

The Gaussian Assumption in Second-Order Estimation Problems in Digital Communications

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

Abstract—This paper deals with the goodness of the Gaussian assumption when designing second-order blind estimation methods in the context of digital communications. The low- and high-signal-to-noise ratio (SNR) asymptotic performance of the maximum likelihood estimator—derived assuming Gaussian transmitted symbols—is compared with the performance of the optimal second-order estimator, which exploits the actual distribution of the discrete constellation. The asymptotic study concludes that the Gaussian assumption leads to the optimal second-order solution if the SNR is very low or if the symbols belong to a multilevel constellation such as quadrature-amplitude modulation (QAM) or amplitude-phase-shift keying (APSK). On the other hand, the Gaussian assumption can yield important losses at high SNR if the transmitted symbols are drawn from a constant modulus constellation such as phase-shift keying (PSK) or continuous-phase modulations (CPM). These conclusions are illustrated for the problem of direction-of-arrival (DOA) estimation of multiple digitally-modulated signals.

Index Terms—Blind estimation, constant modulus, direction-of-arrival (DOA) estimation, estimation bounds, non-data-aided, second-order techniques, self-noise.

I. INTRODUCTION

THIS paper focuses on the parameter estimation problem in digital communications and in particular on the role of the side-information provided by the prior knowledge of the transmitted signal constellation statistics in blind or non-data-aided (NDA) parameter estimation and digital synchronization.

In communication schemes where the received waveform is known, the digital demodulation requires the fine estimation of unknown parameters such as the signal amplitude, symbol timing and the carrier phase and frequency errors, among others [1]. If the transmission is performed in a dispersive channel, the detection task becomes more complex because the channel identification is usually required in order to perform maximum-likelihood (ML) sequence detection [2], [3, Sec. 10-1] at the receiver and/or for an optimal power/bit allocation at the transmitter when a feedback-channel is available in the communication system [4], [5].

In a more general approach, other relevant information about the received signal can be also of interest, as for instance, the propagation channel characterization in positioning

applications implies the estimation of the propagation delay, power-delay profile, angle-of-arrival, and other parameters [6].

In order to assist the receiver in the estimation task, most communication standards include known training sequences. The transmission of these spectrally inefficient training data can be avoided if NDA estimation methods are adopted [1]. In a blind estimation problem, the transmitted symbols become *nuisance parameters* [7, Sec. 10.7] that are modeled as independent, identically distributed (i.i.d.) random variables taking values in a discrete alphabet or constellation.

In this context, the ML theory states that quadratic estimators are asymptotically efficient in a low signal-to-noise ratio (SNR) scenario or if the transmitted symbols are Gaussian distributed [8], meaning that the Cramér–Rao bound (CRB) is attained when the number of observations goes to infinity.

Although the Gaussian distribution is not realistic in digital communications, it is usually adopted as a worst-case assumption for blind parameter estimation [9, Sec. VI] since it is the most uncertain (maximum entropy) variance-constrained distribution. Another attractive feature is that the Gaussian assumption leads to the best *robust*¹ estimator when the data length goes to infinity [9, Sec. VI]. Finally, in some scenarios, the Gaussian assumption is known to supply the optimal second-order estimator for infinite data records [10].

A. Contributions and Related Works

In this paper, the optimality of the Gaussian assumption when designing second-order estimation techniques for digital communication applications is reviewed. The analysis concludes that the Gaussian assumption applies when dealing with complex amplitude modulations such as quadrature-amplitude modulation (QAM) or amplitude phase-shift keying (APSK), independently of the SNR, even for short data records. On the other hand, the Gaussian assumption yields important performance losses in case of *constant modulus* (CM) modulations such as PSK and continuous-phase modulation (CPM) for practical values of the SNR. This means that the Gaussian assumption cannot exploit the CM property whereas the optimal second-order estimator [8] is implicitly using this property to decouple—with no losses—the contribution of all the parameters of interest. These relevant conclusions have been derived from the high-SNR asymptotic analysis of the optimal second-order estimator [8], and they have been validated in the direction-of-arrival (DOA) estimation problem.

Regarding the CM property, the constant modulus algorithm (CMA) was introduced by Treichler *et al.* [11] in the area of

¹An estimator is *robust* if its performance is independent of the statistical distribution of the observation.

Manuscript received July 19, 2006; revised January 13, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Geert Leus. This work has been partially financed by the Spanish/Catalan Science and Technology Commissions and FEDER funds from the European Commission: TEC2004-04526, TIC2003-05482, and 2005SGR-00639.

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

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2007.897872

adaptive equalization, and it has been widely studied in the last decade (see [12], [13], and references therein). The CMA has been later applied to blind parametric estimation problems in which the mixing matrix is structured (parameterized), and the aim is to determine the value of these parameters [14]. In this context, CMA is used to decouple the multivariate estimation task into separated uniparametric estimation problems. However, the CMA separation is carried out without exploiting the signal model structure, and it suffers from noise enhancement at low SNR due to the inherent fourth-order processing. In this sense, this contribution states that second-order parametric estimators, which are optimal at low SNR, can also exploit the CM property at medium-to-high SNR in order to cancel the mutual interference of the received superimposed signals.

B. Paper Organization and Notation

The work is organized as follows. In Section II, the signal model and some basic assumptions are introduced. From this background, in Section III the Gaussian maximum-likelihood (GML) estimator is formulated and evaluated considering Gaussian nuisance parameters. Next, in Section IV, the performance of the GML estimator is computed in case of non-Gaussian nuisance parameters. In addition, the optimal second-order estimator proposed in [8] is reviewed and evaluated. The performance of the Gaussian and optimal second-order estimators is analyzed when the SNR is asymptotically low (Section V) and asymptotically high (Section VI). The asymptotic results are then illustrated in Section VII for the problem of bearing estimation in digital cellular systems. Finally, the main conclusions of the paper are summarized in Section VIII.

The following notation is used in the paper. Superscripts “ T ”, “ $*$ ”, “ H ”, and “ $\#$ ” stand for matrix transpose, matrix conjugate, matrix transpose conjugate and Moore–Penrose pseudoinverse [15], respectively. Uppercase and lowercase boldface (or calligraphic) denote matrices and vectors, respectively. $[\mathbf{v}]_i$ and $[\mathbf{M}]_{i,j}$ are used to refer to the i th element of vector \mathbf{v} and the element in the i th row and j th column of matrix \mathbf{M} , respectively. The symbols $\text{vec}(\mathbf{M})$, $\text{Tr}(\mathbf{M})$, $\text{Re}(\mathbf{M})$, and \otimes stand for the columnwise vectorization, the trace, the real part operators, and the Kronecker product [15], respectively. The symbol \mathbf{I}_N denotes the $N \times N$ identity matrix. Finally, $E_{\mathbf{v}}\{\cdot\}$ stands for the expected value with respect to the random vector \mathbf{v} .

II. PROBLEM STATEMENT

This paper deals with the estimation of a vector of deterministic real parameters $\theta \in \mathbb{R}^P$ from a noisy observation $\mathbf{y} \in \mathbb{C}^M$ having the following linear model:

$$\mathbf{y} = \mathbf{A}(\theta)\mathbf{x} + \mathbf{w}$$

where $\mathbf{x} \in \mathbb{C}^K$ is the vector of unknown transmitted symbols (nuisance parameters), $\mathbf{w} \in \mathbb{C}^M$ is the usual additive white Gaussian noise (AWGN) and, $\mathbf{A}(\theta) \in \mathbb{C}^{M \times K}$ is an arbitrary but known mixing matrix parameterized by θ . This simple model encompasses many estimation problems in the context of digital communications and array signal processing.

Henceforth, we will assume that \mathbf{x} and \mathbf{w} are independent, zero-mean, stationary, spectrally white random vectors of known statistics. Also, we will assume that the noise is circular, i.e., $E\{\mathbf{w}\mathbf{w}^T\} = \mathbf{0}$ [16]. Notice that the model is general because the transmitted symbols correlation can always be included into matrix $\mathbf{A}(\theta)$. On the other hand, if the noise vector were correlated, all the studied estimators would include automatically a whitening matrix to decorrelate the observation noise [8].

III. GAUSSIAN NUISANCE PARAMETERS

If the nuisance parameters are Gaussian distributed, the maximum likelihood (ML) estimator is the one minimizing the following nonlinear cost function [8], [17]:

$$\Lambda_{\text{GML}}(\theta) = \text{Tr} \left(\ln \mathbf{R}(\theta) + \mathbf{R}^{-1}(\theta) \widehat{\mathbf{R}} \right) \quad (1)$$

with $\widehat{\mathbf{R}} = \mathbf{y}\mathbf{y}^H$ the sample covariance matrix and

$$\mathbf{R}(\theta) = \mathbf{A}(\theta)\mathbf{A}^H(\theta) + \sigma_w^2 \mathbf{I}_M$$

its expected value as a function of θ , in which σ_w^2 stands for the variance of the noise samples.

The minimizer of (1) is termed the GML estimator because it is derived assuming Gaussian nuisance parameters.²

The minimum of $\Lambda_{\text{GML}}(\theta)$ can be sought by means of iterative or time-recursive scoring methods based on the following recursion [7, Eq. 7.50], [17, Sec. 4.6], [18]:

$$\begin{aligned} \widehat{\theta}_{k+1} &= \widehat{\theta}_k + \mathbf{e}_{\text{GML}}(\mathbf{y}; \widehat{\theta}_k) \\ &= \widehat{\theta}_k + \mathbf{J}_{\text{GML}}^{-1}(\widehat{\theta}_k) \nabla_{\text{GML}}(\mathbf{y}; \widehat{\theta}_k) \end{aligned} \quad (2)$$

where $\mathbf{e}_{\text{GML}}(\mathbf{y}; \widehat{\theta}_k)$ is the correction term at iterate/time index k and

$$\begin{aligned} \nabla_{\text{GML}}(\mathbf{y}; \theta) &= \frac{d\Lambda_{\text{GML}}(\theta)}{d\theta} \\ \mathbf{J}_{\text{GML}}(\theta) &= -E_{\mathbf{y}} \left\{ \frac{d^2 \Lambda_{\text{GML}}(\theta)}{d\theta d\theta^T} \right\} \\ &= E_{\mathbf{y}} \left\{ \nabla_{\text{GML}}(\mathbf{y}; \theta) \nabla_{\text{GML}}^T(\mathbf{y}; \theta) \right\} \end{aligned}$$

are the gradient and the Fisher information matrix in the Gaussian case. After some algebraic manipulations [19, App. 2.G], it follows that

$$\begin{aligned} \nabla_{\text{GML}}(\mathbf{y}; \theta) &= \mathbf{D}_r^H(\theta) \mathcal{R}^{-1}(\theta) \text{vec} \left(\widehat{\mathbf{R}} - \mathbf{R}(\theta) \right) \\ \mathbf{J}_{\text{GML}}(\theta) &= \mathbf{D}_r^H(\theta) \mathcal{R}^{-1}(\theta) \mathbf{D}_r(\theta) \end{aligned}$$

where

$$[\mathbf{D}_r(\theta)]_p = \text{vec} \left(\frac{\partial \mathbf{R}(\theta)}{\partial \theta_p} \right)$$

is the p th column of matrix $\mathbf{D}_r(\theta)$ and

$$\mathcal{R}(\theta) = \mathbf{R}^*(\theta) \otimes \mathbf{R}(\theta)$$

²If the nuisance parameters are not Gaussian distributed (Section IV), the GML estimator can still be adopted but it is not the true ML solution in this case.

the covariance matrix of the quadratic innovation vector $\text{vec}(\widehat{\mathbf{R}} - \mathbf{R}(\theta))$ in the Gaussian case.

After convergence, if the magnitude of the correction term in (2) is sufficiently small, the estimator operates in the so-called *small-error* regime, in which ML estimators are unbiased and efficient, and hence the CRB theory applies. In order to satisfy the small-error condition, the observation length has to exceed a given threshold that is a decreasing function of the SNR. In the following, we will assume that the observation vector is sufficiently large to verify the small-error condition in the studied SNR interval.

Focusing on the small-error regime and based on the above discussion, it follows that the recursion in (2) converges to the true value of the parameter and attains the (Gaussian) unconditional Cramér–Rao bound (UCRB) [17, Eq. 4.64], as follows:

$$\begin{aligned} \mathbf{B}_{\text{GML}}(\theta) &= E_{\mathbf{y}} \{ \mathbf{e}_{\text{GML}}(\mathbf{y}; \theta) \mathbf{e}_{\text{GML}}^T(\mathbf{y}; \theta) \} \\ &= \mathbf{J}_{\text{GML}}^{-1}(\theta) = (\mathbf{D}_r^H(\theta) \mathcal{R}^{-1}(\theta) \mathbf{D}_r(\theta))^{-1} \\ &= \mathbf{B}_{\text{UCRB}}(\theta). \end{aligned} \quad (3)$$

IV. NON-GAUSSIAN NUISANCE PARAMETERS

If the nuisance parameters are not Gaussian distributed, the small-error performance of the GML estimator presented in last section is no longer given by the UCRB. In particular, we have that

$$\begin{aligned} \mathbf{B}_{\text{GML}}(\theta) &= E_{\mathbf{y}} \{ \mathbf{e}_{\text{GML}}(\mathbf{y}; \theta) \mathbf{e}_{\text{GML}}^T(\mathbf{y}; \theta) \} \\ &= \mathbf{J}_{\text{GML}}^{-1}(\theta) \mathbf{D}_r^H(\theta) \mathcal{R}^{-1}(\theta) \mathbf{Q}(\theta) \\ &\quad \times \mathcal{R}^{-1}(\theta) \mathbf{D}_r(\theta) \mathbf{J}_{\text{GML}}^{-1}(\theta) \end{aligned} \quad (4)$$

where $\mathbf{Q}(\theta)$ is the covariance matrix of the innovation vector $\text{vec}(\widehat{\mathbf{R}} - \mathbf{R}(\theta))$ when the nuisance parameters are not Gaussian distributed. The closed-form expression of $\mathbf{Q}(\theta)$ was derived in [8], [13], obtaining that

$$\mathbf{Q}(\theta) = \mathcal{R}(\theta) + \mathcal{A}(\theta) \mathbf{K} \mathcal{A}^H(\theta) \quad (5)$$

where $\mathcal{A}(\theta) = \mathbf{A}^*(\theta) \otimes \mathbf{A}(\theta)$ and matrix \mathbf{K} contains all the fourth-order cumulants (kurtosis) of the vector of nuisance parameters [8]. The *kurtosis matrix* \mathbf{K} is null in the Gaussian case and provides the complete non-Gaussian information on the nuisance parameters that second-order NDA estimators are able to exploit [8]. In case of circular complex nuisance parameters, \mathbf{K} is given by the following diagonal matrix:

$$\mathbf{K} = (\rho - 2) \text{diag}(\text{vec}(\mathbf{I}_K)) \quad (6)$$

where ρ is the fourth-order moment of the nuisance parameters [8], [13]. It can be shown that $\rho = 1$ for PSK constellations [13], $1 < \rho < 2$ for QAM and APSK constellations and $\rho = 2$ in the Gaussian case. If the nuisance parameters are noncircular, e.g., in case of binary-phase-shift keying (BPSK) or CPM modulations, the exact \mathbf{K} can be obtained numerically.

Finally, if (5) is plugged into (4), the GML covariance matrix can be written as

$$\mathbf{B}_{\text{GML}}(\theta) = \mathbf{B}_{\text{UCRB}}(\theta) + \mathbf{X}_{\text{GML}}(\theta) \quad (7)$$

where the second term

$$\begin{aligned} \mathbf{X}_{\text{GML}}(\theta) &\triangleq \mathbf{J}_{\text{GML}}^{-1}(\theta) \mathbf{D}_r^H(\theta) \mathcal{R}^{-1}(\theta) \mathcal{A}(\theta) \mathbf{K} \mathcal{A}^H(\theta) \\ &\quad \times \mathcal{R}^{-1}(\theta) \mathbf{D}_r(\theta) \mathbf{J}_{\text{GML}}^{-1}(\theta) \end{aligned} \quad (8)$$

is normally negative definite in digital communications due to the sign of matrix \mathbf{K} (6).

The important question that arises at this point is whether there is any other second-order method outperforming the GML estimator when the nuisance parameters are not Gaussian. This question was solved in [8] concluding that the GML estimator (2) can be improved if $\mathcal{R}(\theta)$ is replaced in (2) by the correct innovation covariance matrix $\mathbf{Q}(\theta)$ (5), in which the actual distribution of \mathbf{x} is considered. In that case, the optimum second-order estimator, which was originally named the best quadratic unbiased estimator (BQUE) [20], is computed by means of the following recursion:

$$\begin{aligned} \widehat{\theta}_{k+1} &= \widehat{\theta}_k + \mathbf{e}_{\text{BQUE}}(\mathbf{y}; \widehat{\theta}_k) \\ &= \widehat{\theta}_k + \mathbf{J}_{\text{BQUE}}^{-1}(\widehat{\theta}_k) \nabla_{\text{BQUE}}(\mathbf{y}; \widehat{\theta}_k) \end{aligned} \quad (9)$$

where

$$\begin{aligned} \nabla_{\text{BQUE}}(\mathbf{y}; \theta) &= \mathbf{D}_r^H(\theta) \mathbf{Q}^{-1}(\theta) \text{vec}(\widehat{\mathbf{R}} - \mathbf{R}(\theta)) \\ \mathbf{J}_{\text{BQUE}}(\theta) &= \mathbf{D}_r^H(\theta) \mathbf{Q}^{-1}(\theta) \mathbf{D}_r(\theta) \end{aligned}$$

are the gradient vector and Fisher information matrix in the non-Gaussian case. Consequently, the covariance of any second order estimator is lower bounded by

$$\begin{aligned} \mathbf{B}_{\text{BQUE}}(\theta) &= E_{\mathbf{y}} \{ \mathbf{e}_{\text{BQUE}}(\mathbf{y}; \theta) \mathbf{e}_{\text{BQUE}}^T(\mathbf{y}; \theta) \} \\ &= \mathbf{J}_{\text{BQUE}}^{-1}(\theta) \\ &= (\mathbf{D}_r^H(\theta) \mathbf{Q}^{-1}(\theta) \mathbf{D}_r(\theta))^{-1}. \end{aligned} \quad (10)$$

The last lower bound was already proposed by Porat *et al.* in [21], [22] and extended to noncircular data by Delmas in [23]. In the latter work, a “covariance matching” method is formulated in terms of $\mathbf{Q}(\theta)$ to attain asymptotically $\mathbf{B}_{\text{BQUE}}(\theta)$ in case of non-Gaussian data (see also [21], [22], and [24]). In the context of DOA estimation, other closed-form expressions have been obtained in case of constant-modulus sources [14] and for minimum-shift keying (MSK) and binay/quaternary PSK transmitters [25].

V. LOW-SNR STUDY

When the noise variance goes to infinity ($\sigma_w^2 \rightarrow \infty$), it is rather straightforward to find the following low-SNR asymptotic limits:

$$\begin{aligned} \mathbf{B}_{\text{BQUE}}(\theta) &= \mathbf{B}_{\text{GML}}(\theta) = \mathbf{B}_{\text{UCRB}}(\theta) \\ &= \sigma_w^4 (\mathbf{D}_r^H(\theta) \mathbf{D}_r(\theta))^{-1} + o(\sigma_w^4) \end{aligned} \quad (11)$$

in which the Landau symbol $o(x)$ collects all the terms that converge to zero faster than x [19, Sec. 7.2].

Therefore, it follows that the actual distribution of the nuisance parameters becomes irrelevant at low SNR when designing second-order schemes. Accordingly, the Gaussian—or any other assumption about the distribution of the nuisance

parameters—yields the same asymptotic error covariance matrix at low SNR. Furthermore, since the GML is the true ML estimator at low SNR [19, App. 2.D], (11) is the true CRB when the SNR goes to zero.

VI. HIGH-SNR STUDY

In the low-SNR regime (Section V), GML estimation performs like the optimal second-order estimator [8]. However, for higher SNRs, the optimal second-order estimator makes use of the fourth-order statistical information about the nuisance parameters given by matrix \mathbf{K} [8], whereas this prior information about the signal constellation is not used by the GML scheme, implying a significant degradation in performance in the high-SNR regime. In this sense, the asymptotic analysis of (4) and (10) for $\sigma_w^2 \rightarrow 0$ will supply an upper bound in the performance degradation of the GML criterion with respect to the optimal second-order scheme.

The new asymptotic analysis implies the study of matrices $\mathcal{R}^{-1}(\theta)$ and $\mathbf{Q}^{-1}(\theta)$ in the limit case for $\sigma_w^2 \rightarrow 0$. It can be shown (Appendix I) that $\mathcal{R}^{-1}(\theta)$ becomes

$$\begin{aligned} \mathcal{R}^{-1}(\theta) &= \sigma_w^{-4} [\mathbf{P}_A^{\perp*}(\theta) \otimes \mathbf{P}_A^{\perp}(\theta)] \\ &+ \sigma_w^{-2} [\mathbf{B}^*(\theta) \otimes \mathbf{P}_A^{\perp}(\theta) + \mathbf{P}_A^{\perp*}(\theta) \otimes \mathbf{B}(\theta)] \\ &+ \mathbf{B}^*(\theta) \otimes \mathbf{B}(\theta) - \sigma_w^2 \mathbf{B}^*(\theta) \otimes \mathbf{B}^2(\theta) \\ &- \sigma_w^2 [\mathbf{B}^2(\theta)]^* \otimes \mathbf{B}(\theta) + \Gamma(\theta) + o(\sigma_w^2) \end{aligned} \quad (12)$$

where

$$\begin{aligned} \mathbf{P}_A^{\perp}(\theta) &\triangleq \mathbf{I}_M - \mathbf{A}(\theta)\mathbf{A}^{\#}(\theta) \\ \mathbf{B}(\theta) &\triangleq [\mathbf{A}^{\#}(\theta)]^H \mathbf{A}^{\#}(\theta) \end{aligned} \quad (13)$$

and $\Gamma(\theta)$ stands for the three terms in the expansion of $\mathcal{R}^{-1}(\theta)$ that are irrelevant in the high-SNR study (Appendix I).

The key property of $\mathcal{R}^{-1}(\theta)$ is that the term proportional to σ_w^{-4} is asymptotically orthogonal to $\text{vec}(\mathbf{A}(\theta)\mathbf{M})$, $\text{vec}(\mathbf{M}\mathbf{A}^H(\theta))$, $\mathbf{M} \otimes \mathbf{A}(\theta)$ and $\mathbf{A}^*(\theta) \otimes \mathbf{M}$ for any matrix \mathbf{M} . In particular, this is true for the matrix of derivatives $\mathbf{D}_r(\theta)$ and for $\mathcal{A}(\theta) = \mathbf{A}^*(\theta) \otimes \mathbf{A}(\theta)$. In the same manner, the first term on σ_w^{-2} is asymptotically orthogonal to $\text{vec}(\mathbf{A}(\theta)\mathbf{M})$ and $\mathbf{M} \otimes \mathbf{A}(\theta)$ whereas the second one is orthogonal to $\text{vec}(\mathbf{M}\mathbf{A}^H(\theta))$ and $\mathbf{A}^*(\theta) \otimes \mathbf{M}$. Consequently, it follows that

$$\mathcal{A}^H(\theta)\mathcal{R}^{-1}(\theta) = \mathcal{A}^{\#}(\theta) + o(1). \quad (14)$$

Focusing now on matrix $\mathbf{Q}^{-1}(\theta)$, its asymptotic value depends on the kurtosis matrix \mathbf{K} . The complete study is carried out in Appendix II, obtaining the following asymptotic expression:

$$\mathbf{Q}^{-1}(\theta) = \mathcal{R}^{-1}(\theta) + \sigma_w^{-2} [\mathcal{A}^{\#}(\theta)]^H \mathbf{P}_K^{\perp}(\theta) \mathcal{A}^{\#}(\theta) + o(\sigma_w^{-2}) \quad (15)$$

where $\mathbf{P}_K^{\perp}(\theta)$ is the projector onto the subspace generated by the eigenvectors of \mathbf{K} associated to the eigenvalue -1 .

The second term in the last expression is positive semidefinite, and it becomes zero if and only if $\mathbf{P}_K^{\perp}(\theta) = \mathbf{0}$, i.e., if all the eigenvalues of \mathbf{K} are different from -1 . The rank of $\mathbf{P}_K^{\perp}(\theta)$

is thus essential to assess the gain of considering the kurtosis matrix \mathbf{K} in the design of second-order estimators. The exact expression of $\mathbf{P}_K^{\perp}(\theta)$ is given in Appendix II.

Next, the asymptotic values of $\mathcal{R}^{-1}(\theta)$ and $\mathbf{Q}^{-1}(\theta)$ are used to obtain the high-SNR limit of $\mathbf{B}_{\text{UCRB}}(\theta)$, $\mathbf{B}_{\text{GML}}(\theta)$ and $\mathbf{B}_{\text{BQUE}}(\theta)$.

A. (Gaussian) Unconditional Cramér–Rao Bound

In general, the (Gaussian) UCRB becomes proportional to σ_w^2 when the SNR increases, obtaining the following asymptotic expression:

$$\mathbf{B}_{\text{UCRB}}(\theta) = \sigma_w^2 \mathcal{B}_1^{-1}(\theta) + o(\sigma_w^2) \quad (16)$$

where $\mathcal{B}_1(\theta)$ stands for the high-SNR limit of $\sigma_w^2 \mathbf{D}_r^H(\theta) \mathcal{R}^{-1}(\theta) \mathbf{D}_r(\theta)$. Using (14), the entries of $\mathcal{B}_1(\theta)$ are determined in [19, App. 7.F], obtaining

$$[\mathcal{B}_1(\theta)]_{p,q} = 2\text{ReTr} \left(\frac{\partial \mathbf{A}^H(\theta)}{\partial \theta_p} \mathbf{P}_A^{\perp}(\theta) \frac{\partial \mathbf{A}(\theta)}{\partial \theta_q} \right). \quad (17)$$

Notice that this result requires that $\partial \mathbf{A}(\theta)/\partial \theta_p$ does not lie totally on the subspace generated by the columns of $\mathbf{A}(\theta)$, i.e.,

$$\mathbf{P}_A^{\perp}(\theta) \frac{\partial \mathbf{A}(\theta)}{\partial \theta_p} \neq \mathbf{0} \quad (18)$$

for all the parameters $\theta_1, \dots, \theta_P$. For example, this condition is not verified in the problem of carrier phase synchronization [26] or in those scenarios in which the noise subspace of matrix $\mathbf{A}(\theta)$ is null (i.e., $M \leq K$). In both cases, the constant term $\mathbf{B}^*(\theta) \otimes \mathbf{B}(\theta)$ in (12) has to be considered in order to evaluate the variance floor caused by the *self-noise*³ at high SNR, obtaining that

$$\lim_{\sigma_w^2 \rightarrow 0} \mathbf{B}_{\text{UCRB}}(\theta) = \mathcal{B}_2^{-1}(\theta) \quad (19)$$

where $\mathcal{B}_2(\theta)$ stands for the high-SNR limit of $\mathbf{D}_r^H(\theta) \mathcal{R}^{-1}(\theta) \mathbf{D}_r(\theta)$. After some algebraic manipulations [19, App. 7.G], the entries of $\mathcal{B}_2(\theta)$ are given by

$$\begin{aligned} [\mathcal{B}_2(\theta)]_{p,q} &\triangleq 2\text{ReTr} \left(\frac{\partial \mathbf{A}(\theta)}{\partial \theta_p} \mathbf{A}^{\#}(\theta) \frac{\partial \mathbf{A}(\theta)}{\partial \theta_q} \mathbf{A}^{\#}(\theta) \right) \\ &+ 2\text{ReTr} \left(\frac{\partial \mathbf{A}(\theta)}{\partial \theta_p} \frac{\partial \mathbf{A}^H(\theta)}{\partial \theta_q} \mathbf{B}(\theta) \right). \end{aligned}$$

B. Gaussian Maximum Likelihood

In most estimation problems, the UCRB takes the form given in (16) and self-noise free estimation is possible with Gaussian nuisance parameters. In that case, the asymptotic performance of the GML estimator is exactly the one given in (16) irrespective of the actual distribution of the nuisance parameters, i.e., even if $\mathbf{K} \neq \mathbf{0}$. Formally, we have that

$$\mathbf{B}_{\text{GML}}(\theta) = \mathbf{B}_{\text{UCRB}}(\theta) = \sigma_w^2 \mathcal{B}_1^{-1}(\theta) + o(\sigma_w^2).$$

³*Self-noise* refers to the estimator variability caused by the random nuisance parameters. The self-noise comes from the mutual coupling of the nuisance parameters in the observation model and is related to the intersymbol and multiple access interference in digital communication systems. The self-noise term is manifested as a variance floor when the noise variance goes to zero.

This statement is true because the term $\mathbf{X}_{\text{GML}}(\theta)$ in (7) can be neglected since it depends on σ_w^4 whereas $\mathbf{B}_{\text{UCRB}}(\theta)$ is proportional to σ_w^2 taking into account (14).

Finally, if $\partial\mathbf{A}(\theta)/\partial\theta_p$ and $\mathbf{A}(\theta)$ lie in the same subspace and condition (18) is not held, the GML performance exhibits a variance floor (self-noise) at high SNR that is a function of the kurtosis matrix \mathbf{K} . Using (14) and (19) to simplify $\mathbf{X}_{\text{GML}}(\theta)$ in (8), it follows that the GML variance floor is given by

$$\lim_{\sigma_w^2 \rightarrow 0} \mathbf{B}_{\text{GML}}(\theta) = \mathcal{B}_2^{-1}(\theta) + \mathcal{B}_2^{-1}(\theta) \times \left(\mathbf{D}_r^H(\theta) [\mathcal{A}^\#(\theta)]^H \mathbf{K} \mathcal{A}^\#(\theta) \mathbf{D}_r(\theta) \right) \mathcal{B}_2^{-1}(\theta).$$

C. Best Quadratic Unbiased Estimator

In this section, closed-form expressions are obtained for the ultimate performance of second-order small-error estimators at high SNR. The study of $\mathbf{Q}^{-1}(\theta)$ in Appendix II comes to the conclusion that the Gaussian assumption is optimal at high SNR unless some eigenvalues of the kurtosis matrix \mathbf{K} are equal to -1 . This condition is closely related to the constant modulus property of the nuisance parameters [13]. To validate this result, let us obtain the asymptotic expression of $\mathbf{B}_{\text{BQUE}}(\theta)$ (10) as the noise variance goes to zero ($\sigma_w^2 \rightarrow 0$).

Using the asymptotic value of $\mathbf{Q}^{-1}(\theta)$ (15), it follows that

$$\mathbf{B}_{\text{BQUE}}(\theta) = \sigma_w^2 \left(\mathcal{B}_1(\theta) + \mathbf{D}_r^H(\theta) [\mathcal{A}^\#(\theta)]^H \mathbf{P}_{\mathbf{K}}^\perp(\theta) \mathcal{A}^\#(\theta) \mathbf{D}_r(\theta) \right)^{-1} + o(\sigma_w^2) \quad (20)$$

where $\mathcal{B}_1(\theta)$ stands for the limit of $\mathbf{D}_r^H(\theta) \mathcal{R}^{-1}(\theta) \mathbf{D}_r(\theta)$ (17) and $\mathbf{P}_{\mathbf{K}}^\perp(\theta) \in \mathbb{R}^{K^2 \times K^2}$ is the projector onto the subspace generated by the eigenvectors of \mathbf{K} associated to the eigenvalue -1 (Appendix II). Since the second term inside the inverse is always positive semidefinite, we can state at high SNR that

$$\mathbf{B}_{\text{BQUE}}(\theta) \leq \mathbf{B}_{\text{GML}}(\theta) = \mathbf{B}_{\text{UCRB}}(\theta).$$

Conversely, the second term inside the inverse is zero and, therefore, the Gaussian assumption applies at high SNR in any of the following situations:

- 1) **Signal parameterization.** The Gaussian assumption applies at high SNR if $\partial\mathbf{A}(\theta)/\partial\theta_p$ lies *totally* in the noise subspace of $\mathbf{A}(\theta)$, i.e.,

$$\frac{\partial\mathbf{A}(\theta)}{\partial\theta_p} = \mathbf{P}_{\mathbf{A}}^\perp(\theta) \frac{\partial\mathbf{A}(\theta)}{\partial\theta_p}$$

or, taking into account the definition of $\mathbf{P}_{\mathbf{A}}^\perp(\theta)$ (13)

$$\mathbf{A}^H(\theta) \frac{\partial\mathbf{A}(\theta)}{\partial\theta_p} = \mathbf{0}. \quad (21)$$

In that case, after some simple manipulations, it can be shown that $\mathcal{A}^\#(\theta) \mathbf{D}_r(\theta) = \mathbf{0}$ and, thus, the second term in (20) is strictly zero independently of the nuisance parameters distribution. For example, this condition usually applies in digital synchronization in TDMA systems, or in continuous transmission systems if the observation

length tends to infinity [19, Sec. 6.1.2.]. This condition was already pointed out by Delmas in [10]. By comparing this condition and the one introduced in (18), we can conclude that (21) never applies if the UCRB and GML suffer from self-noise at high SNR because, in that case, $\mathbf{P}_{\mathbf{A}}^\perp(\theta) \partial\mathbf{A}(\theta)/\partial\theta_p = \mathbf{0}$.

- 2) **Nuisance parameters distribution.** Regardless of the signal parametrization, the Gaussian assumption applies at high SNR if all the eigenvalues of the kurtosis matrix \mathbf{K} are different from -1 . In that case, $\mathbf{P}_{\mathbf{K}}^\perp(\theta)$ is strictly zero, and the second term in (20) becomes zero. If the nuisance parameters are drawn from an arbitrary *circular* complex alphabet, the kurtosis matrix is given by (6) and, therefore, the Gaussian assumption always applies *except if* $\rho = 1$. It can be shown that this condition ($\rho = 1$) is solely verified in case of a constant modulus alphabet [13]. Accordingly, in the context of digital communications, the Gaussian assumption applies for any multilevel linear modulation such as QAM or APSK, and it does not apply in case of PSK constellations ($\rho = 1$). If the nuisance parameters are not circular, there is not a simple closed-form expression for the eigenvalues of \mathbf{K} . However, it is found that the kurtosis matrix of some important constant modulus modulations show some eigenvalues equal to -1 . Among them, the CPM family deserves special attention [1], [3]. Other important constant modulus *noncircular* modulations are BPSK [25], [27] and those staggered modulations such as, for example, offset QPSK [28]. Finally, in those scenarios in which the UCRB and the GML exhibit a variance floor at high SNR because (18) is not satisfied, the second term in (20) allows canceling the self-noise, obtaining that

$$\mathbf{B}_{\text{BQUE}}(\theta) = \sigma_w^2 \left(\mathbf{D}_r^H(\theta) [\mathcal{A}^\#(\theta)]^H \mathbf{P}_{\mathbf{K}}^\perp(\theta) \mathcal{A}^\#(\theta) \mathbf{D}_r(\theta) \right)^{-1} + o(\sigma_w^2).$$

This situation arises in the carrier phase estimation problem addressed in [26] and [28], as well as in some singular scenarios in which $\mathbf{A}(\theta)$ is not full-column rank.

VII. CASE STUDY

In this section, the analytical results derived in the paper are illustrated for the problem of DOA estimation in the context of cellular communications. The scenario being analyzed is one in which two static users transmit from the far field towards a base station equipped with a calibrated linear array of M elements spaced half the carrier wavelength. The two users impinge into the array with the same power and with angles-of-arrival $\pm\phi$ degrees with respect to the broadside. We consider that the transmitted signals are digitally modulated, and N snapshots are collected at the matched filter output at one sample per symbol assuming perfect timing synchronization and ISI-free received pulses.

The performance of the GML and BQUE estimators is compared in terms of the aggregated *small-error* variance of the two users, i.e., $\text{Tr}(\mathbf{B}_{\text{GML}}(\theta))$ and $\text{Tr}(\mathbf{B}_{\text{BQUE}}(\theta))$ from (4) and (10), respectively. These figures of merit are evaluated as a function

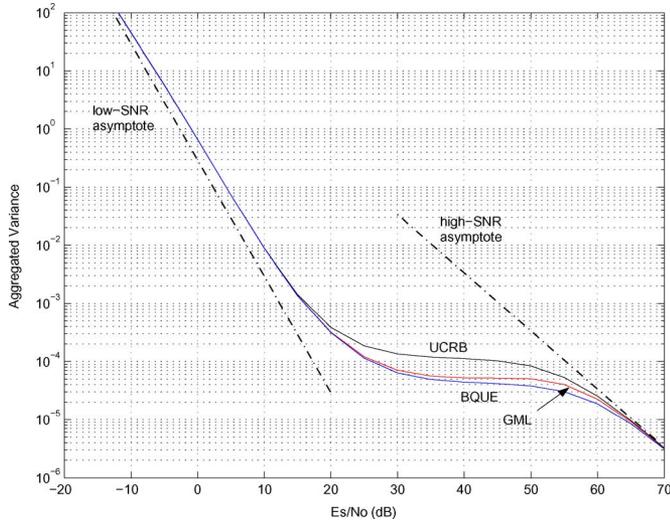


Fig. 1. Aggregated small-error variance for the GML and BQUE DOA estimators in case of having two 16-QAM sources transmitting from $\pm 0.5^\circ$ and an array of $M = 4$ antennas. The (Gaussian) UCRB and the low- and high-SNR asymptotes are plotted as well.

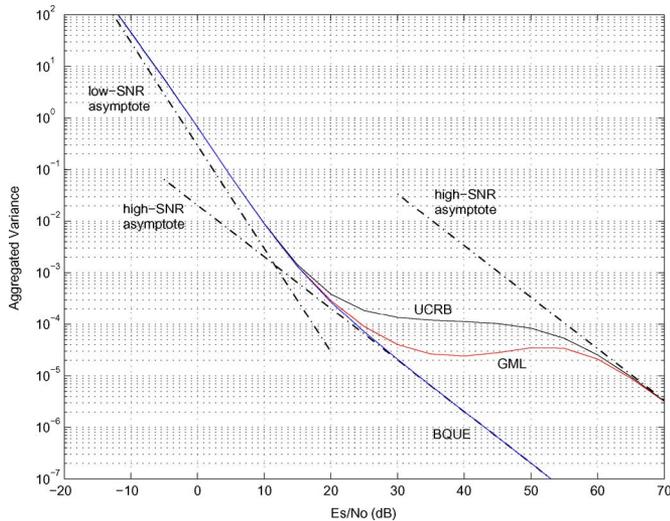


Fig. 2. Aggregated small-error variance for the GML and BQUE DOA estimators in case of having two QPSK sources transmitting from $\pm 0.5^\circ$ and an array of $M = 4$ antennas. The (Gaussian) UCRB and the low- and high-SNR asymptotes are plotted as well.

of the SNR (per user) at the matched filter output (E_s/N_0), the number of antennas (M), and the angular separation 2ϕ . Results will be presented considering a single snapshot taking into account that the variance of the studied estimators is inversely proportional to N . Consequently, the comparative study in this section is still valid for $N \rightarrow \infty$. Anyway, we consider that N is sufficiently large to satisfy the small-error condition for any studied E_s/N_0 .

In Figs. 1 and 2, it is shown how the asymptotic closed-form expressions deduced in the paper predict exactly the low- and high-SNR small-error performance of the studied quadratic estimators. In Fig. 1, the Gaussian assumption is shown to be optimal at low and high SNR in case of 16-QAM users, whereas minor performance losses are observed in the medium SNR range. On the other hand, if the modulation is QPSK (Fig. 2), the Gaussian

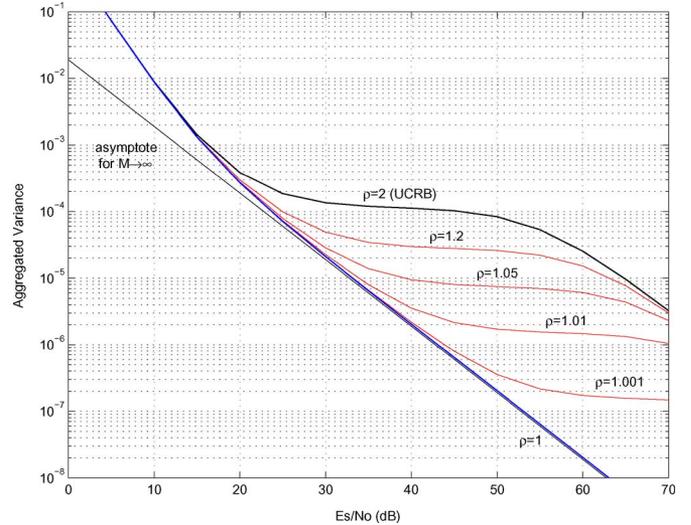


Fig. 3. Aggregated small-error variance for the BQUE DOA estimator for different values of ρ . The simulation parameters are $N = 1$, $M = 4$ and $\phi = 0.5^\circ$.

assumption is found to yield important performance losses when the SNR exceeds a given critical value or threshold determined by the array size and the users angular separation. Regarding the UCRB, it only predicts the low- and high-SNR performance of the GML, whereas it is outperformed for any intermediate SNR.

Additional studies have been carried out for CPM and larger PSK modulations obtaining exactly the same curves as in Fig. 2. The reason is that the studied second-order techniques do not exploit the nonzero improper covariance matrix $E\{\mathbf{y}\mathbf{y}^T\}$ of CPM modulations [23], [25].

It is known that the Gaussian assumption becomes optimal if there is a single user or, alternatively, the number of antennas grows to infinity ($M \rightarrow \infty$) [29, Th. 4.3], [30, Result R8]. In both cases, it follows that

$$\mathbf{B}_{\text{UCRB}}(\boldsymbol{\theta}) = \mathbf{B}_{\text{GML}}(\boldsymbol{\theta}) = \mathbf{B}_{\text{BQUE}}(\boldsymbol{\theta}) = \frac{6}{\pi^2 M^3 N E_s / N_0} \mathbf{I}_P + o(N^{-1} M^{-3}). \quad (22)$$

In Fig. 3, it is shown that the BQUE attains this asymptotic performance—for practical values of the SNR—in case of constant amplitude nuisance parameters ($\rho = 1$) even if there are multiple users and the number of antennas is very small ($M = 4$). This result is coherent with the CRB derived in [14]. On the other hand, if $\rho > 1$, the BQUE performance converges at high SNR to the (Gaussian) UCRB, which corresponds to $\rho = 2$. It can be seen that the larger ρ is, the lower is the E_s/N_0 from which the convergence to the UCRB is manifested. Additional analyses show that the closer the users, the more significant is the loss incurred by the Gaussian assumption in case of constant modulus nuisance parameters, and the lower is the E_s/N_0 from which this loss is manifested.

These conclusions are manifested again when the estimator variance is evaluated as a function of M for very high SNR (Figs. 4 and 5). It can be seen that the UCRB converges to (22) as the number of antennas increases. On the other hand, when the nuisance parameters are constant modulus ($\rho = 1$), the optimal

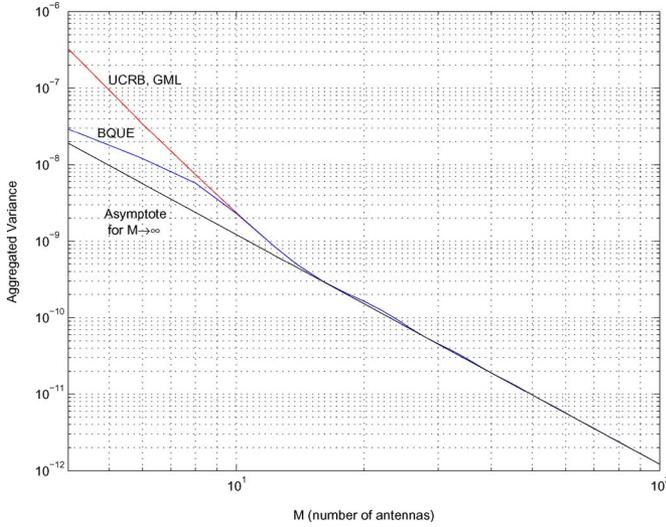


Fig. 4. Aggregated small-error variance for the BQUE DOA estimator as a function of M . The simulation parameters are QPSK, $N = 1$, $E_s/N_0 = 60$ dB and $\phi = 5^\circ$.

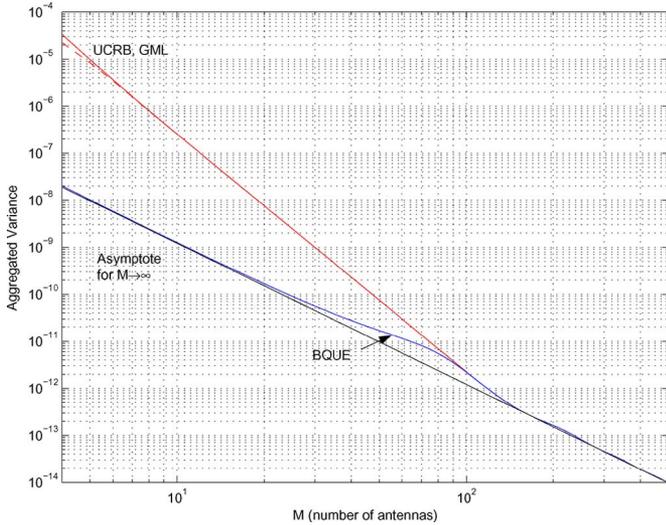


Fig. 5. Aggregated small-error variance for the BQUE DOA estimator as a function of M . The simulation parameters are QPSK, $N = 1$, $E_s/N_0 = 60$ dB and $\phi = 0.5^\circ$.

second-order estimator attains (22) for any value of M , except for an intermediate interval in which the estimator converges to the UCRB. It can be shown that the value of M from which $\mathbf{B}_{\text{BQUE}}(\theta)$ departs from (22) is inversely proportional to the angular separation of the users [19, Sec. 7.5.3]

VIII. CONCLUSION

The Gaussian assumption has been studied in the context of digital communications concluding that it always applies at low SNR whereas it only applies at high SNR for multilevel modulations (e.g., QAM and APSK). Thus, the Gaussian assumption is shown to yield suboptimal second-order estimates at high SNR in case of constant modulus alphabets (e.g., PSK and CPM). This conclusion has been drawn from the eigenanalysis of the kurtosis matrix \mathbf{K} . This matrix appears naturally when the optimal second-order estimator is formulated and it contains the

complete non-Gaussian information about the nuisance parameters that second-order estimators are able to exploit [8]. Regarding the eigenstructure of \mathbf{K} , it has been proven that the Gaussian assumption always applies at high SNR except if some eigenvalues of \mathbf{K} are equal to -1 . In fact, this simple condition is equivalent to the constant modulus property [13]. In the constant-modulus case, matrix \mathbf{K} allows canceling out the mutual interference among the unknown nuisance parameters in the medium-to-high SNR interval.

In the problem of DOA estimation of multiple digitally modulated signals, the constant modulus property can be used to mitigate the multiple access interference and increase the effective array resolution. In that sense, if the constant modulus property is exploited through matrix \mathbf{K} , it is possible to attain the large-array asymptotic performance with a reduced number of antennas [14]. On the other hand, if the Gaussian assumption is imposed by forcing $\mathbf{K} = \mathbf{0}$, important losses are exhibited at high SNR when the array is small and the users impinge from near directions. Notice that these conclusions are also valid if the number of snapshots goes to infinity.

APPENDIX I

HIGH-SNR LIMIT OF $\mathcal{R}^{-1}(\theta)$

If $\mathbf{A}(\theta)$ is full column rank, the asymptotic value of $\mathbf{R}^{-1}(\theta)$ is found using the inversion lemma

$$\begin{aligned} \mathbf{R}^{-1}(\theta) &= (\mathbf{A}(\theta)\mathbf{A}^H(\theta) + \sigma_w^2\mathbf{I}_M)^{-1} \\ &= \sigma_w^{-2}\mathbf{I}_M - \sigma_w^{-2}\mathbf{A}(\theta) (\mathbf{A}^H(\theta)\mathbf{A}(\theta) + \sigma_w^2\mathbf{I}_K)^{-1} \\ &\quad \times \mathbf{A}^H(\theta). \end{aligned} \quad (23)$$

At high SNR, the inverse of $\mathbf{A}^H(\theta)\mathbf{A}(\theta) + \sigma_w^2\mathbf{I}_K$ can be expanded in a Taylor series around $\sigma_w^2 = 0$, yielding⁴

$$\begin{aligned} [\mathbf{A}^H(\theta)\mathbf{A}(\theta)]^{-1} - \sigma_w^2 [\mathbf{A}^H(\theta)\mathbf{A}(\theta)]^{-2} \\ + \sigma_w^4 [\mathbf{A}^H(\theta)\mathbf{A}(\theta)]^{-3} + o(\sigma_w^4). \end{aligned}$$

Finally, after plugging these three terms into (23), the high-SNR limit of $\mathbf{R}^{-1}(\theta)$ is given by

$$\mathbf{R}^{-1}(\theta) = \sigma_w^{-2}\mathbf{P}_\mathbf{A}^\perp(\theta) + \mathbf{B}(\theta) - \sigma_w^2\mathbf{B}^2(\theta) + o(\sigma_w^2)$$

where $\mathbf{P}_\mathbf{A}^\perp(\theta)$ and $\mathbf{B}(\theta)$ were defined in (13) and the identity $[\mathbf{A}^H(\theta)\mathbf{A}(\theta)]^{-1} = \mathbf{A}^\#(\theta)[\mathbf{A}^\#(\theta)]^H$ has been used.

From the asymptotic expression of $\mathbf{R}^{-1}(\theta)$, it follows that the asymptotic value of $\mathcal{R}^{-1}(\theta) = (\mathbf{R}^*(\theta))^{-1} \otimes \mathbf{R}^{-1}(\theta)$ has nine terms. However, three of these terms will be neglected during the high-SNR analysis and, thus, they are not made explicit in (12) for the sake of clarity.

APPENDIX II

HIGH-SNR LIMIT OF $\mathbf{Q}^{-1}(\theta)$

The kurtosis matrix \mathbf{K} is singular in case of complex circular constellations (6) whereas it is full-rank in case of noncircular CPM alphabets [1], [3]. In order to cover both formats, matrix \mathbf{K}

⁴The following relation has been considered to obtain the terms of the Taylor expansion:

$$\frac{\partial \mathbf{M}^{-1}(\lambda)}{\partial \lambda} = -\mathbf{M}^{-1}(\lambda) \frac{\partial \mathbf{M}(\lambda)}{\partial \lambda} \mathbf{M}^{-1}(\lambda).$$

is diagonalized as $\mathbf{V}_K \boldsymbol{\Sigma}_K \mathbf{V}_K^H$ where $\boldsymbol{\Sigma}_K$ is the diagonal matrix containing the nonzero eigenvalues of \mathbf{K} , and \mathbf{V}_K the associated eigenvectors. Then, the inversion lemma is applied, having that

$$\begin{aligned} \mathbf{Q}^{-1}(\boldsymbol{\theta}) &= (\mathcal{A}(\boldsymbol{\theta}) \mathbf{V}_K \boldsymbol{\Sigma}_K \mathbf{V}_K^H \mathcal{A}^H(\boldsymbol{\theta}) + \mathcal{R}(\boldsymbol{\theta}))^{-1} \\ &= \mathcal{R}^{-1}(\boldsymbol{\theta}) - \mathcal{R}^{-1}(\boldsymbol{\theta}) \mathcal{A}(\boldsymbol{\theta}) \\ &\quad \times \mathbf{V}_K \mathbf{M}^{-1} \mathbf{V}_K^H \mathcal{A}^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) \end{aligned} \quad (24)$$

with

$$\mathbf{M} \triangleq \boldsymbol{\Sigma}_K^{-1} + \mathbf{V}_K^H \mathcal{A}^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) \mathcal{A}(\boldsymbol{\theta}) \mathbf{V}_K.$$

Focusing on the inverse of \mathbf{M} , the asymptotic value of $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ is straightforward if $\boldsymbol{\Sigma}_K^{-1} + \mathbf{I}_{K^2}$ is invertible taking into account that the second term of \mathbf{M} is asymptotically constant (14). When this happens, it follows that

$$\begin{aligned} \mathbf{Q}^{-1}(\boldsymbol{\theta}) &= \mathcal{R}^{-1}(\boldsymbol{\theta}) \\ &\quad - [\mathcal{A}^\#(\boldsymbol{\theta})]^H \mathbf{V}_K \left(\boldsymbol{\Sigma}_K^{-1} + \mathbf{I}_{K^2} \right)^{-1} \mathbf{V}_K^H \mathcal{A}^\#(\boldsymbol{\theta}) + o(1) \end{aligned}$$

in which case the second term depending on \mathbf{K} is independent of σ_w^2 and hence negligible at high SNR compared to $\mathcal{R}^{-1}(\boldsymbol{\theta})$ (12). Notice that $\boldsymbol{\Sigma}_K^{-1} + \mathbf{I}_{K^2}$ is full rank if and only if all the eigenvalues of \mathbf{K} are different from -1 . This situation arises in case of multilevel modulations such as QAM and APSK.

However, when the nuisance parameters are drawn from a constant modulus alphabet (e.g., MPSK and CPM), some eigenvalues of \mathbf{K} are equal to -1 and, therefore, $\boldsymbol{\Sigma}_K^{-1} + \mathbf{I}_{K^2}$ is singular. In that case, the high-SNR limit of \mathbf{M}^{-1} is a little more involved because the terms of $\mathcal{R}^{-1}(\boldsymbol{\theta})$ in (12) depending on σ_w^2 must be considered as well, having that

$$\mathbf{M} = \boldsymbol{\Sigma}_K^{-1} + \mathbf{I}_{K^2} - \sigma_w^2 \mathbf{U}(\boldsymbol{\theta}) + o(\sigma_w^2)$$

where $\mathbf{U}(\boldsymbol{\theta})$ is the following full-rank matrix

$$\begin{aligned} \mathbf{U}(\boldsymbol{\theta}) &\triangleq \mathbf{V}_K^H \mathcal{A}^H(\boldsymbol{\theta}) \left[\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{B}^2(\boldsymbol{\theta}) + (\mathbf{B}^2(\boldsymbol{\theta}))^* \otimes \mathbf{B}(\boldsymbol{\theta}) \right] \\ &\quad \times \mathcal{A}(\boldsymbol{\theta}) \mathbf{V}_K \\ &= \mathbf{V}_K^H \left[\mathbf{I}_K \otimes (\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta}))^{-1} \right. \\ &\quad \left. + (\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta}))^{*-1} \otimes \mathbf{I}_K \right] \mathbf{V}_K. \end{aligned}$$

Thus, the inverse \mathbf{M}^{-1} can be solved computing the "economy-size" diagonalization of $\boldsymbol{\Sigma}_K^{-1} + \mathbf{I}_{K^2} = \mathbf{V} \boldsymbol{\Sigma} \mathbf{V}^H$, in which $\boldsymbol{\Sigma}$ is the diagonal matrix containing the eigenvalues of \mathbf{K} different from -1 and \mathbf{V} are the vectors of the canonical basis $\{\mathbf{e}_k\}$ selecting the position of these eigenvalues in $\boldsymbol{\Sigma}_K$. Formally, the k th diagonal entry of $\boldsymbol{\Sigma}_K$ different from -1 is selected by means of the vector \mathbf{e}_k defined as

$$[\mathbf{e}_k]_i \triangleq \begin{cases} 1 & i = k \\ 0 & i \neq k. \end{cases}$$

Then, the inversion lemma can be applied once more to \mathbf{M}^{-1} , obtaining

$$\begin{aligned} \mathbf{M}^{-1} &= (\mathbf{V} \boldsymbol{\Sigma} \mathbf{V}^H - \sigma_w^2 \mathbf{U}(\boldsymbol{\theta}))^{-1} + o(\sigma_w^{-2}) \\ &= -\sigma_w^{-2} \mathbf{P}_{\mathbf{V}}^\perp(\boldsymbol{\theta}) + o(\sigma_w^{-2}) \end{aligned}$$

where $\mathbf{P}_{\mathbf{V}}^\perp(\boldsymbol{\theta})$ is a projector onto the subspace orthogonal to the columns of \mathbf{V} defined as

$$\mathbf{P}_{\mathbf{V}}^\perp(\boldsymbol{\theta}) = \mathbf{U}^{-1}(\boldsymbol{\theta}) \left(\mathbf{I} - \mathbf{V} (\mathbf{V}^H \mathbf{U}^{-1}(\boldsymbol{\theta}) \mathbf{V})^{-1} \mathbf{V}^H \mathbf{U}^{-1}(\boldsymbol{\theta}) \right).$$

Notice that the conventional definition of the orthogonal projector $\mathbf{P}_{\mathbf{V}}^\perp = \mathbf{I} - \mathbf{V} \mathbf{V}^H$ is modified to include the weighting matrix $\mathbf{U}^{-1}(\boldsymbol{\theta})$. Anyway, $\mathbf{P}_{\mathbf{V}}^\perp(\boldsymbol{\theta})$ holds

$$\begin{aligned} \mathbf{P}_{\mathbf{V}}^\perp(\boldsymbol{\theta}) \mathbf{V} &= \mathbf{0} \\ \mathbf{V}^H \mathbf{P}_{\mathbf{V}}^\perp(\boldsymbol{\theta}) &= \mathbf{0} \end{aligned}$$

and, thus, $\mathbf{P}_{\mathbf{V}}^\perp(\boldsymbol{\theta})$ is a projection matrix onto the subspace generated by the eigenvectors of \mathbf{K} associated to the eigenvalue -1 .

Finally, putting together all the above partial results and using (14), we obtain the concluding result presented in (15), in which the name of the orthogonal projector $\mathbf{P}_{\mathbf{V}}^\perp(\boldsymbol{\theta})$ is changed by $\mathbf{P}_{\mathbf{K}}^\perp(\boldsymbol{\theta})$ in order to emphasize the dependence on the kurtosis matrix \mathbf{K} , i.e., $\mathbf{P}_{\mathbf{K}}^\perp(\boldsymbol{\theta}) \triangleq \mathbf{P}_{\mathbf{V}}^\perp(\boldsymbol{\theta})$.

REFERENCES

- [1] U. Mengali and A. D'Andrea, *Synchronization Techniques for Digital Receivers*. New York: Plenum, 1997.
- [2] G. D. Forney, "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 18, pp. 363–378, May 1972.
- [3] J. G. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [4] A. J. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *IEEE Trans. Commun.*, vol. 45, pp. 1218–1230, Oct. 1997.
- [5] G. C. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communications," *IEEE Trans. Commun.*, vol. 46, pp. 357–366, Mar. 1998.
- [6] T. S. Rappaport, J. H. Reed, and B. D. Woerner, "Position location using wireless communications on highways of the future," *IEEE Commun. Mag.*, vol. 34, pp. 33–41, Oct. 1996.
- [7] S. M. Kay, *Fundamentals of Statistical Signal Processing. Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [8] J. Villares and G. Vazquez, "Second-order parameter estimation," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2408–2420, Jul. 2005.
- [9] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Process.*, vol. 40, no. 3, pp. 590–600, Mar. 1992.
- [10] J.-P. Delmas, "Asymptotic performance of second-order algorithms," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 49–57, Jan. 2002.
- [11] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, pp. 459–471, Apr. 1983.
- [12] A. J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1136–1155, May 1996.
- [13] A. J. van der Veen, "Asymptotic properties of the algebraic constant modulus algorithm," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1796–1807, Aug. 2001.
- [14] A. Leshem and A. J. van der Veen, "Direction of arrival estimation for constant modulus signals," *IEEE Trans. Signal Process.*, vol. 47, no. 11, pp. 3125–3129, Nov. 1999.
- [15] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus With Applications in Statistics and Econometrics*. New York: Wiley, 1988.
- [16] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, no. 12, pp. 3473–3482, Dec. 1994.
- [17] 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.
- [18] T. Haykin, *Adaptive Filter Theory*. Englewood Cliffs, NJ: Prentice-Hall Int., 1991.
- [19] J. Villares, "Sample covariance based parameter estimation for digital communications" Ph.D. dissertation, Signal Theory and Communications Dept., Technical University of Catalonia (UPC), Barcelona, Spain, Sep. 2005 [Online]. Available: <http://gps-tsc.upc.es/~comm>

- [20] J. Villares and G. Vázquez, "Best quadratic unbiased estimator (BQUE) for timing and frequency synchronization," presented at the IEEE Workshop on Statistical Signal Processing (SSP) Singapore, Aug. 2001.
- [21] B. Porat and B. Friedlander, "On the limiting behavior of estimates based on sample covariances," presented at the IEEE Int. Conf. Acoustics Speech, Signal Processing (ICASSP) Tokyo, Japan, Apr. 1986.
- [22] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *Int. J. Adapt. Signal Process.*, vol. 3, pp. 191–229, 1989.
- [23] J. P. Delmas, "Asymptotically minimum variance second-order estimation for noncircular signals with application to DOA estimation," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1235–1241, May 2004.
- [24] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing," *Digit. Signal Process.*, vol. 3, pp. 185–210, Jul. 1998, Academic Press.
- [25] J. P. Delmas and H. Abeida, "Statistical resolution limits of DOA for discrete sources," presented at the IEEE Int. Conf. Acoustics Speech, Signal Processing (ICASSP), Toulouse, France, May 2006.
- [26] J. Villares and G. Vázquez, "Self-noise free second-order carrier phase synchronization of MSK-type signals," presented at the IEEE Int. Conf. Communications (ICC), Paris, France, Jun. 2004.
- [27] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Process. Lett.*, vol. 1, pp. 29–31, Feb. 1994.
- [28] J. A. Lopez-Salcedo and G. Vázquez, "Cyclostationary joint phase and timing estimation for staggered modulations," presented at the IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), Montreal, QC, Canada, May 2004.
- [29] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 720–741, May 1989.
- [30] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, pp. 1783–1795, Oct. 1990.



Javier Villares (S'01–M'05) was born in Barcelona, Catalonia, Spain, in 1974. He received the M.S. and Ph.D. degrees in telecommunications engineering from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1999 and 2005, respectively.

He has been a Professor in the Department of Signal Theory and Communications of the Technical University of Catalonia since 2003. His general interests are digital communications, statistical signal processing, and information theory.



Gregori Vázquez (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, Spain, in 1984 and 1988, respectively.

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

Dr. Vázquez is a recipient of the 2003 Best Paper Award of the IEEE Signal Processing Society. He served as a member of the Editorial Board of the *IEEE Signal Processing Magazine*, as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1999 to 2005, and also as a member of the Signal Processing for Communications Technical Committee. He was the Chairman of the Research Program on Communications of the Education and Science Ministry of Spain from 2004 to 2007.