

Second-Order Parameter Estimation

Javier Villares, *Student Member, IEEE*, and Gregori Vázquez, *Senior Member, IEEE*

Abstract—This work provides a general framework for the design of second-order blind estimators without adopting any approximation about the observation statistics or the *a priori* distribution of the parameters. The proposed solution is obtained minimizing the estimator variance subject to some constraints on the estimator bias. The resulting optimal estimator is found to depend on the observation fourth-order moments that can be calculated analytically from the known signal model. Unfortunately, in most cases, the performance of this estimator is severely limited by the residual bias inherent to nonlinear estimation problems. To overcome this limitation, the second-order minimum variance unbiased estimator is deduced from the general solution by assuming accurate prior information on the vector of parameters. This small-error approximation is adopted to design iterative estimators or trackers. It is shown that the associated variance constitutes the lower bound for the variance of any unbiased estimator based on the sample covariance matrix.

The paper formulation is then applied to track the angle-of-arrival (AoA) of multiple digitally-modulated sources by means of a uniform linear array. The optimal second-order tracker is compared with the classical maximum likelihood (ML) blind methods that are shown to be quadratic in the observed data as well. Simulations have confirmed that the discrete nature of the transmitted symbols can be exploited to improve considerably the discrimination of near sources in medium-to-high SNR scenarios.

Index Terms—Blind estimation, direction-of-arrival, estimation bounds, non-data-aided, second-order techniques, tracking.

I. INTRODUCTION

THIS paper addresses the problem of blind parameter estimation. This problem arises in many applications in which we have a *linear* system that is known except for a finite number of parameters that we need to estimate. The estimator is required to find out the value of these parameters based on a *noisy observation* of the system output. In some problems, the designer has knowledge or even control over the system input (e.g., radar, sonar and, sometimes, in control and communications [1, Sec. 1]). However, in other problems, the system input is not available and the designer has only a statistical knowledge about the input signal (e.g., speech recognition, image analysis and, also in control and communications [1, Sec. 1]). In that case, the unknown inputs are regarded as *nuisance parameters* [1, Sec. 10.7] that the estimator has to cope with in order to supply *blind* estimates of the parameters of interest.

Manuscript received June 11, 2003; revised June 30, 2004. This work was supported in part by the European Commission (FEDER) and Spanish/Catalan Government under projects TIC2003-05482, TEC2004-04526, and 2001SGR-00268. The material in this paper was presented in part at the IEEE Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, November 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jan C. de Munck.

The authors are with the Department of Signal Theory and Communications, Technical University of Catalonia (UPC), Campus Nord UPC 08034 Barcelona, Spain (e-mail: javi@gps.tsc.upc.es; gregori@gps.tsc.upc.es).

Digital Object Identifier 10.1109/TSP.2005.849175

The maximum likelihood (ML) principle provides a systematic way for deducing the minimum variance unbiased (MVU) estimator when the observation size is sufficiently large [1]. The ML estimator is the one maximizing the *joint* likelihood function $f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}, \mathbf{x})$, where \mathbf{y} is the observed vector, $\boldsymbol{\theta}$ the vector of parameters of interest, and \mathbf{x} the vector of nuisance parameters. To avoid the joint estimation of all the parameters, including the vector of nuisance unknowns \mathbf{x} , the *unconditional* or stochastic ML (UML) criterion models the vector of nuisance parameters as a random vector and maximizes the marginal of the likelihood function conditioned to \mathbf{x} , that is, $f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}) = E_{\mathbf{x}}\{f_{\mathbf{y}/\mathbf{x}}(\mathbf{y}/\mathbf{x}; \boldsymbol{\theta})\}$, where $E_{\mathbf{x}}\{\cdot\}$ stands for the expected value with respect to \mathbf{x} . Unfortunately, the UML estimator is generally unknown because the expectation with respect to \mathbf{x} cannot be solved in a closed form. As shown in the paper, if the signal-to-noise ratio (SNR) tends to zero, the unconditional likelihood function $f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta})$ is *quadratic* in the observation *with independence* of the statistical distribution of the nuisance parameters $f_{\mathbf{x}}(\mathbf{x})$. This fact has motivated the study of second-order estimators since the low SNR approximation turns out to be realistic in many applications (e.g., modern digital communications).

The interest for second-order estimators is also based on the relevance of other second-order blind techniques studied extensively in the literature and described later in this paper. The first one is the *conditional* maximum likelihood (CML) criterion [2]–[5] that models the nuisance parameters as deterministic unknowns, as opposed to the UML approach, and maximizes the *compressed* likelihood function $f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}, \hat{\mathbf{x}}(\boldsymbol{\theta}))$ in which $\hat{\mathbf{x}}(\boldsymbol{\theta})$ stands for the ML estimate of \mathbf{x} (considering that \mathbf{x} is a continuous variable).

The other quadratic ML criterion considered in the paper is the Gaussian maximum likelihood (GML) [2]–[4] that models the nuisance parameters as Gaussian random variables in order to obtain an analytical solution for the expectation in $E_{\mathbf{x}}\{f_{\mathbf{y}/\mathbf{x}}(\mathbf{y}/\mathbf{x}; \boldsymbol{\theta})\}$. Hence, the GML estimator corresponds to the UML solution when the nuisance parameters are Gaussian distributed or the SNR is asymptotically low. However, higher order techniques are generally required to yield efficient estimates at high SNRs [6]. Moreover, it is proved in the paper that the optimal second-order estimator exploits the fourth-order cumulant matrix of the nuisance parameters, improving in that way the GML performance in some circumstances for medium-to-high SNRs, as shown in the simulations section.

Finally, other relevant second-order methods are those based on the eigendecomposition of the sample covariance matrix of \mathbf{y} such as the Pisarenko, MUSIC, and related variants [7], [8]. It is worth noting that although these techniques are not directly addressed in the paper, there is a clear relationship between these methods and the CML approach, as analyzed in [8, Sec. VI].

From this background, the paper offers a convenient framework for the design of optimal second-order blind estimators without resorting to the usual Gaussian assumption on the nuisance parameters. The optimal second-order estimator is obtained minimizing the estimator variance subject to some constraints on the estimator expected value. This optimization is carried out in the Bayesian sense, that is, averaging with respect to the *prior* distribution of the parameters of interest. Unfortunately, in most estimation problems, the performance of the resulting estimator is severely limited by the residual bias as studied in [9]. Consequently, the use of second-order schemes is normally relegated to implement Newton–Raphson or Scoring based *iterative estimators or trackers* [1, eq. (7.50)], [10] in which, after an initial acquisition stage, the estimator or tracker operates in the small-error regime, which is usually referred to as the *steady-state* in tracking applications [10]. Focusing on the small-error approximation, the optimal second-order iterative estimator or tracker is deduced from the general Bayesian solution by applying an extremely informative prior. The resulting small-error solution supplies the MVU estimator for the problem at hand, and its variance constitutes the lower bound for the variance of any unbiased estimator or tracker based on the sample covariance matrix, including the ML-based and MUSIC-like algorithms introduced above.

The paper is organized as follows: In Section II, the signal model is introduced. Next, in Section III, the second-order constraint is motivated based on the quadratic structure of the aforementioned ML-based estimators. In Section IV, the implementation of iterative estimators and the design of trackers in time-varying scenarios is considered and related to the Crámer–Rao bound (CRB) theory. Then, in Section V, the optimal second-order estimator is deduced from scratch by considering an arbitrary prior for the parameters and the *actual* statistics of the nuisance parameters. Focusing on iterative schemes or trackers, the general solution provided in Section V is particularized in Section VI for the small-error case adopting an impulsive prior. The mathematical deduction is accompanied by a short discussion on identifiability and design considerations in multiuser problems. In Section VII, the problem of angle-of-arrival (AoA) tracking is studied in the case of non-Gaussian sources to illustrate the theoretical results of the paper. Finally, the paper is finished with a section of conclusions.

The following notation is used in the paper. Superscripts T, *, H, and # stand for matrix transpose, matrix conjugate, matrix transpose conjugate, and matrix pseudoinverse, respectively. Uppercase and lowercase boldface (or calligraphic) denote matrices and vectors, respectively. $[\mathbf{v}]_i$, $[\mathbf{M}]_i$, and $[\mathbf{M}]_{i,j}$ are used to refer to the i th element of \mathbf{v} , the i th column of \mathbf{M} , and the element in the i th row and j th column of \mathbf{M} , respectively. The symbols $\text{vec}(\mathbf{M})$, $\text{Tr}(\mathbf{M})$, $\det(\mathbf{M})$, $\|\mathbf{v}\|$, and \otimes stand for the column-wise vectorization, the trace, the determinant, the Euclidean norm and the Kronecker product, respectively. Finally, $E_{\mathbf{v}}\{\cdot\}$ is the expected value with respect to the random vector \mathbf{v} and $E\{\cdot\}$ with respect to all the random terms inside the braces.

II. SIGNAL MODEL

This paper deals with the estimation of the vector of real¹ parameters $\boldsymbol{\theta} \in \mathbb{R}^P$ from a given noisy observation $\mathbf{y}(n) \in \mathbb{C}^M$ that is known to follow the linear model

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

where n is the temporal index, $\mathbf{x}(n) \in \mathbb{C}^K$ is the vector of unknown nuisance parameters at time n , $\mathbf{w}(n) \in \mathbb{C}^M$ is the usual Gaussian noise term, and $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{C}^{M \times K}$ is an arbitrary transfer matrix parameterized by the vector of parameters $\boldsymbol{\theta}$. Henceforth, $\mathbf{x}(n)$ and $\mathbf{w}(n)$ are *circular*,² independent, zero-mean, strict-sense stationary random vectors of known statistics. The nuisance parameters are deemed uncorrelated³ and the noise covariance matrix is given by $\mathbf{R}_w \triangleq E\{\mathbf{w}(n)\mathbf{w}^H(n)\}$, where \mathbf{R}_w is a known, full-rank matrix. In the context of Bayesian estimation, the parameter of interest $\boldsymbol{\theta}$ is modeled as a random variable that is statistically independent of $\mathbf{x}(n)$. In the following, the temporal index n will be omitted when we refer to a generic observation for the sake of clarity.

III. CLASSICAL FORMULATION

The estimation problem formulated in the last section has been studied for a long time. The formal solution is provided by the maximum likelihood (ML) principle that yields *asymptotically* ($M \rightarrow \infty$) efficient and unbiased estimators, meaning that the CRB is attained asymptotically [1, Th. 7.1]. In the introduction, the UML estimator was presented as the maximizer of the (unconditional) likelihood function that, using the signal model in Section II, can be expanded as follows:

$$\begin{aligned} f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}) &= E_{\mathbf{x}}\{f_{\mathbf{y}/\mathbf{x}}(\mathbf{y}/\mathbf{x}; \boldsymbol{\theta})\} \\ &= C_1 E_{\mathbf{x}}\{\exp(-(\mathbf{y} - \mathbf{A}\mathbf{x})^H \mathbf{R}_w^{-1}(\mathbf{y} - \mathbf{A}\mathbf{x}))\} \\ &= C_2 E_{\mathbf{x}}\{\exp(2 \text{Re}(\mathbf{x}^H \mathbf{A}^H \mathbf{R}_w^{-1} \mathbf{y}) \\ &\quad - \mathbf{x}^H \mathbf{A}^H \mathbf{R}_w^{-1} \mathbf{A} \mathbf{x})\} \end{aligned} \quad (2)$$

where C_1 and C_2 are irrelevant factors independent of $\boldsymbol{\theta}$, and the dependence on $\boldsymbol{\theta}$ is omitted from $\mathbf{A}(\boldsymbol{\theta})$ for the sake of clarity.

A. Low-SNR Unconditional Maximum Likelihood (Low-SNR UML)

Unfortunately, the expectation with respect to the unknown vector \mathbf{x} usually precludes deducing an analytical expression for the marginal probability density function (pdf) $f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta})$, except if, for instance, \mathbf{x} is normally distributed. However, the Gaussian assumption may be unrealistic in some estimation problems. For example, in digital communications, the transmitted symbols are nuisance parameters that belong to a discrete alphabet, and the Gaussian assumption is clearly incorrect. To overcome this obstacle, the likelihood function is usually evaluated assuming a very low SNR [4]. The low-SNR constitutes a worst-case situation, yielding robust estimators of $\boldsymbol{\theta}$. Thus, if the noise covariance matrix is rewritten as $\mathbf{R}_w = \sigma_w^2 \mathbf{N}$, in order to make ex-

¹If the vector of parameters is complex, $\boldsymbol{\theta}$ is formed stacking the real and imaginary components. Notice that the complex formulation is equivalent but a bit more confusing.

²A random complex vector \mathbf{v} is circular if $E\{\mathbf{v}\mathbf{v}^T\} = 0$.

³Notice that there is no loss of generality as long as the autocorrelation of $\mathbf{x}(n)$ can always be included in $\mathbf{A}(\boldsymbol{\theta})$.

explicit the variance of the noise samples $\sigma_w^2 \triangleq \text{Tr}(\mathbf{R}_w)/M$, the exponential in (2) can be expanded into a Taylor series around $\sigma_w^{-2} = 0$, resulting in

$$f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}) \simeq C_2 E_{\mathbf{x}} \left\{ 1 + \sigma_w^{-2} \chi(\mathbf{y}; \boldsymbol{\theta}, \mathbf{x}) + \frac{\sigma_w^{-4}}{2} \chi^2(\mathbf{y}; \boldsymbol{\theta}, \mathbf{x}) \right\} \quad (3)$$

with $\chi(\mathbf{y}; \boldsymbol{\theta}, \mathbf{x}) \triangleq 2 \text{Re}(\mathbf{x}^H \mathbf{A}^H \mathbf{N}^{-1} \mathbf{y}) - \mathbf{x}^H \mathbf{A}^H \mathbf{N}^{-1} \mathbf{A} \mathbf{x}$ [4]. Assuming now circular, uncorrelated symbols of unit power, the above expectation is evaluated, obtaining

$$\begin{aligned} E_{\mathbf{x}}\{\chi(\mathbf{y}; \boldsymbol{\theta}, \mathbf{x})\} &= -\text{Tr}(\mathbf{A}^H \mathbf{N}^{-1} \mathbf{A}) \\ E_{\mathbf{x}}\{\chi^2(\mathbf{y}; \boldsymbol{\theta}, \mathbf{x})\} &= 2\text{Tr}(\mathbf{N}^{-1} \mathbf{A} \mathbf{A}^H \mathbf{N}^{-1} \hat{\mathbf{R}}) + \zeta(\boldsymbol{\theta}) \end{aligned} \quad (4)$$

where $\zeta(\boldsymbol{\theta}) \triangleq E_{\mathbf{x}}\{(\mathbf{x}^H \mathbf{A}^H \mathbf{N}^{-1} \mathbf{A} \mathbf{x})^2\}$ has not been expanded because it is negligible compared to $E_{\mathbf{x}}\{\chi(\mathbf{y}; \boldsymbol{\theta}, \mathbf{x})\}$, and

$$\hat{\mathbf{R}} \triangleq \mathbf{y} \mathbf{y}^H \quad (5)$$

is the *sample covariance matrix* of the observation \mathbf{y} .

Finally, bearing in mind that $\ln(1+x) \simeq x$ for $x \simeq 0$ and omitting constant terms, the *low-SNR* log-likelihood function is given by

$$\ln f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}) \simeq -\text{Tr}(\mathbf{A}^H \mathbf{R}_w^{-1} \mathbf{A}) + \text{Tr}(\mathbf{R}_w^{-1} \mathbf{A} \mathbf{A}^H \mathbf{R}_w^{-1} \hat{\mathbf{R}}) \quad (6)$$

proving that the sample covariance matrix $\hat{\mathbf{R}}$ is a *sufficient statistic* for the estimation of $\boldsymbol{\theta}$ under the linear model introduced in (1) if the SNR is asymptotically low and the observed vector is sufficiently large. To be more precise, the log-likelihood function in (6) is an affine transformation of the sample covariance matrix with

$$\begin{aligned} b(\boldsymbol{\theta}) &= -\text{Tr}(\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{R}_w^{-1} \mathbf{A}(\boldsymbol{\theta})) \\ \mathcal{M}(\boldsymbol{\theta}) &= \mathbf{R}_w^{-1} \mathbf{A}(\boldsymbol{\theta}) \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{R}_w^{-1} \end{aligned} \quad (7)$$

the independent term, and the kernel of $\ln f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta})$, respectively.

B. Conditional Maximum Likelihood (CML)

Based on the above deduction, efficient estimators become asymptotically ($M \rightarrow \infty$) quadratic in the observation \mathbf{y} for low SNRs. Nonetheless, the low-SNR approximation usually yields a significant variance floor when applied in high SNR scenarios due to the disturbance of the random nuisance parameters. The variability caused by the nuisance parameters is usually referred to as *self-noise* or *pattern-noise* in digital synchronization [10]. In order to design second-order *self-noise free* estimators, the conditional ML (CML) has been proposed as an alternative to the low-SNR UML estimator [2]–[5]. As sketched in the introduction, the CML estimator maximizes the following compressed log-likelihood function:

$$\begin{aligned} \ln f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}, \hat{\mathbf{x}}) &\simeq -(\mathbf{y} - \mathbf{A} \hat{\mathbf{x}})^H \mathbf{R}_w^{-1} (\mathbf{y} - \mathbf{A} \hat{\mathbf{x}}) \\ &= \text{Tr}(\mathbf{R}_w^{-1} \mathbf{A} (\mathbf{A}^H \mathbf{R}_w^{-1} \mathbf{A})^{-1} \mathbf{A}^H \mathbf{R}_w^{-1} \hat{\mathbf{R}}) \end{aligned} \quad (8)$$

where $\hat{\mathbf{x}} = (\mathbf{A}^H \mathbf{R}_w^{-1} \mathbf{A})^{-1} \mathbf{A}^H \mathbf{R}_w^{-1} \mathbf{y}$ is the ML estimate of the nuisance parameters, assuming that $\mathbf{A}(\boldsymbol{\theta})$ is a tall matrix, i.e., $M > K$ in (1). This condition is necessary to have self-

noise free estimates. Therefore, the CML cost function is also quadratic in the observation \mathbf{y} with $b(\boldsymbol{\theta}) = 0$ and

$$\mathcal{M}(\boldsymbol{\theta}) = \mathbf{R}_w^{-1} \mathbf{A}(\boldsymbol{\theta}) (\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{R}_w^{-1} \mathbf{A}(\boldsymbol{\theta}))^{-1} \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{R}_w^{-1} \quad (9)$$

which is the kernel of $\ln f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}, \hat{\mathbf{x}})$.

The CML solution in (8) is actually projecting the *whitened* observation $\mathbf{R}_w^{-1/2} \mathbf{y}$ onto the orthogonal subspace generated by the columns of $\mathbf{R}_w^{-1/2} \mathbf{A}(\boldsymbol{\theta})$. In that sense, the CML solution (8) coincides with the low-SNR UML solution obtained in (6) if the columns of $\mathbf{R}_w^{-1/2} \mathbf{A}(\boldsymbol{\theta})$ are orthogonal, i.e., $\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{N}^{-1} \mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}$, where \mathbf{I} is the identity matrix. Otherwise, the CML estimator suffers from noise-enhancement precluding its utilization in noisy scenarios [4].

C. Gaussian Maximum Likelihood (GML)

It can be shown that the low-SNR UML and CML solutions are independent of the actual SNR. When the SNR is known, the estimator can be designed specifically for a given SNR so that a tradeoff is established between the contributions of noise and self-noise. Concretely, the estimator is demanded to converge to the low-SNR ML estimator when $\sigma_w^2 \rightarrow \infty$ and to the CML when $\sigma_w^2 \rightarrow 0$. In order to guarantee this asymptotic behavior, the Gaussian assumption about the nuisance parameters distribution is largely adopted, yielding the well-known Gaussian ML (GML) estimator [2]–[4]. Once again, the GML solution is an affine transformation of the sample covariance matrix that, omitting constant terms, is given by

$$\begin{aligned} \ln f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}) &= \ln E_{\mathbf{x}}\{f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}; \mathbf{x})\} \\ &= -\ln \det(\mathbf{R}(\boldsymbol{\theta})) - \text{Tr}(\mathbf{R}^{-1}(\boldsymbol{\theta}) \hat{\mathbf{R}}) \end{aligned} \quad (10)$$

with

$$\mathbf{R}(\boldsymbol{\theta}) \triangleq E\{\hat{\mathbf{R}}\} = E\{\mathbf{y} \mathbf{y}^H\} = \mathbf{A}(\boldsymbol{\theta}) \mathbf{A}^H(\boldsymbol{\theta}) + \mathbf{R}_w \quad (11)$$

which is the covariance matrix of \mathbf{y} .

Therefore, bearing in mind that $\ln \det(\cdot) = \text{Tr} \ln(\cdot)$, it is found that

$$\begin{aligned} b(\boldsymbol{\theta}) &= -\text{Tr}(\ln \mathbf{R}(\boldsymbol{\theta})) \\ \mathcal{M}(\boldsymbol{\theta}) &= -\mathbf{R}^{-1}(\boldsymbol{\theta}) = -(\mathbf{A}(\boldsymbol{\theta}) \mathbf{A}^H(\boldsymbol{\theta}) + \mathbf{R}_w)^{-1} \end{aligned} \quad (12)$$

are the independent term and the kernel of the GML likelihood function (10), respectively.

IV. ITERATIVE ML ESTIMATION AND TRACKING

Generally, the ML-based estimators presented in the last section do not admit an analytical solution, and the iterative maximization of the associated log-likelihood functions is required. Focusing on the UML estimator, which has been deduced for low SNRs in (6) and for Gaussian nuisance parameters in (10), the CRB theorem formulated in [1, Th. 3.2] guarantees that under minor regularity constraints, the following scoring algorithm is *asymptotically* ($M \rightarrow \infty$) efficiency [1, (7.50)]:

$$\hat{\boldsymbol{\theta}}_{k+1} = \hat{\boldsymbol{\theta}}_k + \mathbf{J}^\#(\hat{\boldsymbol{\theta}}_k) \Delta(\hat{\boldsymbol{\theta}}_k) \quad (13)$$

assuming that the initial guess $\hat{\boldsymbol{\theta}}_0$ is sufficiently close to the global maximum in the case of a multimodal problem. In the last equation the integer $k = 0, 1, 2, \dots$ is the iteration index, $\Delta(\boldsymbol{\theta}) \triangleq \partial \ln f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ is the log-likelihood gradient, and

$\mathbf{J}(\boldsymbol{\theta}) \triangleq E\{\boldsymbol{\Delta}(\boldsymbol{\theta})\boldsymbol{\Delta}^H(\boldsymbol{\theta})\}$ is the associated Fisher Information Matrix (FIM). The Moore–Penrose pseudo-inverse $(\cdot)^\#$ is introduced in (13) in order to cover those ill-conditioned cases in which the FIM is singular [11]. Notice that the convergence of (13) is not guaranteed unless the log-likelihood function is quadratic near its maximum. Otherwise, the above scoring method might fluctuate around this point. Nonetheless, $\ln f_{\mathbf{y}}(\mathbf{y}; \boldsymbol{\theta})$ becomes typically quadratic near its maximum when $M \rightarrow \infty$ [1, p. 187].

The above scoring algorithm can be generalized to estimate a given transformation of the parameter [11], [1, Sec. 3.8]. In that case, we obtain that

$$\hat{\boldsymbol{\alpha}}_{k+1} = \mathbf{g}(\hat{\boldsymbol{\theta}}_k) + \mathbf{D}_{\mathbf{g}}(\hat{\boldsymbol{\theta}}_k)\mathbf{J}^\#(\hat{\boldsymbol{\theta}}_k)\boldsymbol{\Delta}(\hat{\boldsymbol{\theta}}_k) \quad (14)$$

where $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$ is the referred transformation, and $\mathbf{D}_{\mathbf{g}}(\boldsymbol{\theta}) \triangleq \partial \mathbf{g}(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}^T$ is the Jacobian of $\mathbf{g}(\boldsymbol{\theta})$. Notice that for the studied low-SNR scenario (6) and/or if the nuisance parameters are Gaussian (10), the above UML scoring method is linear in the sample covariance matrix $\hat{\mathbf{R}} \triangleq \mathbf{y}\mathbf{y}^H$.

After convergence, it is held that $\boldsymbol{\theta}_k \simeq \boldsymbol{\theta}_o$, where $\boldsymbol{\theta}_o$ is the true parameter (*small-error assumption*), and consequently, the covariance matrix of the estimation error is given by

$$\begin{aligned} \mathbf{B}(\boldsymbol{\theta}_o) &\triangleq E\{(\hat{\boldsymbol{\alpha}}_k - \mathbf{g}(\boldsymbol{\theta}_o))(\hat{\boldsymbol{\alpha}}_k - \mathbf{g}(\boldsymbol{\theta}_o))^H\} \\ &= \mathbf{D}_{\mathbf{g}}(\boldsymbol{\theta}_o)\mathbf{J}^\#(\boldsymbol{\theta}_o)\mathbf{D}_{\mathbf{g}}^H(\boldsymbol{\theta}_o) \end{aligned} \quad (15)$$

that coincides with the generalized CRB formulated in [11] [1, Sec. 3.8]. Notice again that (15) is verified if and only if the observation is large enough ($M \rightarrow \infty$), or, alternatively, the SNR goes to infinity. Indeed, both conditions ensure that the estimator converges to the true parameter and not to other spurious maxima (the so-called outliers or large-errors [12]). The occurrence of outliers in nonlinear estimation problems establishes an SNR threshold below which the ML estimator variance departs from the one predicted by the CRB [12]. This implies that the observation size must be large enough to work above the threshold when the SNR is very low (Section III-A).

The above recursions can be modified to derive adaptive *closed-loop* schemes in which the observation $\mathbf{y}(n)$ is updated at each iterate in order to *track* the parameter evolution in nonstationary problems. From (14), the proposed ML tracker is as follows:

$$\hat{\boldsymbol{\alpha}}_{n+1} = \mathbf{g}(\hat{\boldsymbol{\theta}}_n) + \boldsymbol{\Lambda}_\mu \mathbf{D}_{\mathbf{g}}(\hat{\boldsymbol{\theta}}_n)\mathbf{J}^\#(\hat{\boldsymbol{\theta}}_n)\boldsymbol{\Delta}_n(\hat{\boldsymbol{\theta}}_n) \quad (16)$$

where n is the temporal index introduced in (1) and $\boldsymbol{\Lambda}_\mu$ a diagonal matrix containing the forgetting factor or step-size of each parameter. Thus, $\boldsymbol{\Lambda}_\mu$ sets the *normalized* (noise equivalent) loop bandwidth (B_n) associated with each parameter [10, Sec. 3.5.5]. The loop bandwidth determines the maximum variability of the parameters that the closed loop is able to track as well as the closed-loop effective observation time that, approximately, is equal to $N \simeq 0.5/B_n$ instants ([10, (3.5.83)]).

Notice that the log-likelihood gradient $\boldsymbol{\Delta}_n(\boldsymbol{\theta})$ is now time-variant because it depends on the current observation $\mathbf{y}(n)$, whereas $\mathbf{J}(\boldsymbol{\theta})$ is fixed, assuming that both $\mathbf{x}(n)$ and $\mathbf{w}(n)$ are stationary random variables (see Section II).

In (16), the function $\boldsymbol{\alpha}_{n+1} = \mathbf{g}(\boldsymbol{\theta}_n) = (\mathbf{h} \cdot \mathbf{f})(\boldsymbol{\theta}_n)$ might be the composition of any transformation of the parameter $\mathbf{h}(\boldsymbol{\theta})$

and the assumed model for the parameter dynamics, i.e., $\boldsymbol{\theta}_{n+1} = \mathbf{f}(\boldsymbol{\theta}_n)$.

Regarding the tracker performance, the CRB can be attained in the steady-state as $B_n \rightarrow 0$ ($N \rightarrow \infty$) in some estimation problems (e.g., timing synchronization [10, Sec. 7.4.3]). Nonetheless, in other estimation problems, the observation size must be augmented ($M \rightarrow \infty$) to attain the CRB (e.g., parametric spectral estimation in time-series analysis or the dual problem of angle-of-arrival estimation in array signal processing [8, Sec. VIII]).

V. GENERALIZED SECOND-ORDER ESTIMATORS

Thus far, second-order estimation has been motivated starting from the ML principle. The optimality of quadratic ML iterative schemes has been claimed when the SNR approaches zero or the nuisance parameters are Gaussian. Additionally, the quadratic CML and GML solutions have been presented as an alternative to remove the self-noise at high SNRs. In this section, the use of second-order schemes is extended beyond the small-error and Gaussian assumptions accepted previously.

Following the structure indicated in (6), (8), and (10), the generic expression of *any* second-order estimator of the parameter θ_p is given by

$$\hat{\theta}_p = b_p + \text{Tr}(\mathcal{M}_p^H \hat{\mathbf{R}}) = b_p + \mathbf{m}_p^H \hat{\mathbf{r}} \quad (17)$$

where \mathcal{M}_p and b_p are the estimator coefficients corresponding to the p th parameter of $\boldsymbol{\theta}$, and $\hat{\mathbf{r}}$ and \mathbf{m}_p are the column-wise stacking of $\hat{\mathbf{R}}$ and \mathcal{M}_p , that is, $\hat{\mathbf{r}} \triangleq \text{vec}(\hat{\mathbf{R}})$ and $\mathbf{m}_p \triangleq \text{vec}(\mathcal{M}_p)$. This vectorization allows the compacting of the multivariate second-order estimator as follows:

$$\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, \dots, \hat{\theta}_P]^T = \mathbf{b} + \mathbf{M}^H \hat{\mathbf{r}}. \quad (18)$$

The objective in this section is to determine the matrix \mathbf{M} and the vector \mathbf{b} that minimize the estimator variance subject to some design constraints on the estimator mean response:

$$\boldsymbol{\psi}(\boldsymbol{\theta}) \triangleq E\{\hat{\boldsymbol{\theta}}\} = \mathbf{b} + \mathbf{M}^H \mathbf{r}(\boldsymbol{\theta}) \quad (19)$$

where $\mathbf{r}(\boldsymbol{\theta}) \triangleq E\{\hat{\mathbf{r}}\}$ vectorizes (in a column) the covariance matrix of the observed vector (11).

Let us now consider that the estimator is provided with some *side-information* on the value of the unknown parameters. Following a Bayesian approach, this side information will consist of the *a priori* pdf of $\boldsymbol{\theta}$, hereafter called $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. The side information is supposed to be obtained in a previous estimation stage providing both the estimate and its accuracy (e.g., the previous iterate in a recursive implementation) or directly introduced by the designer with the purpose of controlling the estimator mean response (19).

Regarding (19), the estimator will be designed for a given *target* response $\boldsymbol{\psi}_t(\boldsymbol{\theta})$, and the referred constraints will be those obtained by minimizing the mismatch between the actual and intended estimator responses, that is

$$\xi \triangleq E_{\boldsymbol{\theta}} \|\boldsymbol{\psi}(\boldsymbol{\theta}) - \boldsymbol{\psi}_t(\boldsymbol{\theta})\|^2 \quad (20)$$

where $E_{\boldsymbol{\theta}}\{\cdot\}$ is the Bayesian expectation with respect to the prior $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ introduced previously.

Notice that ξ will be zero if and only if $\boldsymbol{\psi}(\boldsymbol{\nu}) = \boldsymbol{\psi}_t(\boldsymbol{\nu})$ for every $\boldsymbol{\nu}$, holding that $f_{\boldsymbol{\theta}}(\boldsymbol{\nu}) > 0$ (except for any null subset of null probability). Normally, this condition is too much stringent if the data vector \mathbf{y} is nonlinear in the parameters, mostly when the range of $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is large. In that case, a perfect match between $\boldsymbol{\psi}(\boldsymbol{\theta})$ and $\boldsymbol{\psi}_t(\boldsymbol{\theta})$ is not possible (i.e., $\xi > 0$) and, as a general solution, one may obtain the vector \mathbf{b} and the matrix \mathbf{M} yielding the weighted least squares (WLS) fitting between the multivariate functions $\boldsymbol{\psi}(\boldsymbol{\theta})$ and $\boldsymbol{\psi}_t(\boldsymbol{\theta})$:

$$\min_{\mathbf{b}, \mathbf{M}} \xi = E_{\boldsymbol{\theta}} \|\mathbf{b} + \mathbf{M}^H \mathbf{r}(\boldsymbol{\theta}) - \boldsymbol{\psi}_t(\boldsymbol{\theta})\|^2 \quad (21)$$

where $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is the weighting function.

If (21) is first solved for the independent term \mathbf{b} and then the resulting \mathbf{b} is substituted into (21), the desired solution for \mathbf{b} and \mathbf{M} is given by

$$\mathbf{b} = \boldsymbol{\psi}_t - \mathbf{M}^H \mathbf{r} \quad (22)$$

$$\tilde{\mathbf{Q}}\mathbf{M} = \mathbf{S} \quad (23)$$

where

$$\mathbf{r} \triangleq E_{\boldsymbol{\theta}}\{\mathbf{r}(\boldsymbol{\theta})\} \quad (24)$$

$$\boldsymbol{\psi}_t \triangleq E_{\boldsymbol{\theta}}\{\boldsymbol{\psi}_t(\boldsymbol{\theta})\} \quad (25)$$

$$\tilde{\mathbf{Q}} \triangleq E_{\boldsymbol{\theta}}\{(\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r})(\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r})^H\} \quad (26)$$

$$\mathbf{S} \triangleq E_{\boldsymbol{\theta}}\{(\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r})(\boldsymbol{\psi}_t(\boldsymbol{\theta}) - \boldsymbol{\psi}_t)^H\}. \quad (27)$$

Generally, the constraints obtained in (23) form an underdetermined system of equations because $R \triangleq \text{rank}(\tilde{\mathbf{Q}}) < M^2$ and \mathbf{S} is contained, by definition, in the column span of $\tilde{\mathbf{Q}}$. Hence, (23) is actually providing R design constraints on the matrix \mathbf{M} that, after the diagonalization of $\tilde{\mathbf{Q}} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^H$, can be formulated as follows:

$$\mathbf{V}^H \mathbf{M} = \mathcal{S} \quad (28)$$

where $\mathcal{S} \triangleq \boldsymbol{\Sigma}^{-1} \mathbf{V}^H \mathbf{S}$, $\boldsymbol{\Sigma} \in \mathbb{R}^{R \times R}$ is the diagonal matrix containing the nonzero eigenvalues of $\tilde{\mathbf{Q}}$, and $\mathbf{V} \in \mathbb{C}^{M^2 \times R}$ are the corresponding eigenvectors.

Therefore, since (23) is only forcing R constraints, the remaining degrees of freedom in \mathbf{M} can be used to optimize the discriminator variance. Specifically, the aim is to minimize the estimator variance in the Bayesian sense subject to the constraints on $\boldsymbol{\psi}(\boldsymbol{\theta})$ obtained in (23) or (28):

$$\min_{\mathbf{b}, \mathbf{M}} E_{\boldsymbol{\theta}} E \|\hat{\boldsymbol{\theta}} - \boldsymbol{\psi}(\boldsymbol{\theta})\|^2 = \min_{\mathbf{M}} E_{\boldsymbol{\theta}} E \|\mathbf{M}^H (\hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\theta}))\|^2$$

subject to $\tilde{\mathbf{Q}}\mathbf{M} = \mathbf{S}$ or $\mathbf{V}^H \mathbf{M} = \mathcal{S}$ (29)

where $E\{\cdot\}$ computes the expectation with respect to \mathbf{w} and \mathbf{x} .

The Bayesian expectation $E_{\boldsymbol{\theta}}\{\cdot\}$ in (21) and (29) plays a prominent role into the deduction because it allows the incorporation of the *a priori* knowledge on the parameter with which the designer is provided. Recovering now the value of \mathbf{b} from (22), the solution to the above constrained optimization problem is given by

$$\begin{aligned} \mathbf{b} &= \boldsymbol{\psi}_t - \mathbf{M}^H \mathbf{r} \\ \mathbf{M} &= \mathbf{P}^H \mathbf{S} = \mathcal{P}^H \mathcal{S} \end{aligned} \quad (30)$$

with \mathbf{P} and \mathcal{P} defined as

$$\mathbf{P} \triangleq (\tilde{\mathbf{Q}}\mathbf{Q}^{-1}\tilde{\mathbf{Q}})^{\#} \tilde{\mathbf{Q}}\mathbf{Q}^{-1} \quad (31)$$

$$\mathcal{P} \triangleq (\mathbf{V}^H \mathbf{Q}^{-1} \mathbf{V})^{-1} \mathbf{V}^H \mathbf{Q}^{-1}. \quad (32)$$

The matrix \mathbf{Q} contains the fourth-order central moments of vector \mathbf{y} and is defined as

$$\begin{aligned} \mathbf{Q} &\triangleq E_{\boldsymbol{\theta}}\{\mathbf{Q}(\boldsymbol{\theta})\} \\ \mathbf{Q}(\boldsymbol{\theta}) &\triangleq E\{(\hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\theta}))(\hat{\mathbf{r}} - \mathbf{r}(\boldsymbol{\theta}))^H\}. \end{aligned} \quad (33)$$

Regarding the inversion in (31) and (32), $\mathbf{Q}(\boldsymbol{\theta})$ and \mathbf{Q} are full-rank iff the noise matrix \mathbf{R}_w is positive definite (11). In Appendix A, a closed-form expression for $\mathbf{Q}(\boldsymbol{\theta})$ is derived by obtaining

$$\mathbf{Q}(\boldsymbol{\theta}) = \mathbf{R}^*(\boldsymbol{\theta}) \otimes \mathbf{R}(\boldsymbol{\theta}) + \mathcal{A}(\boldsymbol{\theta}) \mathbf{K} \mathcal{A}^H(\boldsymbol{\theta}) \quad (34)$$

where $\mathcal{A}(\boldsymbol{\theta}) \triangleq \mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})$, $\mathbf{R}(\boldsymbol{\theta})$ was introduced in (11), and matrix \mathbf{K} is defined as

$$\begin{aligned} \mathbf{K} &\triangleq E_{\mathbf{x}}\{\text{vec}(\mathbf{x}(n)\mathbf{x}^H(n))\text{vec}^H(\mathbf{x}(n)\mathbf{x}^H(n))\} \\ &\quad - \text{vec}(\mathbf{I})\text{vec}^H(\mathbf{I}) - \mathbf{I} \end{aligned} \quad (35)$$

denoting the fourth-order cumulant (kurtosis) matrix of the nuisance parameters \mathbf{x} .

In the case of zero-mean, *circular* complex nuisance parameters, the matrix \mathbf{K} can be calculated analytically (see Appendix B), where we obtain

$$\mathbf{K} = (\rho - 2)\text{diag}\{\text{vec}(\mathbf{I})\} \quad (36)$$

where the scalar $\rho \triangleq E\{|\mathbf{x}_k|^4\}/E^2\{|\mathbf{x}_k|^2\}$ is the fourth- to second-order moment ratio, and $\text{diag}(\cdot)$ converts a vector into a diagonal matrix. Regarding the matrix \mathbf{K} , it is well known that it vanishes for normally distributed nuisance parameters for which $\rho = 2$. Otherwise, the matrix \mathbf{K} provides the complete non-Gaussian information about the nuisance parameters that second-order estimators are able to exploit.

At this point, it is worth realizing that $\mathbf{Q}(\boldsymbol{\theta})$ and \mathbf{K} are calculated analytically from the known signal model introduced in Section II, avoiding the problematic estimation of fourth-order statistics. In case the nuisance parameters are not circular, the expectation in (35) can be computed numerically from the *known* pdf of \mathbf{x} , and this can be done offline. Finally, notice that \mathbf{K} can be estimated rapidly from a few realizations of $f_{\mathbf{x}}(\mathbf{x})$ if the nuisance parameters belong to a discrete alphabet.

Plugging now (30) into (18), the optimal second-order estimator and the associated covariance matrix is given by

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \boldsymbol{\psi}_t + \mathbf{S}^H \mathbf{P} (\hat{\mathbf{r}} - \mathbf{r}) = \boldsymbol{\psi}_t + S^H \mathcal{P} (\hat{\mathbf{r}} - \mathbf{r}) \\ \mathbf{C}(\boldsymbol{\theta}) &\triangleq E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\psi}(\boldsymbol{\theta}))(\hat{\boldsymbol{\theta}} - \boldsymbol{\psi}(\boldsymbol{\theta}))^H\} \\ &= \mathbf{M}^H \mathbf{Q}(\boldsymbol{\theta}) \mathbf{M}. \end{aligned} \quad (37)$$

Therefore, the *Bayesian* counterpart of $\mathbf{C}(\boldsymbol{\theta})$, whose trace was minimized in (29), is obtained after averaging $\mathbf{C}(\boldsymbol{\theta})$ with respect to the parameter prior as follows:

$$\begin{aligned} \mathbf{C} &\triangleq E_{\boldsymbol{\theta}}\{\mathbf{C}(\boldsymbol{\theta})\} = \mathbf{M}^H \mathbf{Q} \mathbf{M} = \mathbf{S}^H (\tilde{\mathbf{Q}}\mathbf{Q}^{-1}\tilde{\mathbf{Q}})^{\#} \mathbf{S} \\ &= S^H (\mathbf{V}^H \mathbf{Q}^{-1} \mathbf{V})^{-1} \mathcal{S}. \end{aligned} \quad (38)$$

Finally, returning to (20), the residual deviation of $\boldsymbol{\psi}(\boldsymbol{\theta})$ with respect to the ideal response $\boldsymbol{\psi}_t(\boldsymbol{\theta})$ can be expressed in any of these alternative forms:

$$\begin{aligned}\xi &= \text{Tr}\{\mathbf{M}^H \tilde{\mathbf{Q}}\mathbf{M} - 2\text{Re}\{\mathbf{M}^H \mathbf{S}\}\} + \sigma_\psi^2 \\ &= \sigma_\psi^2 - \text{Tr}\{\mathbf{S}^H \mathbf{P}\mathbf{S}\} = \sigma_\psi^2 - \text{Tr}\{\mathbf{S}^H \boldsymbol{\Sigma}\mathbf{S}\} \\ &= \sigma_\psi^2 - \text{Tr}\{\mathbf{S}^H \tilde{\mathbf{Q}}^\# \mathbf{S}\}\end{aligned}\quad (39)$$

where $\sigma_\psi^2 \triangleq E_{\boldsymbol{\theta}}\|\boldsymbol{\psi}_t(\boldsymbol{\theta}) - \boldsymbol{\psi}_t\|^2$, and (22), (23), and (30) have been used. Notice that any matrix \mathbf{M} solving (23) or (28) yields the same value of ξ . Simplifications are done taking into account that $\mathbf{P}\tilde{\mathbf{Q}}\mathbf{P}^H = \mathbf{P}$. Furthermore, the last equality is verified because \mathbf{P} can be replaced by any matrix holding $\mathbf{P}\tilde{\mathbf{Q}}\mathbf{P}^H = \mathbf{P}$ as, for example, the pseudo-inverse $\tilde{\mathbf{Q}}^\#$.

VI. SMALL-ERROR APPROXIMATION

Thus far, the variability of $\boldsymbol{\theta}$ has been considered by means of the prior $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ introduced in Section V. In this section, the asymptotic case in which this variability is very small is studied ($\boldsymbol{\theta} \simeq \boldsymbol{\theta}_o$). This *small-error assumption* is commonly adopted when designing iterative schemes or *trackers* in nonlinear estimation problems, as explained in Section IV. After convergence, the algorithm is assumed to operate in the neighborhood of the true solution, where $\boldsymbol{\psi}(\boldsymbol{\theta})$ is approximately linear in the parameter, irrespective of the received signal parametrization. In that case, second-order estimators are known to be efficient at low SNRs or if the nuisance parameters are Gaussian, as claimed in Section III.

Next, the general Bayesian estimator deduced in Section V is particularized for a very informative prior $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$, holding that $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) < \varepsilon$ for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_o$ with ε arbitrarily small. This impulsive prior conveniently models the small-error regime described in Section IV. In that case, the Taylor expansion around $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ can be used to approximate the expected value of those complex matrices appearing in (37).

Thus, if $\mathcal{F}(\boldsymbol{\theta})$ is a generic complex matrix depending on the vector of parameters $\boldsymbol{\theta}$, its mean value in the neighborhood of $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ can be approximated as follows:

$$E_{\boldsymbol{\theta}}\{\mathcal{F}(\boldsymbol{\theta})\} \simeq \mathcal{F}(\boldsymbol{\theta}_o) + \frac{1}{2} \sum_{p,q=1}^P \left. \frac{\partial^2 \mathcal{F}(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} [\mathbf{C}_{\boldsymbol{\theta}}]_{p,q} \quad (40)$$

where the linear term is omitted, taking into account that $E_{\boldsymbol{\theta}}\{\boldsymbol{\theta}\} \triangleq \boldsymbol{\theta}_o$ by definition, and $\mathbf{C}_{\boldsymbol{\theta}}$ is the *a priori* covariance matrix of the parameter:

$$\mathbf{C}_{\boldsymbol{\theta}} \triangleq E_{\boldsymbol{\theta}}\{(\boldsymbol{\theta} - \boldsymbol{\theta}_o)(\boldsymbol{\theta} - \boldsymbol{\theta}_o)^H\}. \quad (41)$$

In Appendix C, \mathbf{r} , $\boldsymbol{\psi}_t$, $\tilde{\mathbf{Q}}$, \mathbf{S} (24)–(27) and \mathbf{Q} (33) are approximated using (40), obtaining

$$\mathbf{r} \simeq \mathbf{r}(\boldsymbol{\theta}_o) \triangleq \mathbf{r}_o \quad (42)$$

$$\boldsymbol{\psi}_t \simeq \boldsymbol{\psi}_t(\boldsymbol{\theta}_o) \triangleq \boldsymbol{\theta}_o \quad (43)$$

$$\tilde{\mathbf{Q}} \simeq \mathbf{D}_r \mathbf{C}_{\boldsymbol{\theta}} \mathbf{D}_r^H \quad (44)$$

$$\mathbf{S} \simeq \mathbf{D}_r \mathbf{C}_{\boldsymbol{\theta}} \mathbf{D}_\psi^H \quad (45)$$

$$\mathbf{Q} \simeq \mathbf{Q}(\boldsymbol{\theta}_o) \triangleq \mathbf{Q}_o \quad (46)$$

where

$$\mathbf{D}_r \triangleq \left. \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \quad (47)$$

$$\mathbf{D}_\psi \triangleq \left. \frac{\partial \boldsymbol{\psi}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}. \quad (48)$$

Finally, under the small-error assumption, the prior is concentrated in $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ so that $\mathbf{C}_{\boldsymbol{\theta}}$ in (41) collapses at this point, becoming proportional to a given matrix $\mathbf{C}_{\boldsymbol{\theta}}^0$, which is defined as

$$\mathbf{C}_{\boldsymbol{\theta}}^0 \triangleq \lim_{\delta \rightarrow 0} \frac{1}{\delta} \mathbf{C}_{\boldsymbol{\theta}} \quad (49)$$

with $\delta \triangleq \|\boldsymbol{\theta} - \boldsymbol{\theta}_o\|$ the radius of the infinitesimal ball in which the prior is defined.

Evidently, under the small-error assumption, the fitting of $\boldsymbol{\psi}(\boldsymbol{\theta})$ to $\boldsymbol{\psi}_t(\boldsymbol{\theta})$ inside the referred ball is much easier. Returning to (20), it is easy to show that a perfect matching is achieved ($\xi = 0$) iff the derivatives of $\boldsymbol{\psi}_t(\boldsymbol{\theta})$ and $\boldsymbol{\psi}(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ are identical:

$$\mathbf{M}^H \mathbf{D}_r = \mathbf{D}_\psi. \quad (50)$$

For the time being, the target response $\boldsymbol{\psi}_t(\boldsymbol{\theta})$ is supposed to verify the above equality for at least one matrix \mathbf{M} . Therefore, solving again the minimization problem in (29) under the constraints on \mathbf{b} and \mathbf{M} obtained in (22) and (50), respectively, the optimal *small-error* estimator is given by

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_o + \mathbf{D}_\psi (\mathbf{D}_r^H \mathbf{Q}_o^{-1} \mathbf{D}_r)^\# \mathbf{D}_r^H \mathbf{Q}_o^{-1} (\hat{\mathbf{r}} - \mathbf{r}_o) \quad (51)$$

where \mathbf{r}_o and \mathbf{Q}_o were defined in (42) and (46), and the Moore–Penrose pseudoinverse is maintained to cover those cases in which \mathbf{D}_r is singular (50). Finally, the covariance matrix of the estimation error is the following:

$$\begin{aligned}\mathbf{C}(\boldsymbol{\theta}_o) &= E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)^H\} \\ &= \mathbf{D}_\psi (\mathbf{D}_r^H \mathbf{Q}_o^{-1} \mathbf{D}_r)^\# \mathbf{D}_\psi^H.\end{aligned}\quad (52)$$

Regarding the obtained solution, it is remarkable that the estimator covariance matrix in (52) has the same structure as the CRB in (15), where

$$\mathbf{J}_2 \triangleq \mathbf{D}_r^H \mathbf{Q}_o^{-1} \mathbf{D}_r \quad (53)$$

plays the same role than the Fisher information matrix (FIM) in the case of *second-order* estimators. Therefore, (52) can be seen as the particularization of the CRB to second-order estimation techniques. In fact, (52) is a member of the class of quadratic covariance bounds analyzed in [13] following the representation proposed in [14]. Note that \mathbf{J}_2 has been proved to coincide with the FIM of the problem when the SNR is asymptotically low or the nuisance parameters are Gaussian (see Section III).

In general, it can be affirmed that

$$E\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_o)^H\} \geq \mathbf{C}(\boldsymbol{\theta}_o) \geq \mathbf{B}(\boldsymbol{\theta}_o) \quad \forall \boldsymbol{\theta}_o \quad (54)$$

for any unbiased estimator $\hat{\boldsymbol{\theta}}$ based on the sample covariance matrix $\hat{\mathbf{R}} = \mathbf{y}\mathbf{y}^H$, where $\mathbf{M}_1 \geq \mathbf{M}_2$ means that $\mathbf{M}_1 - \mathbf{M}_2$ is semipositive definite, and $\mathbf{B}(\boldsymbol{\theta}_o)$ is the CRB of $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\psi}_t(\boldsymbol{\theta})$ (15). As stated before, the second identity in (54) is solely

verified when the SNR tends to zero or the nuisance parameters are Gaussian random variables.

A. Second-Order Identifiability

Regarding the solution obtained in (51), two different cases can be distinguished, depending on whether \mathbf{D}_r is full-rank or not. In the first case, (50) is always satisfied *whatever* the content of \mathbf{D}_ψ . On the other hand, the second case requires that \mathbf{D}_ψ lies in the row span of \mathbf{D}_r . If not, (50) would be inconsistent, and no solution would exist for the given set of target derivatives (\mathbf{D}_ψ). In fact, $R = \text{rank}(\mathbf{D}_r)$ states the number of derivatives that can be forced or, equivalently, how many parameters can be discriminated *separately* from the considered signal model. Hence, the rank of \mathbf{D}_r is strongly related to the *local* (small-error) identifiability of the multivariate problem at hand.

Regarding the singular case, the authors proved in [11] that a finite variance estimator does not exist if (50) is not satisfied. Alternatively, the same conclusion can be drawn following the geometrical interpretation carried out in [15]. Therefore, if the desired target response $\psi_t(\boldsymbol{\theta})$ does not satisfy (50), the designer has to proceed as in (21) to obtain the best approximation to $\psi_t(\boldsymbol{\theta})$ holding (50). Thus, substituting (44) and (45) into (23), the system of constraints for the small-error scenario would be

$$\mathbf{D}_r \mathbf{C}_\theta^0 \mathbf{D}_r^H \mathbf{M} = \mathbf{D}_r \mathbf{C}_\theta^0 \mathbf{D}_\psi^H \quad (55)$$

where the *a priori* covariance matrix \mathbf{C}_θ^0 (49) is used to carry out the WLS fitting proposed in (21). Otherwise, if \mathbf{D}_r is full-rank, \mathbf{C}_θ^0 is not profitable, showing that Bayesian estimators cannot improve deterministic ones when the small-error assumption applies.

Two main circumstances lead to the singularity of \mathbf{D}_r : a wrong problem statement and/or the estimator finite resolution. In the first case, the estimator ambiguity is attributable to one of the following issues: 1) There are more parameters than data samples ($M < P$); 2) the P parameters are not linearly independent so that the model is “overparameterized”; or 3) the inherent phase ambiguity of second-order estimators is due to the insensitivity of $\hat{\mathbf{R}}$ to any phase-offset in the received signal.⁴

On the other hand, although the problem is correctly formulated, the estimator is unable to resolve two parameters of the same nature when their values are very similar. For example, this problem arises when tracking multiuser parameters, e.g., time-of-arrival, or AoA in array signal processing (see Section VII-A). It is worth noting that this situation, contrary to the ambiguities related before, cannot be predicted beforehand; therefore, it is not possible to guarantee (50) all the time. Therefore, the constraints in (55) must be used instead of those in (50), and the general estimator in (37) must be considered using the small-error matrices in (42)–(46).

However, from the designer viewpoint, the use of (55) may be problematic because the estimator would reduce automatically the rank of \mathbf{D}_ψ when entering into a singular situation (e.g., if two users cross each other, as studied in Section VII-A), changing the value of \mathbf{D}_ψ . In the next section, this problem is overcome by setting the value of the cross derivatives free.

⁴The signal modulus would be also ambiguous if the noise variance σ_w^2 were not known, as we have assumed throughout the paper.

B. Generalized Second-Order Constrained Estimators

Thus far, the estimator is designed to have a target mean response $\psi_t(\boldsymbol{\theta})$, yielding the constraints in (50) when working under the small-error regime. Whereas the diagonal terms in \mathbf{D}_ψ are related to the estimator bias in the neighborhood of $\boldsymbol{\theta} = \boldsymbol{\theta}_o$, the cross-terms reflect the coupling between parameters or, in other words, the interparameter interference (IPI). The classical unbiased solution forces $\mathbf{D}_\psi = \mathbf{I}$ in order to yield uncorrelated estimates with no IPI. However, strictly speaking, unbiased estimators are only required to fix the diagonal terms to 1, that is, $\text{diag}(\mathbf{D}_\psi) = \text{diag}(\mathbf{I})$, since the IPI contribution is zero-mean under the small-error approximation and, therefore, can only increase the estimator variance. Moreover, in noisy scenarios, the IPI-free condition usually causes noise-enhancement, whereas if the cross-terms in (50) are kept free, the estimator automatically makes a tradeoff among noise, self-noise, and IPI in order to minimize the overall variance.

Therefore, the proposed second-order *unbiased* estimator of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}_o + \text{diag}(\mathbf{D}_\psi) \text{diag}^{-1}(\mathbf{J}_2) \mathbf{D}_r^H \mathbf{Q}_o^{-1} (\hat{\mathbf{r}} - \mathbf{r}_o) \quad (56)$$

where $\text{diag}(\cdot)$ stands here for the diagonal matrix built from the diagonal of the argument, and $\mathbf{J}_2 \triangleq \mathbf{D}_r^H \mathbf{Q}_o^{-1} \mathbf{D}_r$ is the second-order FIM introduced in (53).

On the other hand, if the aim is estimating a given transformation of the parameter $\boldsymbol{\alpha} \triangleq \psi_t(\boldsymbol{\theta})$ instead of $\boldsymbol{\theta}$ itself, the cross-terms significance changes radically, and assuming \mathbf{D}_ψ is full matrix (all the elements different from 0), any *unbiased* estimator of $\boldsymbol{\alpha}$ is required to fulfill (50), leading to the original small-error solution in (51). In general, if \mathbf{D}_ψ is sparse, only the constraints in (50) corresponding to nonzero elements of \mathbf{D}_ψ have to be imposed to obtain unbiased estimators of $\boldsymbol{\alpha}$.

VII. ANGLE-OF-ARRIVAL TRACKING

In this section, the second-order framework presented in the paper is applied to an illustrative multivariate estimation problem: AoA tracking in the context of wireless communications [16]. In digital communications, the transmitted symbols are unknown unless the receiver is assisted with known training data to facilitate the estimation of certain relevant parameters. When the symbols are unknown, blind methods are required to extract these parameters of interest from the received signal [3], [2], [8]. In that case, the transmitted symbols are nuisance parameters, degrading the estimator performance. These nuisance parameters are clearly non-Gaussian since they are drawn from a discrete alphabet (digital constellation).

The optimal AoA second-order tracker deduced in Section VI is compared with the other quadratic ML-based methods described in Section III. Simulations have confirmed that second-order tracking can be improved if the discrete nature of the nuisance parameters is exploited by means of the fourth-order cumulant matrix \mathbf{K} (36). The optimal second-order scheme will be named in the sequel the Best Quadratic Unbiased Estimator (BQUE) [17] since it constitutes a logical extension of the Best Linear Unbiased Estimator (BLUE) [1, Sec. 6]. Moreover, simulations have shown the interest of the

MUI-resistant AoA tracker formulated in Section VI-B in noisy scenarios when the angular separation of the sources is small.

Simulations will be contrasted with two well-known theoretical lower bounds: the Modified CRB (MCRB) [4], [10], [18] and the *Gaussian* Unconditional CRB (UCRB) [2]–[4]. The MCRB is a useful bound employed in communications that may be attained when the nuisance parameters are known. In the context of array signal processing, this happens when we have a temporal reference (training data). On the other hand, the UCRB has been shown in some problems to be a lower bound for the performance of second-order blind estimators [4]. However, as shown through the following examples, it can happen that for medium-to-high SNRs, the estimator variance violates the UCRB when the nuisance parameters are not Gaussian. The MCRB and UCRB generic expressions are given next for completeness:

$$\begin{aligned} \mathbf{B}_m(\boldsymbol{\theta}) &= 0.5 \mathbf{D}_\psi \mathbf{R}_e^{-1} (\mathcal{D}^H(\boldsymbol{\theta}) (\mathbf{I} \otimes \mathbf{R}_w^{-1}) \mathcal{D}(\boldsymbol{\theta})) \mathbf{D}_\psi^H \\ \mathbf{B}_u(\boldsymbol{\theta}) &= \mathbf{D}_\psi [\mathbf{D}_r^H(\boldsymbol{\theta}) (\mathbf{R}^*(\boldsymbol{\theta}) \otimes \mathbf{R}(\boldsymbol{\theta}))^{-1} \mathbf{D}_r(\boldsymbol{\theta})]^{-1} \mathbf{D}_\psi^H \end{aligned} \quad (57)$$

with $\mathcal{D}(\boldsymbol{\theta}) \triangleq [\text{vec}\{\partial \mathbf{A}(\boldsymbol{\theta})/\partial \theta_1\}, \dots, \text{vec}\{\partial \mathbf{A}(\boldsymbol{\theta})/\partial \theta_P\}]$. Notice that these bounds must be computed with $M = N \simeq 0.5/B_n$, which is the effective observation size in the case of closed-loop trackers (Section IV).

A. Signal Model for AoA Tracking

Thus far, the study has been limited to stationary scenarios in which the parameter $\boldsymbol{\theta}$ is a static variable. In the context of mobile communications, it is important to obtain an accurate estimate of the different users angular position but also to track these parameters as the transmitters move around the base station. Consequently, it is necessary to estimate both the angle and the angular speed of every source transmitting toward the base station. Higher derivatives of the AoA (acceleration and so on) will be disregarded from the study for clarity. Therefore, based on the small-error estimators obtained in (51) and (56), a closed-loop scheme (tracker) as the one suggested in (16) will be implemented in order to track the parameter $\boldsymbol{\theta}_n$ evolution. To do so, we have to incorporate the parameter dynamics (speed, acceleration, etc.) into the model.

Formally, let us consider the problem of tracking the AoA of P narrowband sources impinging into a uniform linear array composed of M antennas spaced $\lambda/2$ m, where λ is the shared signal wavelength. Let us consider that all the transmitters are visible from the base station array and that they do not experience multipath propagation. Let $\boldsymbol{\phi}(t) \in [-\pi, \pi)^P$ be the temporal evolution of the P AoAs in radians and $\boldsymbol{\phi}'(t) \triangleq \partial \boldsymbol{\phi}(t)/\partial t$ the respective derivatives accounting for the angular speed. Let us assume that the acceleration and higher derivatives are negligible during the observation time, that is, $\partial^i \boldsymbol{\phi}(t)/\partial t^i = \mathbf{0}$ for $i > 1$. Furthermore, let us assume that the bandwidth of $\boldsymbol{\phi}(t)$ does not exceed $1/2T$, where T is the symbol period. In that case, $\boldsymbol{\phi}(nT)$ holds the sampling theorem, and the P trajectories can be ideally reconstructed from their samples $\boldsymbol{\phi}_n \triangleq \boldsymbol{\phi}(nT)$,

yielding to the following discrete-time dynamical model or state equation

$$\boldsymbol{\phi}_{n-l} = \boldsymbol{\phi}_n - l \boldsymbol{\phi}'_n \quad (58)$$

where the angular speed $\boldsymbol{\phi}'_n$ is normalized to the symbol period T . Therefore, the composed vector of parameters that must be estimated to track the users without having any systematic pursuit error is

$$\boldsymbol{\theta}_{n+1} \triangleq \begin{bmatrix} \boldsymbol{\phi}_{n+1} \\ \boldsymbol{\phi}'_{n+1} \end{bmatrix} = \mathbf{F} \boldsymbol{\theta}_n \quad (59)$$

with

$$\mathbf{F} \triangleq \begin{bmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}. \quad (60)$$

Consequently, the optimal second-order AoA tracker is given by

$$\hat{\boldsymbol{\theta}}_{n+1} = \mathbf{F} \hat{\boldsymbol{\theta}}_n + \boldsymbol{\Lambda}_\mu \mathbf{F} \mathbf{J}_2^\#(\hat{\boldsymbol{\theta}}_n) \mathbf{D}_r^H \mathbf{Q}^{-1}(\hat{\boldsymbol{\theta}}_n) (\hat{\mathbf{r}} - \mathbf{r}(\hat{\boldsymbol{\theta}}_n)) \quad (61)$$

using the small-error expression obtained in (51) with $\mathbf{D}_\psi = \mathbf{F}$. First of all, notice that this expression coincides with (16) when the SNR tends to zero if $\mathbf{g}(\boldsymbol{\theta}_n) = \mathbf{F} \boldsymbol{\theta}_n$.

The above solution forces to zero all the cross-derivatives of $\boldsymbol{\psi}(\boldsymbol{\theta})$, including the IPI terms associated with the interference from other users (see Section VI-B). This interference is referred to as multiuser or multiple access interference (MUI or MAI) in the literature. Thus, the second-order AoA tracker in (61) will be referred to as *MUI-free* hereafter.

On the other hand, following the reasoning in Section VI.B, it is not strictly necessary to cancel out the cross-derivatives corresponding to different users because the tracker optimization will remove the MUI contribution if the SNR is sufficiently high. Likewise, if the SNR is low, the MUI term will be automatically ignored to not enhance the noise contribution. Hence, it is only necessary to decouple the estimates of $\boldsymbol{\phi}_n$ and $\boldsymbol{\phi}'_n$ in order to have unbiased estimates of $\boldsymbol{\theta}_{n+1}$ (59). If not, AoA estimation errors would yield angular speed deviations and vice versa. To avoid this, we have to constrain these cross-derivatives to zero as indicated next:

$$\left. \frac{\partial [\hat{\boldsymbol{\phi}}_n]_p}{\partial [\hat{\boldsymbol{\phi}}_n']_p} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} = \left. \frac{\partial [\hat{\boldsymbol{\phi}}_n']_p}{\partial [\hat{\boldsymbol{\phi}}_n]_p} \right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n} = 0 \quad p = 1, \dots, P \quad (62)$$

while the rest of cross-derivatives is liberated. We will refer to this solution as the *MUI-resistant* AoA tracker in the sequel.

To complete the signal model, the received signal is introduced into the matched filter and then sampled at one sample per symbol in order to collect L snapshots. Independent snapshots are obtained, assuming that the actual modulation is ISI-free and that the P signals are perfectly synchronized. Notice that the problem dynamics (angle and angular velocity) require processing of $L \geq 2$ snapshots.

According to the considerations above, the snapshot recorded at time $n-l$ is given by

$$\mathbf{y}_{n-l} = \mathbf{A}_{n-l}(\boldsymbol{\theta}_n) \mathbf{x}_{n-l} + \mathbf{w}_{n-l} \quad (63)$$

where \mathbf{x}_{n-l} is the vector containing the symbols transmitted by the P users at time $n-l$, \mathbf{w}_{n-l} are the *white* noise samples ($\mathbf{R}_w = \sigma_w^2 \mathbf{I}$), and the p th column of $\mathbf{A}_{n-l}(\boldsymbol{\theta}_n)$

$$[\mathbf{A}_{n-l}(\boldsymbol{\theta}_n)]_p \triangleq \begin{bmatrix} 1 \\ \exp[j\pi \sin([\phi_n]_p - l[\phi'_n]_p)] \\ \vdots \\ \exp[j\pi(M-1) \sin([\phi_n]_p - l[\phi'_n]_p)] \end{bmatrix} \quad (64)$$

is the steering vector associated with the p th source at time $n-l$, where $j \triangleq \sqrt{-1}$. Notice that $\mathbf{A}_{n-l}(\boldsymbol{\theta}_n)$ incorporates the known dynamical model (58).

In order to reproduce the vectorial model in (1), the L snapshots are stacked to build the following spatio-temporal observation:

$$\mathbf{y}(n) \triangleq \begin{bmatrix} \mathbf{y}_n \\ \vdots \\ \mathbf{y}_{n-L+1} \end{bmatrix} = \mathbf{A}(\boldsymbol{\theta}_n) \mathbf{x}(n) + \mathbf{w}(n) \quad (65)$$

where $\mathbf{x}(n)$ and $\mathbf{w}(n)$ are constructed as $\mathbf{y}(n)$, and the transfer matrix $\mathbf{A}(\boldsymbol{\theta}_n)$ is given by

$$\mathbf{A}(\boldsymbol{\theta}_n) \triangleq \begin{bmatrix} \mathbf{A}_n(\boldsymbol{\theta}_n) & & \\ & \ddots & \\ & & \mathbf{A}_{n-L+1}(\boldsymbol{\theta}_n) \end{bmatrix}. \quad (66)$$

As stated before, once the signal model has been determined, we find the set of constituent matrices in (51) and (56). To conclude, the derivatives of the steering vectors are provided by

$$\begin{aligned} \frac{\partial[\mathbf{A}_{n-l}(\boldsymbol{\theta}_n)]_{p,m}}{\partial[\phi_n]_q} &= j\pi m \cos([\phi_n]_p - l[\phi'_n]_p) \\ &\quad \times \exp[j\pi \sin([\phi_n]_p - l[\phi'_n]_p)] \cdot \delta(p-q) \\ \frac{\partial[\mathbf{A}_{n-l}(\boldsymbol{\theta}_n)]_{p,m}}{\partial[\phi'_n]_q} &= -j\pi ml \cos([\phi_n]_p - l[\phi'_n]_p) \\ &\quad \times \exp[j\pi \sin([\phi_n]_p - l[\phi'_n]_p)] \cdot \delta(p-q) \end{aligned} \quad (67)$$

for all $p, q \in \{1, \dots, P\}$, where $\delta(\cdot)$ stands for the Kronecker delta.

B. Numerical Results

Two independent sources transmitting from the far-field to a uniform linear array composed of $M = 4$ antennas are simulated. The received power is assumed to be the same for simplicity. Both signals are quadrature phase shift keying (QPSK) modulated, and two snapshots ($L = 2$) are recorded at the matched-filter output.

The figure of merit considered in this section is the estimator normalized steady-state variance defined as

$$\text{var}(\hat{\boldsymbol{\theta}}) \triangleq \frac{E\|\hat{\boldsymbol{\phi}}_n - \boldsymbol{\phi}_n\|^2}{P\Delta\phi^2} \quad (68)$$

where $\Delta\phi \triangleq ([\phi_n]_2 - [\phi_n]_1)/2$ is half of the sources separation. The variance will be plotted as a function of the SNR per source at the matched-filter output $\gamma \triangleq E_s/N_0 = \sigma_w^{-2}$, where E_s is the received symbol energy and N_0 the noise double-sided spectral density.

Two AoA trackers forcing a different set of constraints on $\boldsymbol{\psi}(\boldsymbol{\theta})$ will be tested:

- 1) *MUI-free AoA tracker*: $\mathbf{M}^H \mathbf{D}_r = \mathbf{D}_\psi = \mathbf{I}$;
- 2) *MUI-resistant AoA tracker*: $\text{diag}(\mathbf{M}^H \mathbf{D}_r) = \text{diag}(\mathbf{D}_\psi) = \text{diag}(\mathbf{I})$, and the cross terms in (62) are set to zero.

Two different scenarios have been simulated in order to illustrate and validate the main conclusions of the paper.

Simulation 1:

Two Users Crossing: Fig. 1 shows that the MUI-free AoA tracker (left plot) loses tracking as the two sources approach each other due to the noise enhancement observed when the SNR is low (SNR = 10 dB). This situation arises because, when the users are transmitting from similar angles, the matrix \mathbf{D}_r becomes nearly singular and the estimator variance in (52) augments suddenly.

On the other hand, the MUI-resistant AoA tracker (right plot) overcomes this critical situation because it does not try to remove the MUI term associated with the cross derivatives of \mathbf{D}_r when the noise contribution is dominant (low SNR). Following the explanation in Section VI-B, the MUI-resistant AoA tracker liberates the cross derivatives in \mathbf{D}_r while the users are crossing, and matrix \mathbf{D}_r is badly conditioned. In this manner, the tracker does not enhance the noise contribution and is able to remain “locked” during the crossing.

Simulation 2:

Steady-State Variance for Two Near Sources: The steady-state variance of the MUI-free AoA tracker is plotted as a function of the SNR, considering that we have two still users separated 10° (Fig. 2) and 1° (Fig. 3). The noise equivalent loop bandwidth B_n (see Section IV) has been selected in order to guarantee the small-error condition for all the simulated SNRs (Section VI). For the studied setup, the noise enhancement caused by the source’s proximity is found to be negligible. This fact makes the two suggested implementations (MUI-resistant and MUI-free) to be practically equivalent in the simulated scenarios. Theoretically, the performance of the MUI-free estimator is very limited at low SNRs when the two sources are close, as shown in Fig. 4, whereas its competitor (MUI-resistant) achieves the single user performance, whatever the simulated SNR. Fig. 4 illustrates the potential gain that the MUI-resistant alternative offers in terms of steady-state variance when the problem is badly conditioned and the observations are very corrupted by the noise.

In Figs. 2 and 3, the optimal second-order tracker has been compared with the ML-based trackers formulated in Section III. The first conclusion is that the low-SNR approximation appears to be useless in these critical scenarios for the SNRs of interest. The underlying motive is the so-called self-noise, i.e., the variance floor caused by the nuisance parameters at high SNR (Section III). The self-noise is really irrelevant when the SNR

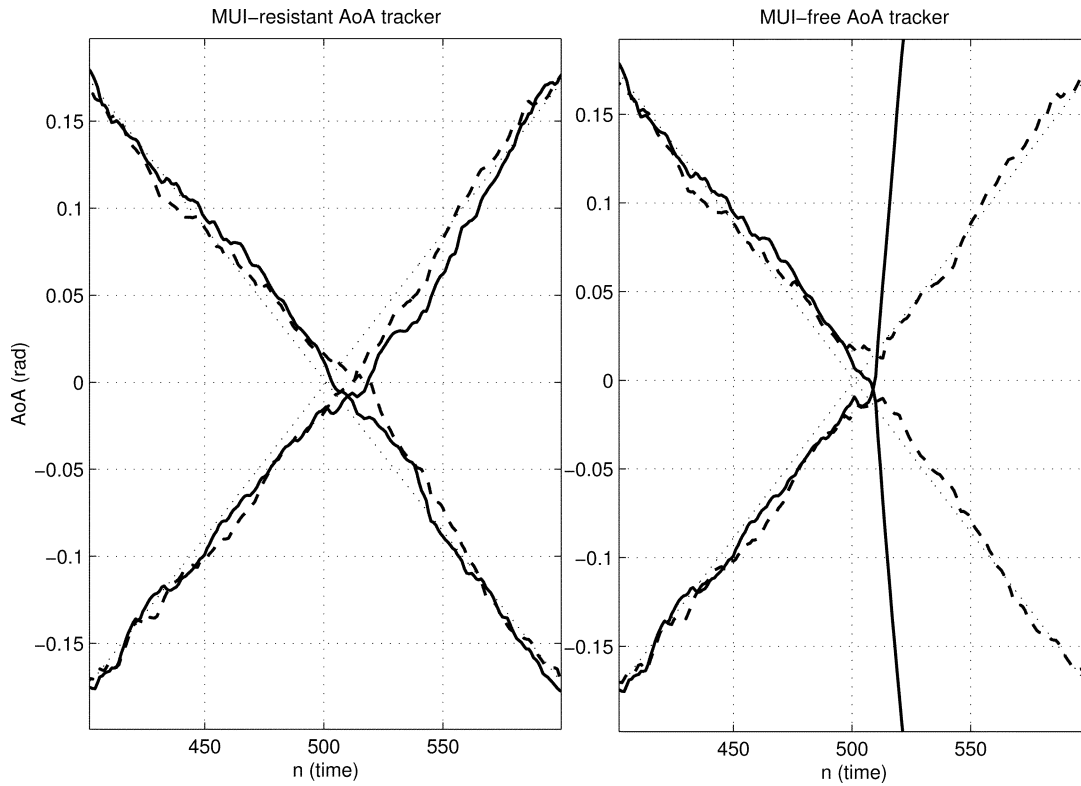


Fig. 1. AoA tracking of two users whose trajectories cross at time instant $n = 500$. The output of the MUI-free and MUI-resistant trackers is plotted on the left- and right-hand sides, respectively. Two simulations are run with two different outcomes for the MUI-free tracker: Tracking is lost (solid line) or the two sources are interchanged (dashed line). The signal SNR is fixed to 10 dB in both cases.

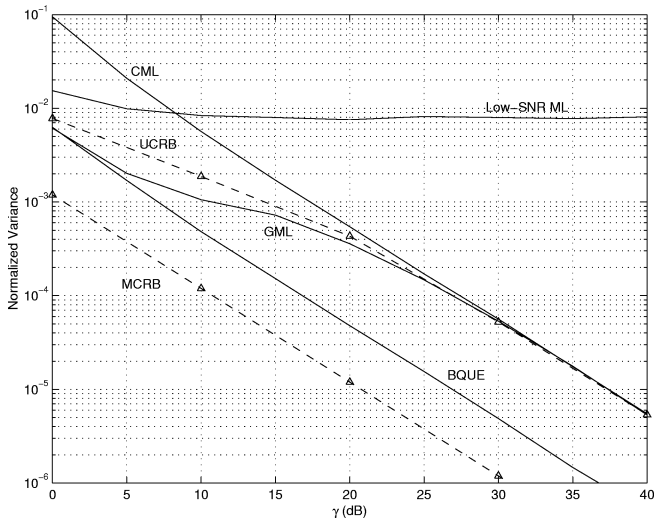


Fig. 2. Steady-state variance of the AoA tracker for two sources located at $\pm 5^\circ$ degrees from the broadside. The loop-bandwidth is set to $B_{\text{n}}=1.25 \cdot 10^{-3}$ and the MUI-free estimator is simulated.

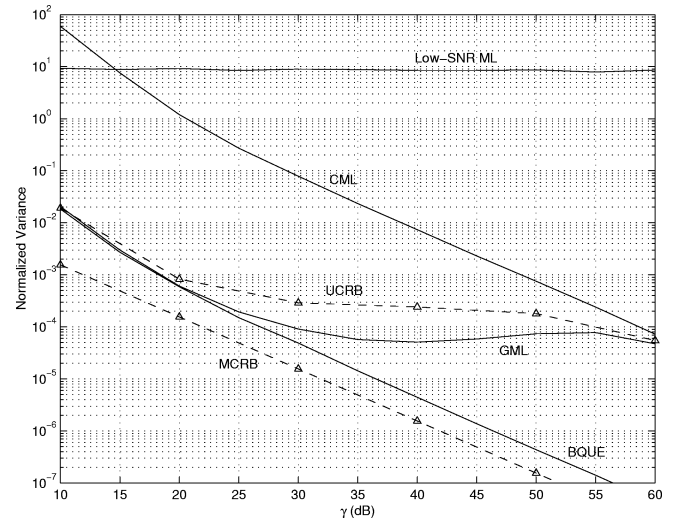


Fig. 3. Steady-state variance of the AoA tracker for two sources located at $\pm 5^\circ$ degrees from the broadside. The loop-bandwidth is set to $B_{\text{n}}=1.65 \cdot 10^{-4}$ and the MUI-free estimator is simulated.

tends to zero, but it becomes dominant as soon as the SNR is increased. Notice that in the AoA estimation problem at hand, the so-called *self-noise* is generated by the random symbols (nuisance parameters) from the user of interest as well as the other interfering users. Therefore, the MUI and self-noise contributions are strongly connected in this case study.

To overcome the low-SNR UML variance floor, the CML tracker was proposed in Section III-B. The CML is able to yield self-noise free estimates, but it suffers from noise enhancement

when the SNR is low because it tries to decorrelate the nuisance parameters from the different users.

Regarding the GML AoA tracker presented in Section III-C, the convergence to the CML solution for high SNRs and to the low-SNR UML solution for low SNRs (if the x-axis were expanded) is observed. Between these two asymptotic extremes, the GML adjusts its coefficients, depending on the actual SNR to minimize the joint contribution of the noise and the self-noise.

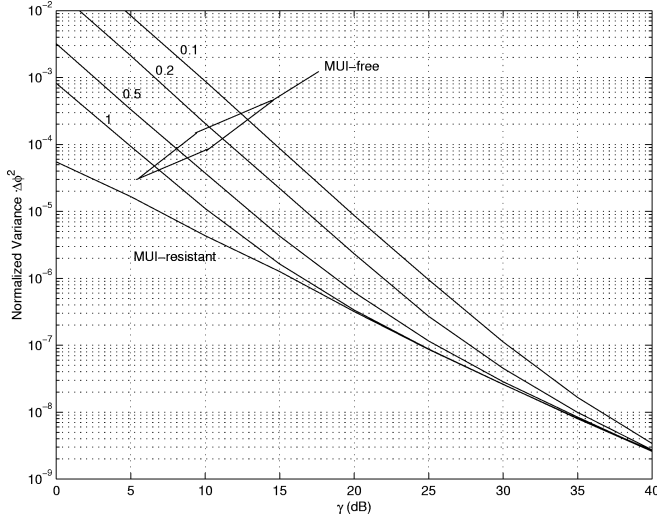


Fig. 4. Normalized variance as a function of the SNR for the MUI-free and MUI-resistant AoA trackers when the sources are separated 0.1° , 0.2° , 0.5° , or 1° , as indicated for each curve. The tracker loop bandwidth is set to $B_n = 1.25 \cdot 10^{-3}$.

Indeed, the GML solution is found to be the best quadratic estimator or tracker based uniquely on the second-order moments of the nuisance parameters, i.e., $\mathbf{K} = \mathbf{0}$ in (35). Nonetheless, when comparing the variance of the GML and BQUE AoA trackers in Figs. 2 and 3, it is confirmed that second-order estimation is improved for medium-to-high SNRs if the fourth-order statistical knowledge on the nuisance parameters ($\mathbf{K} \neq \mathbf{0}$) is exploited. The resulting gain is shown to be greater when the angular separation is reduced if one compares the left- and right-hand side plots in Fig. 3. Moreover, when the loop bandwidth B_n is small (Fig. 2), the BQUE performance is rather close to the one predicted by the MCRB in the case of known nuisance parameters, and it definitely constitutes the lower bound for the variance of any unbiased estimator based on the sample covariance matrix.

Surprisingly, the UCRB bound is violated by the GML estimator out of the aforementioned asymptotic cases because the nuisance parameters are actually non-Gaussian (QPSK discrete symbols), and the UCRB is based on the Gaussian assumption.

VIII. CONCLUSION

The problem of parameter estimation in the presence of random nuisance parameters is dealt with in the paper. It is shown that the unconditional ML solution, which is known to be efficient for large data records, is generally unknown unless the SNR is asymptotically low or the nuisance parameters are Gaussian distributed. In both cases, the sample covariance matrix is proved to be (asymptotically) a sufficient statistic for estimating the parameters of interest. From this background, the paper formulates the optimal *second-order* estimator for a given SNR and an arbitrary distribution of the nuisance parameters. The Bayesian philosophy is adopted in order to provide the estimator with any *a priori* knowledge about the parameters of interest. The resulting estimator is shown to exploit the fourth-order cumulants (i.e., kurtosis) of the nuisance parameters that admit a simple closed form or are easily

computed offline. Simulations have confirmed the importance of considering this fourth-order information when tracking the AoA of two near sources, even if the SNR is relatively low.

Unfortunately, only when the covariance matrix is linear in the parameter of interest, the obtained solution is efficient, meaning that the CRB is attained. Otherwise, the bias impairment relegates the utilization of second-order schemes to iterative schemes (or trackers). Accordingly, the optimal second-order iterative estimator or tracker is deduced from the general Bayesian solution, considering that the *small-error* approximation is verified once the algorithm has converged to the true solution. The resulting small-error estimator is found to attain the CRB when the SNR is low or the nuisance parameters are Gaussian. In addition, the proposed solution is the best quadratic unbiased estimator (BQUE) with independence of the actual SNR or the nuisance parameter's distribution. Moreover, the variance of the proposed estimator constitutes a new, realizable lower bound for the variance of any unbiased estimation technique based on the sample covariance matrix. In that sense, the paper is actually providing an alternative CRB theory for sample covariance-based parameter estimators introducing the second-order Fisher information matrix and the concept of second-order efficiency.

APPENDIX A DEDUCTION OF $\mathbf{Q}(\theta)$

The expression of $\mathbf{Q}(\theta)$ in (33) can be written as follows:

$$\mathbf{Q}(\theta) = E\{\hat{\mathbf{r}}\hat{\mathbf{r}}^H\} - \mathbf{r}\mathbf{r}^H \quad (69)$$

where $\hat{\mathbf{r}} = \text{vec}\{(\mathbf{A}\mathbf{x} + \mathbf{w})(\mathbf{A}\mathbf{x} + \mathbf{w})^H\} = \text{vec}\{\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A}^H + \mathbf{A}\mathbf{x}\mathbf{w}^H + \mathbf{w}\mathbf{x}^H\mathbf{A}^H + \mathbf{w}\mathbf{w}^H\}$.

Taking into account the circularity and zero mean of the noise term, i.e., $E\{\mathbf{w}\} = \mathbf{0}$, $E\{\mathbf{w}\mathbf{w}^T\} = \mathbf{0}$ and $E\{w_i w_j^* w_k\} = E\{w_i^* w_j^* w_k\} = 0 \forall i, j, k$, only six terms, out of the 16 in $\hat{\mathbf{r}}\hat{\mathbf{r}}^H$, survive to the expectation in (69). They are classified next as follows:

- signal \times signal :
 $\text{vec}(\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A}^H)\text{vec}^H(\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A}^H)$
- signal \times noise :
 $\text{vec}(\mathbf{A}\mathbf{x}\mathbf{w}^H)\text{vec}^H(\mathbf{A}\mathbf{x}\mathbf{w}^H)$
 $+ \text{vec}(\mathbf{w}\mathbf{x}^H\mathbf{A}^H)\text{vec}^H(\mathbf{w}\mathbf{x}^H\mathbf{A}^H)$
 $+ \text{vec}(\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A}^H)\text{vec}^H(\mathbf{w}\mathbf{w}^H)$
 $+ \text{vec}(\mathbf{w}\mathbf{w}^H)\text{vec}^H(\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A}^H)$
- noise \times noise :
 $\text{vec}(\mathbf{w}\mathbf{w}^H)\text{vec}^H(\mathbf{w}\mathbf{w}^H)$.

Then, using the following three properties [9, ch. 2]

$$\text{vec}(\mathbf{A}\mathbf{B}\mathbf{C}^H) = (\mathbf{C}^* \otimes \mathbf{A})\text{vec}(\mathbf{B}) \quad (70)$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{A}\mathbf{C} \otimes \mathbf{B}\mathbf{D} \quad (71)$$

$$\begin{aligned} \text{vec}(\mathbf{a}\mathbf{b}^H)\text{vec}^H(\mathbf{a}\mathbf{b}^H) &= (\mathbf{b}^* \otimes \mathbf{a})(\mathbf{b}^* \otimes \mathbf{a})^H \\ &= (\mathbf{b}\mathbf{b}^H)^* \otimes \mathbf{a}\mathbf{a}^H \end{aligned} \quad (72)$$

and, bearing in mind that $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$, one obtains that

$$\begin{aligned} E\{\hat{\mathbf{r}}\hat{\mathbf{r}}^H\} &= \mathcal{A}E\{\text{vec}(\mathbf{x}\mathbf{x}^H)\text{vec}^H(\mathbf{x}\mathbf{x}^H)\}\mathcal{A}^H \\ &+ \mathbf{R}_w^* \otimes \mathbf{A}\mathbf{A}^H + (\mathbf{A}\mathbf{A}^H)^* \otimes \mathbf{R}_w \\ &+ \text{vec}(\mathbf{A}\mathbf{A}^H)\text{vec}^H(\mathbf{R}_w) + \text{vec}(\mathbf{R}_w)\text{vec}^H(\mathbf{A}\mathbf{A}^H) \\ &+ \mathbf{R}_w^* \otimes \mathbf{R}_w + \text{vec}(\mathbf{R}_w)\text{vec}^H(\mathbf{R}_w) \end{aligned} \quad (73)$$

where $\mathcal{A} \triangleq \mathbf{A}^* \otimes \mathbf{A}$, and the following property of Gaussian vectors is used (Appendix B):

$$\begin{aligned} E\{\text{vec}(\mathbf{w}\mathbf{w}^H)\text{vec}^H(\mathbf{w}\mathbf{w}^H)\} \\ = \mathbf{R}_w^* \otimes \mathbf{R}_w + \text{vec}(\mathbf{R}_w)\text{vec}^H(\mathbf{R}_w). \end{aligned} \quad (74)$$

Therefore, grouping terms in (73) and bearing in mind (11), the following result is obtained:

$$\begin{aligned} E\{\hat{\mathbf{r}}\hat{\mathbf{r}}^H\} &= \mathcal{A}E\{\text{vec}(\mathbf{x}\mathbf{x}^H)\text{vec}^H(\mathbf{x}\mathbf{x}^H)\}\mathcal{A}^H \\ &- (\mathbf{A}\mathbf{A}^H)^* \otimes \mathbf{A}\mathbf{A}^H - \text{vec}(\mathbf{A}\mathbf{A}^H)\text{vec}^H(\mathbf{A}\mathbf{A}^H) \\ &+ \mathbf{R}^* \otimes \mathbf{R} + \text{vec}(\mathbf{R})\text{vec}^H(\mathbf{R}). \end{aligned} \quad (75)$$

Finally, using once more (71) and (70) in order to write the negative terms above as a function of \mathcal{A} and plugging this result into (69), we get the expression proposed in (34).

APPENDIX B FOURTH-ORDER MOMENTS

In this section, the fourth-order moments of a generic *zero-mean, circular*, possibly non-Gaussian vector $\mathbf{v} \in \mathbb{C}^L$ are deduced. The resulting L^4 terms are ordered in the following matrix:

$$\mathbf{Q}_v = E\{\text{vec}(\mathbf{v}\mathbf{v}^H)\text{vec}^H(\mathbf{v}\mathbf{v}^H)\} \quad (76)$$

whose elements are given by

$$\begin{aligned} [\mathbf{Q}_v]_{i+Lj, k+Ll} &= E\{v_i v_j^* v_k^* v_l\} \\ &= E\{v_i v_j^*\}E\{v_k^* v_l\} + E\{v_i v_k^*\}E\{v_j^* v_l\} \\ &- (E\{\|v_i\|^4\} - 2E^2\{\|v_i\|^2\})\delta(i, j, k, l) \end{aligned} \quad (77)$$

where $i, j, k, l \in \{0, \dots, L-1\}$, and $\delta(i, j, k, l)$ is the Kronecker delta of multiple dimensions.

If all these elements are arranged in \mathbf{Q}_v , the following three terms are identified:

$$\mathbf{Q}_v = \text{vec}(\mathbf{R}_v)\text{vec}^H(\mathbf{R}_v) + \mathbf{R}_v^* \otimes \mathbf{R}_v + \text{diag}(\mathbf{\Gamma}) \quad (78)$$

where $\mathbf{R}_v \triangleq E\{\mathbf{v}\mathbf{v}^H\}$, and $\mathbf{\Gamma}$ is the diagonal matrices with $[\mathbf{\Gamma}]_{i,i} \triangleq E\{\|v_i\|^4\} - 2E^2\{\|v_i\|^2\}$.

If the elements of \mathbf{v} are identically distributed, $\mu \triangleq E\{\|v_i\|^2\}$ and $\rho \triangleq E\{\|v_i\|^4\}/\mu^2$ do not depend on i , and thus, the third term can be simplified to obtain that

$$\mathbf{Q}_v = \text{vec}(\mathbf{R}_v)\text{vec}^H(\mathbf{R}_v) + \mathbf{R}_v^* \otimes \mathbf{R}_v + \mu^2(\rho-2)\text{diag}(\mathbf{I}). \quad (79)$$

In particular, the fourth-order moments of \mathbf{x} in (36) are given by (79), bearing in mind that the symbols autocorrelation is $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$, and thus, $\mathbf{R}_v = \mathbf{I}$, $\mu = 1$. On the other hand, if \mathbf{v} is normally distributed, as the noise vector \mathbf{w} in the adopted signal model, the third term in (79) can be removed, taking into account that $\rho = 2$ in the Gaussian case, hence proving (74).

APPENDIX C SMALL-ERROR MATRICES

Let us define the matrices $\mathbf{S}(\boldsymbol{\theta})$ and $\tilde{\mathbf{Q}}(\boldsymbol{\theta})$ as the arguments inside the brackets in (26) and (27):

$$\begin{aligned} \mathbf{S}(\boldsymbol{\theta}) &\triangleq (\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r})(\boldsymbol{\psi}_t(\boldsymbol{\theta}) - \boldsymbol{\psi}_t)^H \\ \tilde{\mathbf{Q}}(\boldsymbol{\theta}) &\triangleq (\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r})(\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r})^H. \end{aligned} \quad (80)$$

Regarding matrix $\mathbf{S}(\boldsymbol{\theta})$, it is easy to show that

$$\begin{aligned} \mathbf{S}(\boldsymbol{\theta}_o) &= 0 \\ \frac{\partial^2 \mathbf{S}(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} &= \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_p} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \left(\frac{\partial \boldsymbol{\psi}_t(\boldsymbol{\theta})}{\partial \theta_q} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \right)^H \\ &+ \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_q} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \left(\frac{\partial \boldsymbol{\psi}_t(\boldsymbol{\theta})}{\partial \theta_p} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \right)^H \\ &= [\mathbf{D}_r]_p [\mathbf{D}_\psi]_q^H + [\mathbf{D}_r]_q [\mathbf{D}_\psi]_p^H \end{aligned} \quad (81)$$

since the pair of terms depending on $\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r}_o$ and $\boldsymbol{\psi}_t(\boldsymbol{\theta}) - \boldsymbol{\psi}_t$ vanish at $\boldsymbol{\theta} = \boldsymbol{\theta}_o$.

Then, (45) is obtained after substituting into (40) the following result:

$$\begin{aligned} \sum_{p,q=1}^P \frac{\partial^2 \mathbf{S}(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} [\mathbf{C}_\boldsymbol{\theta}]_{p,q} \\ = \sum_{p,q=1}^P [\mathbf{D}_r]_p [\mathbf{D}_\psi]_q^H [\mathbf{C}_\boldsymbol{\theta}]_{p,q} + \sum_{p,q=1}^P [\mathbf{D}_r]_q [\mathbf{D}_\psi]_p^H [\mathbf{C}_\boldsymbol{\theta}]_{p,q} \\ = \mathbf{D}_r \mathbf{C}_\boldsymbol{\theta} \mathbf{D}_\psi^H + \mathbf{D}_r \mathbf{C}_\boldsymbol{\theta}^T \mathbf{D}_\psi^H = 2\mathbf{D}_r \mathbf{C}_\boldsymbol{\theta} \mathbf{D}_\psi^H \end{aligned} \quad (82)$$

where the last equality is verified because $\mathbf{C}_\boldsymbol{\theta}$ is symmetric. Proceeding in the same way with matrix $\tilde{\mathbf{Q}}$, it is found that

$$\begin{aligned} \tilde{\mathbf{Q}}(\boldsymbol{\theta}_o) &= 0 \\ \frac{\partial^2 \tilde{\mathbf{Q}}(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} &= \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_p} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \left(\frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_q} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \right)^H \\ &+ \frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_q} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \left(\frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \theta_p} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \right)^H \\ &= [\mathbf{D}_r]_p [\mathbf{D}_r]_q^H + [\mathbf{D}_r]_q [\mathbf{D}_r]_p^H. \end{aligned} \quad (83)$$

Then, (44) is deduced after plugging into (40) the following expression:

$$\begin{aligned} & \sum_{p,q=1}^P \left. \frac{\partial^2 \tilde{Q}(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} [\mathbf{C}\boldsymbol{\theta}]_{p,q} \\ &= \sum_{p,q=1}^P [\mathbf{D}_r]_p [\mathbf{D}_r]_q^H [\mathbf{C}\boldsymbol{\theta}]_{p,q} + \sum_{p,q=1}^P [\mathbf{D}_r]_q [\mathbf{D}_r]_p^H [\mathbf{C}\boldsymbol{\theta}]_{p,q} \\ &= \mathbf{D}_r \mathbf{C}\boldsymbol{\theta} \mathbf{D}_r^H + \mathbf{D}_r \mathbf{C}\boldsymbol{\theta}^T \mathbf{D}_r^H = 2\mathbf{D}_r \mathbf{C}\boldsymbol{\theta} \mathbf{D}_r^H. \end{aligned} \quad (84)$$

REFERENCES

- [1] S. M. Kay, *Fundamentals of Statistical Signal Processing. Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [2] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.
- [3] B. Ottersten, M. Viberg, and P. Stoica, "Exact and large sample maximum likelihood techniques for parameter estimation and detection," in *Radar Array Processing*. New York: Springer-Verlag, 1993, ch. 4, pp. 99–151.
- [4] G. Vázquez and J. Riba, "Non-data-aided digital synchronization," in *Signal Processing Advances in Wireless Communications*. Englewood Cliffs, NJ: Prentice-Hall, 2000, vol. 2, ch. 9.
- [5] J. Riba, J. Sala, and G. Vázquez, "Conditional maximum likelihood recovery: Estimators and bounds," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 835–850, Apr. 2001.
- [6] J. Villares, G. Vázquez, and J. Riba, "Fourth-order nondata-aided synchronization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Salt Lake City, UT, May 2001.
- [7] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs, NJ: Prentice Hall, 1997.
- [8] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Crámer–Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 720–741, May 1989.
- [9] J. Villares and G. Vázquez, "Sample covariance matrix parameter estimation: Carrier frequency, a case study," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Hong Kong, Apr. 2003.
- [10] U. Mengali and A. D'Andrea, *Synchronization Techniques for Digital Receivers*. New York: Plenum, 1997.
- [11] P. Stoica and T. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 87–90, Jan. 2001.
- [12] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, "Extended Ziv–Zakai lower bound for vector parameter estimation," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 624–637, Mar. 1997.
- [13] L. T. McWhorter and L. L. Scharf, "Properties of quadratic covariance bounds," in *Proc. 27th Asilomar Conf. Signals, Syst., Comput.*, Monticello, IL, Nov. 1993.
- [14] E. Weinstein and A. J. Weiss, "A general class of lower bounds in parameter estimation," *IEEE Trans. Inf. Theory*, vol. 34, no. 2, pp. 338–342, Mar. 1988.
- [15] L. T. McWhorter and L. L. Scharf, "Geometry of the Cramer–Rao bound," in *Proc. 6th IEEE Workshop Statist. Signal Array Process.*, Victoria, BC, Canada, Oct. 1992.
- [16] J. Villares and G. Vázquez, "Second-order DOA estimation from digitally modulated signals," in *Proc. 37th IEEE Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 2003.
- [17] —, "Best quadratic unbiased estimator (BQUE) for timing and frequency synchronization," in *Proc. 11th IEEE Workshop Statist. Signal Process.*, Singapore, Aug. 2001.
- [18] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "The modified Cramer–Rao bound and its application to synchronization problems," *IEEE Trans. Commun.*, vol. 46, no. 11, pp. 1391–1399, Nov. 1998.
- [19] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. New York: Wiley, 1988.



Javier Villares (S'01) was born in Barcelona, Catalonia, Spain, in 1974. He received the M.S. degree in telecommunications engineering from the Technical University of Catalonia (UPC), Barcelona, in 1999.

He joined the Department of Signal Theory and Communications, UPC, in 1999, where he was a research assistant until September 2003 under a grant from the Catalan Government. Since September 2003, he has been an Assistant Professor at UPC. His research interests are in digital communications, statistical signal processing, and information theory.



Gregori Vazquez (SM'98) was born in Barcelona, Spain, in 1961. He received the M.S. and Ph.D. degrees in telecommunications engineering from the Technical University of Catalonia (UPC), Barcelona, in 1984 and 1988, respectively.

He is a Professor with the Department of Signal Theory and Communications, UPC. His general interests are statistical signal processing and digital communications.

Dr. Vazquez received the 2003 Best Paper Award of the IEEE Signal Processing Society. He has been serving as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING since 1999, also serving as a member of the Editorial Board of the IEEE SIGNAL PROCESSING MAGAZINE and as a member of the Signal Processing for Communications Technical Committee. He is the Chairman of the Research Program on Communications of the Spanish Science and Technology Ministry.