimates belonging to  $\boldsymbol{\theta}_u$  shows that a perfect oblique projection operator appears, explaining somehow why the criterion is optimal and why only the unknown parameters contribute to the optimality; even if the weighted matrix uses the whole set of parameters.

#### IV. CONCLUSION

Some works have been proposed for treating the problem of estimating DOA under prior knowledge. The optimal solution to this problem does not employ any deflation process whereas the criterion focuses solely on the parameters of interest. In the present work we have proposed to better understand this by giving the theoretical variance for the estimates of interest. Particularly, the minimum variance is implicitly obtained by the use of oblique projection operators associated with their first order derivatives. The main contribution herein was therefore to propose theoretical expressions for these operators and to show the optimality of the Prior-knowledge criterion.

## REFERENCES

- [1] D. Linebarger, R. DeGroat, E. Dowling, P. Stoica, and G. Fudge, "Incorporating a Priori Information Into MUSIC-Algorithms and Analysis," *Signal Processing*, vol. 46, no. 1, 1995.
- [2] P. Wirfält, G. Bouleux, M. Jansson, and P. Stoica, "Subspace-based Frequency Estimation Utilizing Prior Information," *IEEE Statistical Signal Processing Workshop*, June 2011.
- [3] X.-P. Mao, A.-J. Liu, H.-J. Hou, H. Hong, R. Guo, and W.-B. Deng, "Oblique Projection Polarisation Filtering for Interference Suppression in High-Frequency surface Wave Radar," *IET Journal on Radar, Sonar & Navigation*, vol. 6, pp. 71–80, Feb. 2012.
- [4] B. Cao, Q. Zhang, D. Liang, S. Wen, L. Jin, and Y. Zhang, "Blind Adaptive Polarization Filtering Based on Oblique Projection," *IEEE International Conference on Communications*, pp. 1–5, May 2010.
- [5] R. Boyer and G. Bouleux, "Oblique Projections for Direction-Of-Arrival Estimation with Prior Knowledge," *IEEE Trans. on Signal Processing*, vol. 56, pp. 1374–1387, April 2008.
- [6] G. Bouleux, P. Stoica, and R. Boyer, "An Optimal Prior knowledge-based DOA Estimation Method," *17th European Signal Processing Conference*, Aug. 2009.
- [7] P. Wirfält, M. Jansson, G. Bouleux, and P. Stoica, "Prior Knowledge-based Direction of Arrival Estimation," *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2540–2543, May 2011.
- [8] R. Behrens and L. Scharf, "Signal Processing Applications of Oblique Projection Operators," *IEEE Trans. on Signal Processing*, vol. 42, no. 6, pp. 1413–1424, 1994.
- [9] M. Viberg and B. Ottersten, "Sensor Array Processing Based on Subspace Fitting," *IEEE Trans. on signal processing*, vol. 39, no. 5, 1991.
- [10] G. H. Golub and V. Pereyra, "The differentiation of pseudoinverses and nonlinear least squares problems whose variables separate," *SIAM Journal on Numer. Anal.*, vol. 10, pp. pp 413–432, 1973.
- [11] P. Stoica and K. Sharman, "Maximum Likelihood Methods for Direction Of Arrival Estimation," *IEEE Trans. on Acoustics Speech and Signal Processing*, vol. 38, pp. 1132–1143, 1990.
- [12] Y. Bresler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," *IEEE ASSP Magazine*, vol. 34, pp. 1081–1089, 1986.