Chapter 7

Asymptotic Studies

In this chapter, some asymptotic results are provided for the second-order estimators formulated in Chapters 3 and 4. In the first sections, their asymptotic performance is evaluated when the SNR is very low or very high. The low-SNR study concludes that the nuisance parameters distribution is irrelevant in a noisy scenario. In that case, the Gaussian assumption is shown to yield efficient estimators. On the other hand, the high-SNR asymptotic study is useful to bound the loses incurred when the Gaussian assumption is applied in spite of having non-Gaussian nuisance parameters. The most important conclusion is that the Gaussian assumption leads to optimal second-order schemes unless the nuisance parameters belong to a constant modulus constellation, such as MPSK or CPM. Therefore, the Gaussian assumption applies for very important constellations in digital communications such as QAM or APK.

The theoretical study is accompanied with some simulations for the problem of bearing estimation in case of digitally-modulated signals (Section 7.5.1). Numerical results are also provided in Section 7.5.2 for the problem of feedforward second-order frequency estimation initially addressed in Section 4.5. The same asymptotic study was carried out in Section 6.2 for the carrier phase estimation problem in case of noncircular transmissions.

In the second part of the chapter, the asymptotic performance of the second-order small-error estimators in Chapter 4 is evaluated when the data record grows to infinity. Asymptotic expressions are deduced for a vast majority of estimation problems in digital communications, such as timing and frequency synchronization, channel impulse response estimation and time-of-arrival estimation, among others. In that case, the large sample asymptotic expressions become a function of the spectra of the received waveform and its derivatives. In this context, a simple condition is obtained that identifies whether the Gaussian assumption yields optimal second-order schemes or not. From this result, the Gaussian assumption is proved to be optimal for timing and frequency synchronization. Some simulations are supplied in Section 7.5.2 that validate this last conclusion.

Asymptotic expressions are also obtained for the DOA estimation problem when the spatio-temporal observation grows indefinitely. If the number of antennas increases, it is shown that the covariance of the estimation error is asymptotically independent of the sources statistical distribution and, therefore, the Gaussian assumption can be applied to obtain efficient DOA estimators. On the other hand, the Gaussian assumption is found to yield an important loss if *the number of sensors is small and multiple constant-modulus signals* (e.g., MPSK or CPM) impinge into the array *from near directions.* These conclusions are validated numerically in Section 7.5.3.

7.1 Introduction

Let us summarize first the main results obtained in Chapters 3 and 4. As it was shown therein, any second-order estimator of $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$ is an affine transformation of the sample covariance matrix, having the following form:

$$\widehat{\boldsymbol{\alpha}} = \mathbf{g} + \mathbf{M}^H \left(\widehat{\mathbf{r}} - \mathbf{r}\right)$$

where $\mathbf{g} = E_\theta \left\{\mathbf{g}(\boldsymbol{\theta})\right\}$ and $\mathbf{r} = E_\theta E \left\{\widehat{\mathbf{r}}\right\}$ are the a priori knowledge about the parameter $\boldsymbol{\alpha}$ and the quadratic observation $\widehat{\mathbf{r}} = \mathrm{vec}(\mathbf{y}\mathbf{y}^H)$, respectively.

Based on the linear signal model presented in Section 2.4, matrix $\mathbf{M}$ was optimized in Chapters 3 and 4 by adopting different criteria. For the *large-error* MMSE and minimum variance second-order estimators studied in Chapter 3, matrix $\mathbf{M}$ was given by

$$\begin{aligned} \mathbf{M}_{mse} &\triangleq \left(\widetilde{\mathbf{Q}} + \mathbf{Q}\right)^{-1} \mathbf{S} \\ \mathbf{M}_{var} &\triangleq \mathbf{Q}^{-1}\widetilde{\mathbf{Q}} \left(\widetilde{\mathbf{Q}}\mathbf{Q}^{-1}\widetilde{\mathbf{Q}}\right)^{\#} \mathbf{S} \end{aligned} \tag{7.1}$$

where $\widetilde{\mathbf{Q}}$ was introduced in (3.23), and $\mathbf{Q}$ is the Bayesian expectation, $E_\theta \left\{\cdot\right\}$, of matrix

$$\mathbf{Q}\left(\boldsymbol{\theta}\right) = \mathcal{R}\left(\boldsymbol{\theta}\right) + \mathcal{A}\left(\boldsymbol{\theta}\right) \mathbf{K} \mathcal{A}^H\left(\boldsymbol{\theta}\right),$$

with $\mathbf{K}$ the fourth-order cumulant matrix in (3.11) and

$$\begin{aligned} \mathcal{A}\left(\boldsymbol{\theta}\right) &= \mathbf{A}^*\left(\boldsymbol{\theta}\right) \otimes \mathbf{A}\left(\boldsymbol{\theta}\right) \\ \mathcal{R}\left(\boldsymbol{\theta}\right) &= \mathbf{R}^*\left(\boldsymbol{\theta}\right) \otimes \mathbf{R}\left(\boldsymbol{\theta}\right) \\ \mathbf{R}\left(\boldsymbol{\theta}\right) &= \mathbf{A}\left(\boldsymbol{\theta}\right) \mathbf{A}^H\left(\boldsymbol{\theta}\right) + \mathbf{R}_w. \end{aligned} \tag{7.2}$$

On the other hand, the optimum second-order *small-error* estimator was obtained in Chapter 4, having that

$$\mathbf{M}_{bque}\left(\boldsymbol{\theta}\right) \triangleq \mathbf{Q}^{-1}\left(\boldsymbol{\theta}\right) \mathbf{D}_r\left(\boldsymbol{\theta}\right) \left(\mathbf{D}_r^H\left(\boldsymbol{\theta}\right) \mathbf{Q}^{-1}\left(\boldsymbol{\theta}\right) \mathbf{D}_r\left(\boldsymbol{\theta}\right)\right)^{-1} \mathbf{D}_g^H\left(\boldsymbol{\theta}\right) \tag{7.3}$$

where $\boldsymbol{\theta}$ stands henceforth for the actual value of the parameter and

$$[\mathbf{D}_r(\boldsymbol{\theta})]_p = \text{vec}\left(\frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_p}\right) = \text{vec}\left(\frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_p}\mathbf{A}^H(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta})\frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_p}\right) \tag{7.4}$$

is the derivative of $\mathbf{R}(\boldsymbol{\theta})$ with respect to the p-th parameter, i.e., $\theta_p \triangleq [\boldsymbol{\theta}]_p$.

The MSE matrices for the above estimators are given by[1]

$$\begin{aligned}
\Sigma_{mse} &\triangleq \Sigma_g - \mathbf{S}^H\left(\mathbf{Q}+\widetilde{\mathbf{Q}}\right)^{-1}\mathbf{S} \\
\Sigma_{var} &\triangleq \Sigma_g + \mathbf{S}^H\left(\widetilde{\mathbf{Q}}\mathbf{Q}^{-1}\widetilde{\mathbf{Q}}\right)^{\#}\mathbf{S} - \mathbf{S}^H\widetilde{\mathbf{Q}}^{\#}\mathbf{S} \\
\mathbf{B}_{bque}(\boldsymbol{\theta}) &\triangleq \mathbf{D}_g(\boldsymbol{\theta})\left(\mathbf{D}_r^H(\boldsymbol{\theta})\mathbf{Q}^{-1}(\boldsymbol{\theta})\mathbf{D}_r(\boldsymbol{\theta})\right)^{-1}\mathbf{D}_g^H(\boldsymbol{\theta}).
\end{aligned} \tag{7.5}$$

where $\mathbf{D}_g(\boldsymbol{\theta}) = \partial \mathbf{g}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^T$ and

$$\Sigma_g \triangleq E_{\boldsymbol{\theta}}\left\{(\mathbf{g}(\boldsymbol{\theta})-\mathbf{g})(\mathbf{g}(\boldsymbol{\theta})-\mathbf{g})^H\right\}$$

stands for the prior covariance matrix.

Finally, when the above estimators are deduced under the Gaussian assumption (i.e., $\mathbf{K}=\mathbf{0}$ and $\mathbf{Q}=\mathcal{R}$), their performance is given by the following MSE matrices:

$$\begin{aligned}
\Sigma'_{mse} &\triangleq \Sigma_g - \mathbf{S}^H\left(\widetilde{\mathbf{Q}}+\mathcal{R}\right)^{-1}\mathbf{S} + \mathbf{X}_{mse}(\mathbf{K}) \\
\Sigma'_{var} &\triangleq \Sigma_g + \mathbf{S}^H\left(\widetilde{\mathbf{Q}}\mathcal{R}^{-1}\widetilde{\mathbf{Q}}\right)^{\#}\mathbf{S} - \mathbf{S}^H\widetilde{\mathbf{Q}}^{\#}\mathbf{S} + \mathbf{X}_{var}(\mathbf{K}) \\
\mathbf{B}_{gml}(\boldsymbol{\theta}) &\triangleq \mathbf{B}_{UCRB}(\boldsymbol{\theta}) + \mathbf{X}_{gml}(\mathbf{K}).
\end{aligned} \tag{7.6}$$

where

$$\mathbf{B}_{UCRB}(\boldsymbol{\theta}) \triangleq \mathbf{D}_g(\boldsymbol{\theta})\left(\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathbf{D}_r(\boldsymbol{\theta})\right)^{-1}\mathbf{D}_g^H(\boldsymbol{\theta}) \tag{7.7}$$

is the well-known (Gaussian) unconditional CRB (Section 2.6.1) and, $\mathbf{X}_{mse}(\mathbf{K})$, $\mathbf{X}_{var}(\mathbf{K})$, $\mathbf{X}_{gml}(\mathbf{K})$ are the terms depending on the kurtosis matrix $\mathbf{K}$, which are given by

$$\begin{aligned}
\mathbf{X}_{mse}(\mathbf{K}) &\triangleq \mathbf{S}^H\left(\widetilde{\mathbf{Q}}+\mathcal{R}\right)^{-1}E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})\mathbf{K}\mathcal{A}^H(\boldsymbol{\theta})\right\}\left(\widetilde{\mathbf{Q}}+\mathcal{R}\right)^{-1}\mathbf{S} \\
\mathbf{X}_{var}(\mathbf{K}) &\triangleq \mathbf{S}^H\left(\widetilde{\mathbf{Q}}\mathcal{R}^{-1}\widetilde{\mathbf{Q}}\right)^{\#}\widetilde{\mathbf{Q}}\mathcal{R}^{-1}E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})\mathbf{K}\mathcal{A}^H(\boldsymbol{\theta})\right\}\mathcal{R}^{-1}\widetilde{\mathbf{Q}}\left(\widetilde{\mathbf{Q}}\mathcal{R}^{-1}\widetilde{\mathbf{Q}}\right)^{\#}\mathbf{S} \\
\mathbf{X}_{gml}(\mathbf{K}) &\triangleq \mathbf{D}_g(\boldsymbol{\theta})\left(\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathbf{D}_r(\boldsymbol{\theta})\right)^{-1}\left(\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta})\mathbf{K}\mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathbf{D}_r(\boldsymbol{\theta})\right) \\
&\quad \left(\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathbf{D}_r(\boldsymbol{\theta})\right)^{-1}\mathbf{D}_g^H(\boldsymbol{\theta})
\end{aligned} \tag{7.8}$$

It will be shown in next sections that $\mathbf{X}_{gml}(\mathbf{K})$ is always negligible for very low or high SNR. Nonetheless, the GML estimator might outperform the associated UCRB if the SNR is

[1]The "MSE matrix" is defined as $E_{\boldsymbol{\theta}}E\left\{\mathbf{e}\mathbf{e}^H\right\}$ where $\mathbf{e}$ stands for the considered estimation error [Kay93b].

moderate and $\mathbf{X}_{gml}(\mathbf{K})$ is negative. This behaviour has been observed, for example, in the DOA estimation problem in Section 6.5. On the other hand, in this chapter, the Gaussian assumption is proved to yield the optimal second-order estimator when the SNR goes to zero or if the amplitude of the nuisance parameters is not constant and the SNR goes to infinity. Finally, regarding the large-error MMSE and minimum variance estimators, $\mathbf{X}_{var}(\mathbf{K})$ and $\mathbf{X}_{mse}(\mathbf{K})$ are irrelevant at low SNR but they are determinant at high SNR because they are able to reduce the variance floor.

Before going into detail, let us decompose the noise covariance matrix $\mathbf{R}_w$ as $\sigma_w^2\mathbf{N}$ in order to make explicit the dependence on the noise variance σ_w^2. Assuming that the noise is stationary, the diagonal entries of $\mathbf{R}_w$ are precisely the noise variance σ_w^2. Formally, the noise variance is given by

$$\sigma_w^2 \triangleq \mathrm{Tr}(\mathbf{R}_w)/M.$$

and, therefore, $\mathbf{N} = \mathbf{R}_w/\sigma_w^2$ has unitary diagonal entries, by definition. Furthermore, in next sections, it will be useful to consider the following fourth-order matrix:

$$\mathcal{N} \triangleq \mathbf{N}^* \otimes \mathbf{N}.$$

7.2 Low SNR Study

When the noise variance goes to infinity ($\sigma_w^2 \to \infty$), the inverse of $\mathbf{R}(\boldsymbol{\theta})$, $\mathbf{Q}(\boldsymbol{\theta})$ and $\mathcal{R}(\boldsymbol{\theta})$ take the following asymptotic form:

$$\begin{aligned}\mathbf{R}^{-1}(\boldsymbol{\theta}) &= \sigma_w^{-2}\mathbf{N}^{-1} + o\left(\sigma_w^{-2}\right)\\ \mathbf{Q}^{-1}(\boldsymbol{\theta}), \mathcal{R}^{-1}(\boldsymbol{\theta}) &= \sigma_w^{-4}\mathcal{N}^{-1} + o\left(\sigma_w^{-4}\right)\end{aligned}$$

assuming that $\mathbf{N}$ is full-rank. The Landau symbol $o(x)$ is introduced to consider all those terms that converge to zero faster than x. On the other hand, the rest of matrices appearing in (7.5) and (7.6) are independent of σ_w^2. Specifically, the noise variance does not affect the value of $\mathbf{A}(\boldsymbol{\theta})$, $\mathcal{A}(\boldsymbol{\theta})$, $\widetilde{\mathbf{Q}}$, $\mathbf{S}$, $\mathbf{D}_r(\boldsymbol{\theta})$, $\mathbf{D}_g(\boldsymbol{\theta})$, $\mathbf{K}$ and Σ_g.

Therefore, the MSE matrices in (7.5) and (7.6) have the following asymptotic expressions at low SNR:

$$\begin{aligned}\Sigma_{mse}, \Sigma'_{mse} &= \Sigma_g - \sigma_w^{-4}\mathbf{S}^H\mathcal{N}^{-1}\mathbf{S} + o\left(\sigma_w^{-4}\right)\\ \Sigma_{var}, \Sigma'_{var} &= \sigma_w^4\mathbf{S}^H\left(\widetilde{\mathbf{Q}}\mathcal{N}^{-1}\widetilde{\mathbf{Q}}\right)^{\#}\mathbf{S} + o\left(\sigma_w^4\right)\end{aligned} \tag{7.9}$$

$$\mathbf{B}_{UCRB}(\boldsymbol{\theta}), \mathbf{B}_{gml}(\boldsymbol{\theta}), \mathbf{B}_{bque}(\boldsymbol{\theta}) = \sigma_w^4\mathbf{D}_g(\boldsymbol{\theta})\left(\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{N}^{-1}(\boldsymbol{\theta})\mathbf{D}_r(\boldsymbol{\theta})\right)^{-1}\mathbf{D}_g^H(\boldsymbol{\theta}) + o\left(\sigma_w^4\right).$$

taking into account that $\mathbf{X}_{mse}(\mathbf{K})$ in (7.8) is proportional to σ_w^{-8} and, $\mathbf{X}_{var}(\mathbf{K})$ and $\mathbf{X}_{gml}(\mathbf{K})$ are constant. Notice that the fourth-order matrix $\mathbf{K}$ does not appear in none of the above

asymptotic expressions. This implies that the actual distribution of the nuisance parameters becomes irrelevant at low SNR when designing second-order schemes. Moreover, *any assumption* about the distribution of the nuisance parameters yields the same MSE expressions in (7.9).

To complete the analysis, the asymptotic expression of the studied second-order estimators is provided next:

$$\begin{aligned}
\mathbf{M}_{mse}, \mathbf{M}'_{mse} &= \sigma_w^{-4} \mathcal{N}^{-1} \mathbf{S} + o\left(\sigma_w^{-4}\right) \\
\mathbf{M}_{var}, \mathbf{M}'_{var} &= \mathcal{N}^{-1} \widetilde{\mathbf{Q}} \left(\widetilde{\mathbf{Q}} \mathcal{N}^{-1} \widetilde{\mathbf{Q}} \right)^{\#} \mathbf{S} + o(1) \\
\mathbf{M}_{bque}(\boldsymbol{\theta}), \mathbf{M}_{gml}(\boldsymbol{\theta}) &= \mathcal{N}^{-1}(\boldsymbol{\theta}) \mathbf{D}_r(\boldsymbol{\theta}) \left(\mathbf{D}_r^H(\boldsymbol{\theta}) \mathcal{N}^{-1}(\boldsymbol{\theta}) \mathbf{D}_r(\boldsymbol{\theta}) \right)^{-1} \mathbf{D}_g^H(\boldsymbol{\theta}) + o(1)
\end{aligned} \tag{7.10}$$

where $\mathbf{M}'_{mse}$ and $\mathbf{M}'_{var}$ correspond to the minimum variance and MMSE estimators obtained under the Gaussian assumption.

In Appendix 7.A, it is shown that $\mathbf{M}_{bque}(\boldsymbol{\theta})$ and $\mathbf{M}_{gml}(\boldsymbol{\theta})$ in (7.10) coincide with the scoring method that implements the low-SNR ML estimator deduced in Section 2.4.1. Due to the asymptotic efficiency of the ML estimator (Section 2.3.2), if the GML and BQUE estimators converge to the ML solution at low SNR, we can state that the GML and BQUE estimators become asymptotically efficiency as the SNR goes to zero. As it was discussed in Section 2.3.2, the "asymptotic" condition is satisfied whenever the estimator operates in the *small-error* regime or, equivalently, the actual SNR exceeds the SNR threshold. Accordingly, in the studied low SNR scenario ($\sigma_w^2 \to \infty$), the asymptotic condition requires that the observation length goes to infinity ($M \to \infty$) in order to attain the small-error regime.

Likewise, because the GML is efficient at low SNR, the associated (Gaussian) UCRB (7.9) becomes the *true* CRB at low SNR if and only if the observation size goes to infinity (small-error). Notice that both the UCRB and the true CRB are proportional to σ_w^{-4} at low SNR, as it was reported in [Ste01] for the problem of timing synchronization.

7.3 High SNR Study

In low SNR conditions, the Gaussian assumption has been proved to yield optimal second-order estimators. However, when the SNR increases, the optimal second-order estimators listed in (7.1) and (7.3) exploit the fourth-order statistical information about the nuisance parameters contained in matrix $\mathbf{K}$. When the Gaussian assumption is adopted and this information is omitted ($\mathbf{K} = \mathbf{0}$), the performance of the studied second-order estimators degrades at high SNR. In this section, this loss is upper bounded by evaluating the asymptotic performance of the aforementioned estimation methods when the noise variance goes to zero.

In Appendix 7.B, the asymptotic value of $\mathbf{R}^{-1}(\boldsymbol{\theta})$ and $\mathcal{R}^{-1}(\boldsymbol{\theta})$ as the noise variance goes to

zero is calculated, obtaining

$$\mathbf{R}^{-1}(\boldsymbol{\theta}) = \sigma_w^{-2}\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) + \mathbf{B}(\boldsymbol{\theta}) - \sigma_w^2\mathbf{B}(\boldsymbol{\theta})\mathbf{N}\mathbf{B}(\boldsymbol{\theta}) + O\left(\sigma_w^4\right) \tag{7.11}$$

$$\begin{aligned}\mathcal{R}^{-1}(\boldsymbol{\theta}) = {} & \sigma_w^{-4}\left[\mathbf{P}_{\mathbf{A}}^{\perp *}(\boldsymbol{\theta}) \otimes \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\right] \\ & + \sigma_w^{-2}\left[\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) + \mathbf{P}_{\mathbf{A}}^{\perp *}(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta})\right] \\ & + \mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta}) \\ & - \sigma_w^2\left(\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta})\mathbf{N}\mathbf{B}(\boldsymbol{\theta}) + \left[\mathbf{B}(\boldsymbol{\theta})\mathbf{N}\mathbf{B}(\boldsymbol{\theta})\right]^* \otimes \mathbf{B}(\boldsymbol{\theta})\right) \\ & + O\left(\sigma_w^4\right)\end{aligned} \tag{7.12}$$

where the Landau symbol $O(x)$ includes all the terms that converge to zero as x or faster. The asymptotic value of $\mathbf{R}^{-1}(\boldsymbol{\theta})$ and $\mathcal{R}^{-1}(\boldsymbol{\theta})$ is given in terms of the following matrices:

$$\mathbf{A}^{\#}(\boldsymbol{\theta}) \triangleq \left(\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\right)^{-1}\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1} \tag{7.13}$$

$$\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \triangleq \mathbf{N}^{-1}\left[\mathbf{I}_M - \mathbf{A}(\boldsymbol{\theta})\mathbf{A}^{\#}(\boldsymbol{\theta})\right] \tag{7.14}$$

$$\mathbf{B}(\boldsymbol{\theta}) \triangleq \left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H\mathbf{A}^{\#}(\boldsymbol{\theta}) \tag{7.15}$$

where $\mathbf{A}^{\#}(\boldsymbol{\theta})$ and $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})$ are variations of the Moore-Penrose pseudoinverse and the projector onto the null subspace of $\mathbf{A}(\boldsymbol{\theta})$, respectively. The original definitions are altered to include the whitening matrix $\mathbf{N}^{-1}$ in case of correlated noise samples (i.e., $\mathbf{N} \neq \mathbf{I}_M$). Although abusing of notation, the above matrices retain all the properties of the original definition, that is,

$$\begin{gathered}\mathbf{A}^{\#}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}_K \\ \mathbf{A}(\boldsymbol{\theta})\mathbf{A}^{\#}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta}) \\ \mathbf{A}^{\#}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta})^{\#}\mathbf{A}(\boldsymbol{\theta}) = \mathbf{A}^{\#}(\boldsymbol{\theta}) \\ \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) = \mathbf{0} \\ \mathbf{A}^H(\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) = \mathbf{0}.\end{gathered}$$

On the other hand, the asymptotic value of $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ depends on the kurtosis matrix $\mathbf{K}$. The complete study is carried out in Appendix 7.C when $\mathbf{K}$ is full-rank and in Appendix 7.D when $\mathbf{K}$ is singular. In these appendices, $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ is proved to have the following asymptotic expression:

$$\mathbf{Q}^{-1}(\boldsymbol{\theta}) = \mathcal{R}^{-1}(\boldsymbol{\theta}) + \sigma_w^{-2}\left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})\mathcal{A}^{\#}(\boldsymbol{\theta}) + O(1) \tag{7.16}$$

where the pseudoinverse of $\mathcal{A}(\boldsymbol{\theta}) = \mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})$ (7.2) is defined as follows

$$\mathcal{A}^{\#}(\boldsymbol{\theta}) \triangleq \left(\mathcal{A}^H(\boldsymbol{\theta})\mathcal{N}^{-1}\mathcal{A}(\boldsymbol{\theta})\right)^{-1}\mathcal{A}^H(\boldsymbol{\theta})\mathcal{N}^{-1} = \left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^* \otimes \mathbf{A}^{\#}(\boldsymbol{\theta}).$$

and $\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})$ stands for the projector onto the subspace generated by the eigenvectors of $\mathbf{K} = \mathbf{V}_K\Sigma_K\mathbf{V}_K^H$ associated to the eigenvalue -1.

The second term in (7.16) is positive semidefinite and it becomes zero if and only if $\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta}) = \mathbf{0}$, i.e., if all the eigenvalues of $\mathbf{K}$ are different from -1. The rank of $\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})$ is thus determinant to assess the potential benefit of considering the kurtosis matrix $\mathbf{K}$ in the design of second-order estimators. The exact expression of $\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})$ is given in Appendix 7.C ($\mathbf{K}$ full-rank) and in Appendix 7.D ($\mathbf{K}$ singular). In Section 7.3.3, the study of $\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})$ will be addressed with more detail.

It is worth realizing that all the above asymptotic results are implicitly assuming that the transfer matrix $\mathbf{A}(\boldsymbol{\theta})$ is full column rank. This will be the baseline for the asymptotic studies in this section. In addition, some indications are given in Appendix 7.E to carry out the asymptotic study when the rank of $\mathbf{A}(\boldsymbol{\theta})$ is lower than the number of columns K.

7.3.1 (Gaussian) Unconditional Cramér-Rao Bound

The (Gaussian) UCRB is widely used to lower bound the performance of second-order estimators. Thus far, it is proved that the UCRB is a valid second-order lower bound when the SNR goes to zero or if the nuisance parameters are actually Gaussian. Nonetheless, this is not generally true. Indeed, the UCRB is shown to be outperformed *at high SNR* by the optimal second-order small-error estimator proposed in Chapter 4. Likewise, the GML estimator usually outperforms the UCRB for intermediate SNRs.

In this section, the high-SNR limit of $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$ when the noise variance goes to zero is derived. It is shown that $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$ becomes proportional to σ_w^2 at high SNR and, therefore, self-noise free estimates are feasible when the nuisance parameters are Gaussian. Formally, we have that

$$\begin{aligned}\mathbf{B}_{UCRB}(\boldsymbol{\theta}) &= \mathbf{D}_g(\boldsymbol{\theta})\left(\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathbf{D}_r(\boldsymbol{\theta})\right)^{-1}\mathbf{D}_g^H(\boldsymbol{\theta}) \\ &= \sigma_w^2\mathbf{D}_g(\boldsymbol{\theta})\mathcal{B}_1^{-1}(\boldsymbol{\theta})\mathbf{D}_g^H(\boldsymbol{\theta}) + O\left(\sigma_w^4\right)\end{aligned} \tag{7.17}$$

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

$$\left[\mathcal{B}_1(\boldsymbol{\theta})\right]_{p,q} = 2\,\mathrm{Re}\,\mathrm{Tr}\left(\frac{\partial\mathbf{A}^H(\boldsymbol{\theta})}{\partial\theta_p}\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_q}\right). \tag{7.18}$$

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

$$\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_p} \neq \mathbf{0}$$

for all the parameter $\theta_1, \ldots, \theta_P$. This abnormal situation takes place if the noise subspace of matrix $\mathbf{A}(\boldsymbol{\theta})$ is null (Appendix 7.D) but also in the problem of carrier phase synchronization

addressed in Section 6.2. In both cases, the constant term $\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta})$ in (7.12) has to be considered in order to evaluate the *variance floor* at high SNR, having that

$$\lim_{\sigma_w^2 \to 0} \mathbf{B}_{UCRB}(\boldsymbol{\theta}) = \mathbf{D}_g(\boldsymbol{\theta}) \mathcal{B}_2^{-1}(\boldsymbol{\theta}) \mathbf{D}_g^H(\boldsymbol{\theta}) \tag{7.19}$$

where $\mathcal{B}_2(\boldsymbol{\theta})$ stands for the high-SNR limit of $\mathbf{D}_r^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) \mathbf{D}_r(\boldsymbol{\theta})$. The entries of $\mathcal{B}_2(\boldsymbol{\theta})$ are determined in Appendix 7.G, obtaining

$$[\mathcal{B}_2(\boldsymbol{\theta})]_{p,q} \triangleq 2 \operatorname{Re} \operatorname{Tr} \left(\frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{A}^{\#}(\boldsymbol{\theta}) \frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_q} \mathbf{A}^{\#}(\boldsymbol{\theta}) + \frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_p} \frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_q} \mathbf{B}(\boldsymbol{\theta}) \right). \tag{7.20}$$

7.3.2 Gaussian Maximum Likelihood

In most estimation problems, the UCRB takes the form in equation (7.17) 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 computed in (7.17) irrespective of the actual distribution of the nuisance parameters, i.e., even if $\mathbf{K} \neq \mathbf{0}$. Formally, we have that

$$\mathbf{B}_{gml}(\boldsymbol{\theta}), \mathbf{B}_{UCRB}(\boldsymbol{\theta}) = \sigma_w^2 \mathbf{D}_g(\boldsymbol{\theta}) \mathcal{B}_1^{-1}(\boldsymbol{\theta}) \mathbf{D}_g^H(\boldsymbol{\theta}) + O\left(\sigma_w^4\right) \tag{7.21}$$

with $\mathcal{B}_1(\boldsymbol{\theta})$ given in (7.18).

This statement is true because the term $\mathbf{X}_{gml}(\mathbf{K})$ in (7.6) can be neglected since it depends on σ_w^4 whereas $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$ is proportional to σ_w^2. Notice that $\mathbf{X}_{gml}(\mathbf{K})$ is proportional to σ_w^4 because

$$\mathcal{R}^{-1}(\boldsymbol{\theta}) \mathcal{A}(\boldsymbol{\theta}) = [\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta})] \mathcal{A}(\boldsymbol{\theta}) + O\left(\sigma_w^2\right)$$

is asymptotically constant, as pointed out in Appendix 7.B.

Finally, if $\partial \mathbf{A}(\boldsymbol{\theta}) / \partial \theta_p$ and $\mathbf{A}(\boldsymbol{\theta})$ were linearly dependent, the GML performance would exhibit a variance floor at high SNR that would be a function of the kurtosis matrix $\mathbf{K}$. Using equation (7.6), it follows that the GML variance floor would be equal to

$$\begin{aligned} \lim_{\sigma_w^2 \to 0} \mathbf{B}_{gml}(\boldsymbol{\theta}) = & \mathbf{D}_g(\boldsymbol{\theta}) \mathcal{B}_2^{-1}(\boldsymbol{\theta}) \mathbf{D}_g^H(\boldsymbol{\theta}) \\ & + \mathbf{D}_g(\boldsymbol{\theta}) \mathcal{B}_2^{-1}(\boldsymbol{\theta}) \left(\mathbf{D}_r^H(\boldsymbol{\theta}) \left[\mathcal{A}^{\#}(\boldsymbol{\theta}) \right]^H \mathbf{K} \mathcal{A}^{\#}(\boldsymbol{\theta}) \mathbf{D}_r(\boldsymbol{\theta}) \right) \mathcal{B}_2^{-1}(\boldsymbol{\theta}) \mathbf{D}_g^H(\boldsymbol{\theta}) \end{aligned} \tag{7.22}$$

where $\mathbf{B}_{UCRB}(\boldsymbol{\theta}) = \mathbf{D}_g(\boldsymbol{\theta}) \mathcal{B}_2^{-1}(\boldsymbol{\theta}) \mathbf{D}_g^H(\boldsymbol{\theta})$ is the variance floor in case of Gaussian nuisance parameters (7.19) and the second term corresponds to $\mathbf{X}_{gml}(\boldsymbol{\theta})$ in (7.6).

7.3.3 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 in Appendix 7.C and Appendix 7.D 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. It seems that this condition is related to the constant modulus of the nuisance parameters. This important result suggests to classify the nuisance parameters distribution according to the eigendecomposition of $\mathbf{K}$. With this purpose, let us first obtain the asymptotic expression of $\mathbf{B}_{bque}(\boldsymbol{\theta})$ (7.5) as the noise variance goes to zero, i.e., $\sigma_w^2 \to 0$.

Using the asymptotic value of $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ in (7.16), it follows that

$$\begin{aligned}\mathbf{D}_r^H(\boldsymbol{\theta})\,\mathbf{Q}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta}) = \mathbf{D}_r^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta}) \\ +\sigma_w^{-2}\mathbf{D}_r^H(\boldsymbol{\theta})\left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H \mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})\,\mathcal{A}^{\#}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta}) + O(1) \quad (7.23)\end{aligned}$$

where $\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta}) \in \mathbb{R}^{K^2\times K^2}$ denotes the projector onto the subspace generated by the eigenvectors of $\mathbf{K}$ associated to the eigenvalue -1.

Using now the asymptotic expression of $\mathbf{D}_r^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})$ in (7.17), it follows that

$$\begin{aligned}\mathbf{B}_{bque}(\boldsymbol{\theta}) = \sigma_w^2\mathbf{D}_g(\boldsymbol{\theta}) \quad (7.24)\\ \left(\mathcal{B}_1(\boldsymbol{\theta}) + \mathbf{D}_r^H(\boldsymbol{\theta})\left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H \mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})\,\mathcal{A}^{\#}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})\right)^{-1}\mathbf{D}_g^H(\boldsymbol{\theta}) + O(\sigma_w^4)\end{aligned}$$

where the second term inside the inverse is always positive semidefinite and, therefore, we can state at high SNR that

$$\mathbf{B}_{bque}(\boldsymbol{\theta}) \leq \mathbf{B}_{gml}(\boldsymbol{\theta}) = \mathbf{B}_{UCRB}(\boldsymbol{\theta}).$$

The second term of (7.24) 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 $\partial\mathbf{A}(\boldsymbol{\theta})/\partial\theta_p$ lies *totally* in the noise subspace of $\mathbf{A}(\boldsymbol{\theta})$, i.e,

$$\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_p} = \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_p} \quad (7.25)$$

or, taking into account the definition of $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})$ in (7.14),

$$\mathbf{A}^H(\boldsymbol{\theta})\,\mathbf{N}^{-1}\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_p} = \mathbf{0}.$$

In that case, after some simple manipulations, it can be shown that

$$\begin{aligned}\left[\mathcal{A}^{\#}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})\right]_p &= \mathrm{vec}\left(\mathbf{A}^{\#}(\boldsymbol{\theta})\frac{\partial\mathbf{R}(\boldsymbol{\theta})}{\partial\theta_p}\left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H\right)\\ &= \mathrm{vec}\left(\frac{\partial\mathbf{A}^H(\boldsymbol{\theta})}{\partial\theta_p}\left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H + \mathbf{A}^{\#}(\boldsymbol{\theta})\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_p}\right) = \mathbf{0}\end{aligned}$$

and, thus, the second term in (7.24) is strictly zero independently of the nuisance parameters distribution. For example, this condition applies in digital synchronization as the observation length goes to infinity (Section 7.4.4).

By comparing this condition and the one introduced in Section 7.3.1, we can conclude that the condition (7.25) never applies if the UCRB and GML suffer from self-noise at high SNR since, in that case, $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\,\partial\mathbf{A}(\boldsymbol{\theta})/\partial\theta_p = \mathbf{0}$.

2. **Nuisance parameters distribution.** Regardless of the signal parameterization, the Gaussian assumption applies at high SNR if all the eigenvalues of the kurtosis matrix $\mathbf{K}$ are different from -1. In that case, $\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})$ is strictly zero, and the second term in (7.24) becomes zero.

   If the nuisance parameters are drawn from an arbitrary *circular* complex alphabet, the kurtosis matrix is given by $\mathbf{K} = (\rho - 2)\,\mathrm{diag}(\mathrm{vec}(\mathbf{I}_K))$ (3.12) 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 constant modulus alphabets. Accordingly, in the context of digital communications, the Gaussian assumption applies for any multilevel linear modulation such as QAM or APK. On the other hand, it does not apply in case of any complex MPSK modutation holding that $\rho = 1$.

   If the nuisance parameters are not circular, there is not a closed-form expression for the eigenvalues of $\mathbf{K}$. However, it is found that the kurtosis matrix of some important *constant-modulus* noncircular modulations has some eigenvalues equal to -1. Among them, an special attention is given in this thesis to the CPM modulation. Other important constant-modulus noncircular modulations are the BPSK and those constant-modulus staggered modulations such as the offset QPSK [Pro95].

   Finally, in those scenarios in which the UCRB (7.19) and the GML (7.22) exhibit a variance floor at high SNR because $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\,\partial\mathbf{A}(\boldsymbol{\theta})/\partial\theta_p = \mathbf{0}$, the Gaussian assumption fails when the nuisance parameters have constant modulus. In that case, the second term in (7.24) allows cancelling the self-noise because

$$\mathbf{D}_r^H(\boldsymbol{\theta})\,\mathbf{Q}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta}) = \mathcal{B}_2(\boldsymbol{\theta}) + \sigma_w^{-2}\mathbf{D}_r^H(\boldsymbol{\theta})\left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H \mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})\,\mathcal{A}^{\#}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta}) + O(1)$$

   and, therefore, the constant term $\mathcal{B}_2(\boldsymbol{\theta})$ (7.20) can be neglected when compared to the second term, that is proportional to σ_w^{-2}. Using this result, the asymptotic variance of the optimal second-order estimator is given by

$$\mathbf{B}_{bque}(\boldsymbol{\theta}) = \sigma_w^2\mathbf{D}_g(\boldsymbol{\theta})\left(\mathbf{D}_r^H(\boldsymbol{\theta})\left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H \mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})\,\mathcal{A}^{\#}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})\right)^{-1}\mathbf{D}_g^H(\boldsymbol{\theta}) + O\left(\sigma_w^4\right)$$

   This situation arises in the carrier phase estimation problem studied in Section 6.2 as well as in the scenarios simulated in Section 4.5 in which $\mathbf{A}(\boldsymbol{\theta})$ is not full-column rank.

7.3.4 Large Error Estimators

In this section, the asymptotic performance of the Bayesian estimators in (7.1) is analyzed when the SNR goes to infinity. The result of this asymptotic study depends on the influence of the Bayesian expectation $E_{\boldsymbol{\theta}}\{\cdot\}$ on the following matrices:

$$\begin{aligned}
\mathbf{R} &\triangleq E_{\boldsymbol{\theta}}\{\mathbf{R}(\boldsymbol{\theta})\} = \mathbf{G} + \sigma_w^2\mathbf{N} \\
\mathcal{R} &\triangleq E_{\boldsymbol{\theta}}\{\mathcal{R}(\boldsymbol{\theta})\} = E_{\boldsymbol{\theta}}\{\mathbf{R}^*(\boldsymbol{\theta}) \otimes \mathbf{R}(\boldsymbol{\theta})\} \\
&= E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})\mathcal{A}^H(\boldsymbol{\theta})\right\} + \sigma_w^2\mathbf{U} + \sigma_w^4\mathcal{N} \\
\mathbf{Q} &\triangleq E_{\boldsymbol{\theta}}\{\mathbf{Q}(\boldsymbol{\theta})\} = \mathcal{R} + E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})\mathbf{K}\mathcal{A}^H(\boldsymbol{\theta})\right\} \\
&= E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})(\mathbf{I}_{K^2} + \mathbf{K})\mathcal{A}^H(\boldsymbol{\theta})\right\} + \sigma_w^2\mathbf{U} + \sigma_w^4\mathcal{N}
\end{aligned} \tag{7.26}$$

with the following definitions:

$$\mathbf{G} \triangleq E_{\boldsymbol{\theta}}\left\{\mathbf{A}(\boldsymbol{\theta})\mathbf{A}^H(\boldsymbol{\theta})\right\} \tag{7.27}$$

$$\mathbf{U} \triangleq [\mathbf{G}^* \otimes \mathbf{N} + \mathbf{N}^* \otimes \mathbf{G}] \tag{7.28}$$

The Bayesian expectation always increases the rank of these matrices. Even if the prior distribution is rather informative, these matrices become rapidly full rank. Therefore, let us consider that $\mathbf{G}$, and hence $\mathbf{U}$, are eventually full rank. In that case, the MSE matrices in (7.5) and (7.6) converge to the following limits at high SNR (Appendix 7.H):

$$\begin{aligned}
\lim_{\sigma_w^2\to 0}\Sigma_{mse} &= \Sigma_g - \mathbf{S}^H\mathbf{B}_{T_1}\mathbf{S} \\
\lim_{\sigma_w^2\to 0}\Sigma_{var} &= \Sigma_g + \mathbf{S}^H\left(\widetilde{\mathbf{Q}}\mathbf{B}_{T_2}\widetilde{\mathbf{Q}}\right)^{\#}\mathbf{S} - \mathbf{S}^H\widetilde{\mathbf{Q}}^{\#}\mathbf{S} \\
\lim_{\sigma_w^2\to 0}\Sigma'_{mse} &= \Sigma_g - \mathbf{S}^H\mathbf{B}_{T_3}\mathbf{S} + \mathbf{X}_{mse}(\mathbf{K}) \\
\lim_{\sigma_w^2\to 0}\Sigma'_{var} &= \Sigma_g + \mathbf{S}^H\left(\widetilde{\mathbf{Q}}\mathbf{B}_{T_4}\widetilde{\mathbf{Q}}\right)^{\#}\mathbf{S} - \mathbf{S}^H\widetilde{\mathbf{Q}}^{\#}\mathbf{S} + \mathbf{X}_{var}(\mathbf{K})
\end{aligned} \tag{7.29}$$

where

$$\begin{aligned}
\mathbf{X}_{mse}(\mathbf{K}) &= \mathbf{S}^H\mathbf{B}_{T_3}E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})\mathbf{K}\mathcal{A}^H(\boldsymbol{\theta})\right\}\mathbf{B}_{T_3}\mathbf{S} \\
\mathbf{X}_{var}(\mathbf{K}) &= \mathbf{S}^H\left(\widetilde{\mathbf{Q}}\mathbf{B}_{T_4}\widetilde{\mathbf{Q}}\right)^{\#}\widetilde{\mathbf{Q}}\mathbf{B}_{T_4}E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})\mathbf{K}\mathcal{A}^H(\boldsymbol{\theta})\right\}\mathbf{B}_{T_4}\widetilde{\mathbf{Q}}\left(\widetilde{\mathbf{Q}}\mathbf{B}_{T_4}\widetilde{\mathbf{Q}}\right)^{\#}\mathbf{S}
\end{aligned} \tag{7.30}$$

and $\mathbf{B}_T$ is computed as

$$\mathbf{B}_T \triangleq \mathbf{U}^{-1}\mathbf{V}_T\left(\mathbf{V}_T^H\mathbf{U}^{-1}\mathbf{V}_T\right)^{-1}\Sigma_T^{-1}\left(\mathbf{V}_T^H\mathbf{U}^{-1}\mathbf{V}_T\right)^{-1}\mathbf{V}_T^H\mathbf{U}^{-1} \tag{7.31}$$

with $\mathbf{V}_T\Sigma_T\mathbf{V}_T^H$ the "economy-size" diagonalization of the specific matrix $\mathbf{T}$ considered in (7.29):

$$\begin{aligned}
\mathbf{T}_1 &\triangleq E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})(\mathbf{I}_{K^2} + \mathbf{K})\mathcal{A}^H(\boldsymbol{\theta})\right\} + \widetilde{\mathbf{Q}} \\
\mathbf{T}_2 &\triangleq E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})(\mathbf{I}_{K^2} + \mathbf{K})\mathcal{A}^H(\boldsymbol{\theta})\right\} \\
\mathbf{T}_3 &\triangleq E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})\mathcal{A}^H(\boldsymbol{\theta})\right\} + \widetilde{\mathbf{Q}} \\
\mathbf{T}_4 &\triangleq E_{\boldsymbol{\theta}}\left\{\mathcal{A}(\boldsymbol{\theta})\mathcal{A}^H(\boldsymbol{\theta})\right\}.
\end{aligned} \tag{7.32}$$

Taking a glance at (7.29), one observes that the terms

$$\mathbf{S}^H \left(\widetilde{\mathbf{Q}}\mathbf{B}_{T_2}\widetilde{\mathbf{Q}}\right)^{\#} \mathbf{S}$$
$$\mathbf{S}^H \left(\widetilde{\mathbf{Q}}\mathbf{B}_{T_4}\widetilde{\mathbf{Q}}\right)^{\#} \mathbf{S} + \mathbf{X}_{var}\left(\mathbf{K}\right)$$

in Σ'_{var} and Σ_{var} correspond to the *self-noise* whereas the term $\Sigma_g - \mathbf{S}^H\widetilde{\mathbf{Q}}^{\#}\mathbf{S}$ is the estimator *bias* at high SNR. The self-noise terms were found to vanish when the observation time was infinite for the problem of blind frequency estimation in Section 3.4. On the other hand, the bias term could not be cancelled extending the observation time in the problem of blind frequency estimation (Section 3.4).

7.4 Large Sample Study

In this section, the asymptotic performance of the second-order estimators deduced in Chapter 4 is evaluated when the number of observed samples goes to infinity ($M \rightarrow \infty$). Notice that M can be increased by augmenting either the sampling rate (N_{ss}) or the observation interval ($N_s = M/N_{ss}$).

In the first case, the sampling theorem states that it is enough to take $N_{ss} = 2$ samples per symbol for those modulations with an excess of band smaller than 100%. However, when the observation window is too short, the observed spectrum becomes wider due to well-known smearing and leakage effects [Sto97]. The proposed estimators deal with this problem by applying the best temporal window according to the known signal model and the adopted optimization criterion. Nonetheless, if the vector of nuisance parameters is longer than the number of observed samples, it is not possible to avoid the variance floor at high SNR unless N_{ss} is increased (see Appendix 7.E). This problem is only relevant when the observation time is really short, as it has been considered in this dissertation so far. If the observation time N_s is augmented, the problem of spectral aliasing becomes rapidly negligible and $N_{ss} = 2$ becomes sufficient.

The importance of the sampling rate was also evidenced in Section 3.4 for the problem of carrier frequency-offset synchronization. It was shown therein that the estimator bias can *only* be cancelled if N_{ss} goes to infinity. Surprisingly, the bias term cannot be removed by only increasing N_s. However, this sort of arguments are specific to the frequency estimation problem and should be revised for other estimation problems.

Considering in the sequel that N_{ss} is fixed, asymptotic expressions are given in this section for the *small-error* second-order estimators deduced in Chapter 4 as the observation length goes to infinity ($M \rightarrow \infty$) . The study for the large error estimators in Chapter 3 is omitted because it is less insightful due to the role of the Bayesian expectation (see Section 7.3.4).

In the large sample case, a unified analysis is not feasible because the results depend on the actual parameterization and on how $\mathbf{A}(\boldsymbol{\theta})$ grows when $M \to \infty$. In this section, the problems of non-data-aided synchronization in Section 6.1, blind time-of-arrival estimation in Section 6.3, blind channel identification in Section 6.4 and DOA estimation in Section 6.5 are considered.

Before addressing the asymptotic study for the aforementioned estimation problems, the covariance matrices $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$, $\mathbf{B}_{gml}(\boldsymbol{\theta})$ and $\mathbf{B}_{bque}(\boldsymbol{\theta})$ in Section 7.1 are now restated in terms of the following matrices

$$\begin{aligned} \mathbf{B}(\boldsymbol{\theta}) &\triangleq \mathbf{A}^H(\boldsymbol{\theta})\,\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta}) \\ \mathbf{B}_p(\boldsymbol{\theta}) &\triangleq \frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_p}\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta}) \\ \mathbf{B}_{p,q}(\boldsymbol{\theta}) &\triangleq \frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_p}\mathbf{N}^{-1}\frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_q}. \end{aligned} \tag{7.33}$$

for $p, q = 1, \ldots, P$. These matrices collect all the scalar products between two columns of $\mathbf{A}(\boldsymbol{\theta})$ and $\partial\mathbf{A}(\boldsymbol{\theta})/\partial\theta_p$ –normalized by means of $\mathbf{N}^{-1}$. It is shown in the following subsections that these $K \times K$ matrices determine enterely the performance of second-order estimators. Thus, it is only necessary to study the asymptotic value of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ as the number of observations goes to infinity ($M \to \infty$) or, in other words, as the dimension of the column space of $\mathbf{A}(\boldsymbol{\theta})$ and $\partial\mathbf{A}(\boldsymbol{\theta})/\partial\theta_p$ increases wihtout limit.

For the sake of clarity, we will consider hereafter that $\mathbf{g}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ and, in most cases, the noise term will be assumed white, i.e., $\mathbf{N} = \mathbf{I}_M$.

7.4.1 (Gaussian) Unconditional Cramér-Rao Bound

After some simplifications, the (Gaussian) UCRB in (7.7) can be restated as

$$\begin{aligned} \left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q} &= \left[\mathbf{D}_r^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})\right]_{p,q} \\ &= 2\,\mathrm{Re}\,\mathrm{Tr}\left\{\left(\frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_p}\mathbf{R}^{-1}(\boldsymbol{\theta})\,\mathbf{A}(\boldsymbol{\theta})\right)\left(\frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_q}\mathbf{R}^{-1}(\boldsymbol{\theta})\,\mathbf{A}(\boldsymbol{\theta})\right)\right. \\ &\quad \left. + \left(\mathbf{A}^H(\boldsymbol{\theta})\,\mathbf{R}^{-1}(\boldsymbol{\theta})\,\mathbf{A}(\boldsymbol{\theta})\right)\left(\frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_p}\mathbf{R}^{-1}(\boldsymbol{\theta})\frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_q}\right)\right\}. \end{aligned} \tag{7.34}$$

using the algebraic properties in (7.54) from Appendix 7.A.

Then, if the inversion lemma is applied to arrange the inverse of $\mathbf{R}(\boldsymbol{\theta})$ as

$$\mathbf{R}^{-1}(\boldsymbol{\theta}) = \sigma_w^{-2}\mathbf{N}^{-1} - \sigma_w^{-2}\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\left(\mathbf{A}^H(\boldsymbol{\theta})\,\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta}) + \mathbf{I}_K\right)^{-1}\mathbf{A}^H(\boldsymbol{\theta})\,\mathbf{N}^{-1},$$

it follows that the entries of $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ become a function of the following three matrices:

$$\begin{aligned}
\mathbf{X}(\boldsymbol{\theta}) &\triangleq \sigma_w^2 \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{R}^{-1}(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta}) = \mathbf{B}(\boldsymbol{\theta}) - \mathbf{B}(\boldsymbol{\theta}) \left(\mathbf{B}(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}_K\right)^{-1} \mathbf{B}(\boldsymbol{\theta}) \\
\mathbf{X}_p(\boldsymbol{\theta}) &\triangleq \sigma_w^2 \frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{R}^{-1}(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta}) = \mathbf{B}_p(\boldsymbol{\theta}) - \mathbf{B}_p(\boldsymbol{\theta}) \left(\mathbf{B}(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}_K\right)^{-1} \mathbf{B}(\boldsymbol{\theta}) \\
\mathbf{X}_{p,q}(\boldsymbol{\theta}) &\triangleq \sigma_w^2 \frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{R}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_q} = \mathbf{B}_{p,q}(\boldsymbol{\theta}) - \mathbf{B}_p(\boldsymbol{\theta}) \left(\mathbf{B}(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}_K\right)^{-1} \mathbf{B}_q^H(\boldsymbol{\theta})
\end{aligned} \tag{7.35}$$

where $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ were introduced in (7.33).

Therefore, plugging (7.35) into (7.34), it follows that

$$\left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q} = 2\sigma_w^{-4} \operatorname{Re} \operatorname{Tr} \left\{ \mathbf{X}_p(\boldsymbol{\theta}) \mathbf{X}_q(\boldsymbol{\theta}) + \mathbf{X}(\boldsymbol{\theta}) \mathbf{X}_{p,q}(\boldsymbol{\theta}) \right\}. \tag{7.36}$$

7.4.2 Gaussian Maximum Likelihood

In this section, the GML covariance matrix $\mathbf{B}_{gml}(\boldsymbol{\theta})$ is restated in terms of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$. Bearing in mind that $\mathbf{D}_g(\boldsymbol{\theta}) = \mathbf{I}_P$, it follows that (7.6) can be written as

$$\mathbf{B}_{gml}(\boldsymbol{\theta}) = \mathbf{B}_{UCRB}(\boldsymbol{\theta}) + \mathbf{X}_{gml}(\mathbf{K}) = \mathbf{B}_{UCRB}(\boldsymbol{\theta}) + \mathbf{B}_{UCRB}(\boldsymbol{\theta}) \Psi(\mathbf{K}) \mathbf{B}_{UCRB}(\boldsymbol{\theta}). \tag{7.37}$$

where

$$\Psi(\mathbf{K}) \triangleq \mathbf{D}_r^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) \mathcal{A}(\boldsymbol{\theta}) \mathbf{K} \mathcal{A}^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) \mathbf{D}_r(\boldsymbol{\theta}).$$

Next, we will prove that $\Psi(\mathbf{K})$ is also a function of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ (7.33). Taking into account the definitions of $\mathcal{A}(\boldsymbol{\theta})$ and $\mathcal{R}(\boldsymbol{\theta})$ in (7.2), the associative property of the Kronecker product yields

$$\mathcal{A}^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) = \left[\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{R}^{-1}(\boldsymbol{\theta})\right]^* \otimes \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{R}^{-1}(\boldsymbol{\theta}).$$

Then, using again the matrix properties in (7.54), it can be seen that

$$\left[\mathcal{A}^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) \mathbf{D}_r(\boldsymbol{\theta})\right]_p = \operatorname{vec}\left(\mathbf{X}(\boldsymbol{\theta}) \mathbf{X}_p(\boldsymbol{\theta}) + \mathbf{X}_p^H(\boldsymbol{\theta}) \mathbf{X}(\boldsymbol{\theta})\right)$$

where $\mathbf{X}(\boldsymbol{\theta})$ and $\mathbf{X}_p(\boldsymbol{\theta})$ were introduced in (7.35).

Therefore, $\Psi(\mathbf{K})$ can be written as

$$\begin{aligned}
[\Psi(\mathbf{K})]_{p,q} &= \sigma_w^{-4} \operatorname{vec}^H(\mathbf{Y}_p(\boldsymbol{\theta})) \mathbf{K} \operatorname{vec}(\mathbf{Y}_q(\boldsymbol{\theta})) \\
&= \sigma_w^{-4} \operatorname{vec}^H(\mathbf{Y}_p(\boldsymbol{\theta})) \mathbf{V}_K \Sigma_K \mathbf{V}_K^H \operatorname{vec}(\mathbf{Y}_q(\boldsymbol{\theta}))
\end{aligned} \tag{7.38}$$

where

$$\mathbf{Y}_p(\boldsymbol{\theta}) \triangleq \mathbf{X}(\boldsymbol{\theta}) \mathbf{X}_p(\boldsymbol{\theta}) + \mathbf{X}_p^H(\boldsymbol{\theta}) \mathbf{X}(\boldsymbol{\theta}) \tag{7.39}$$

and $\mathbf{V}_K \Sigma_K \mathbf{V}_K^H$ is the "economy-size" diagonalization of $\mathbf{K}$.

To conclude this analysis, $\Psi(\mathbf{K})$ can be further simplified when the nuisance parameters are *circular*. In that case, the kurtosis matrix $\mathbf{K}$ is equal to $(\rho - 2)\,\mathrm{diag}\,(\mathrm{vec}\,(\mathbf{I}_K))$ (3.12), and the eigenvalues and eigenvectors of $\mathbf{K}$ are given by

$$\begin{aligned} \Sigma_K &= (\rho - 2)\,\mathbf{I}_K \\ [\mathbf{V}_K]_k &= \mathrm{vec}\left(\mathbf{e}_k \mathbf{e}_k^H\right) \end{aligned} \tag{7.40}$$

where $\mathbf{e}_k \in \mathbb{R}^K$ is defined as

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

In [Mag98, Ex.4, p.62], it is shown that $\mathbf{V}_K$ has the following interesting properties:

$$\begin{aligned} \mathbf{V}_K^H\,\mathrm{vec}(\mathbf{A}) &= \mathrm{diag}\,(\mathbf{A}) \\ [(\mathbf{A} \otimes \mathbf{B})\,\mathbf{V}_K]_k &= [\mathbf{A}]_k \otimes [\mathbf{B}]_k \\ \mathbf{V}_K^H\,(\mathbf{A} \otimes \mathbf{B})\,\mathbf{V}_K &= \mathbf{A} \odot \mathbf{B} \end{aligned} \tag{7.41}$$

for any pair of matrices $\mathbf{A}$ and $\mathbf{B}$ of appropriate size.

Taking into account the first property in (7.41), (7.42) becomes

$$\begin{aligned} [\Psi(\mathbf{K})]_{p,q} &= \sigma_w^{-4}\,\mathrm{diag}^H\,(\mathbf{Y}_p(\boldsymbol{\theta}))\,\Sigma_K\,\mathrm{diag}\,(\mathbf{Y}_q(\boldsymbol{\theta})) \\ &= \sigma_w^{-4}\,(\rho - 2)\,\mathrm{Tr}\,(\mathbf{Y}_p(\boldsymbol{\theta}) \odot \mathbf{Y}_q(\boldsymbol{\theta})) \end{aligned} \tag{7.42}$$

using (7.40) and the following identity:

$$\mathrm{diag}^H\,(\mathbf{A})\,\mathrm{diag}\,(\mathbf{B}) = \mathrm{Tr}\,(\mathbf{A}^* \odot \mathbf{B}) = \mathrm{Tr}\left(\mathbf{A}^H \odot \mathbf{B}\right).$$

Regarding now the definition of $\mathbf{Y}_p(\boldsymbol{\theta})$ (7.39), it follows that $\mathrm{diag}\,(\mathbf{Y}_q(\boldsymbol{\theta}))$ is always real-valued because

$$\mathrm{diag}\,(\mathbf{Y}_q(\boldsymbol{\theta})) = 2\,\mathrm{Re}\left\{\mathrm{diag}\left(\mathbf{X}_p^H(\boldsymbol{\theta})\,\mathbf{X}(\boldsymbol{\theta})\right)\right\}.$$

7.4.3 Best Quadratic Unbiased Estimator

In this section, the same analysis is carried out for the optimal second-order estimator. The aim is also to formulate $\mathbf{B}_{bque}(\boldsymbol{\theta})$ in terms of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ (7.33). To begin with, the inversion lemma is applied to $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ obtaining

$$\mathbf{B}_{bque}^{-1}(\boldsymbol{\theta}) = \mathbf{D}_r^H(\boldsymbol{\theta})\,\mathbf{Q}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta}) = \mathbf{D}_r^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta}) + \Gamma(\mathbf{K}) \tag{7.43}$$

where the first term corresponds to $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ in (7.36), and

$$\begin{aligned}\Gamma(\mathbf{K}) \triangleq & -\mathbf{D}_r^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathcal{A}(\boldsymbol{\theta})\,\mathbf{V}_K\left(\mathbf{V}_K^H\mathcal{A}^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathcal{A}(\boldsymbol{\theta})\,\mathbf{V}_K+\Sigma_K^{-1}\right)^{-1}\\ & \mathbf{V}_K^H\mathcal{A}^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})\end{aligned} \tag{7.44}$$

is the term depending on the kurtosis matrix[2]. The "economy-size" diagonalization of $\mathbf{K}$ is introduced to encompass those problems in which $\mathbf{K}$ is singular (e.g., CPM).

Next, $\Gamma(\mathbf{K})$ is formulated in terms of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ (7.33),

$$\begin{aligned}\left[\Gamma(\mathbf{K})\right]_{p,q} = & -\sigma_w^{-4}\,\mathrm{vec}^H\left(\mathbf{Y}_p(\boldsymbol{\theta})\right)\\ & \mathbf{V}_K\left(\mathbf{V}_K^H\left[\mathbf{X}^*(\boldsymbol{\theta})\otimes\mathbf{X}(\boldsymbol{\theta})\right]\mathbf{V}_K+\sigma_w^4\Sigma_K^{-1}\right)^{-1}\mathbf{V}_K^H\,\mathrm{vec}\left(\mathbf{Y}_p(\boldsymbol{\theta})\right),\end{aligned} \tag{7.45}$$

using the following identities:

$$\begin{aligned}\left[\mathcal{A}^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})\right]_p &= \sigma_w^{-4}\,\mathrm{vec}\left(\mathbf{Y}_p(\boldsymbol{\theta})\right)\\ \mathcal{A}^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathcal{A}(\boldsymbol{\theta}) &= \sigma_w^{-4}\left[\mathbf{X}^*(\boldsymbol{\theta})\otimes\mathbf{X}(\boldsymbol{\theta})\right].\end{aligned}$$

Unfortunately, the analysis of (7.45) is really involved except for those circular alphabets holding $\mathbf{K}=(\rho-2)\,\mathrm{diag}\left(\mathrm{vec}\left(\mathbf{I}_K\right)\right)$ (3.12). In that case, from (7.40) and (7.41), we have

$$\begin{aligned}\mathbf{V}_K^H\,\mathrm{vec}\left(\mathbf{Y}_p(\boldsymbol{\theta})\right) &= \mathrm{diag}\left(\mathbf{Y}_p(\boldsymbol{\theta})\right)\\ \mathbf{V}_K^H\left[\mathbf{X}^*(\boldsymbol{\theta})\otimes\mathbf{X}(\boldsymbol{\theta})\right]\mathbf{V}_K &= \mathbf{X}^*(\boldsymbol{\theta})\odot\mathbf{X}(\boldsymbol{\theta}).\end{aligned}$$

Accordingly, if $\rho\neq 2$, the non-Gaussian term $\Gamma(\mathbf{K})$ is given by

$$\left[\Gamma(\mathbf{K})\right]_{p,q} = -\sigma_w^{-4}\,\mathrm{diag}^H\left(\mathbf{Y}_p(\boldsymbol{\theta})\right)\left(\mathbf{X}^*(\boldsymbol{\theta})\odot\mathbf{X}(\boldsymbol{\theta})+\sigma_w^4(\rho-2)^{-1}\mathbf{I}_K\right)^{-1}\mathrm{diag}\left(\mathbf{Y}_q(\boldsymbol{\theta})\right). \tag{7.46}$$

Regarding the last expression, the following important conclusion arises. If the nuisance parameters are circular (3.12), the Gaussian assumption applies *independently of the SNR and the nuisance parameters distribution* if

$$\mathrm{diag}\left(\mathbf{Y}_p(\boldsymbol{\theta})\right) = 2\,\mathrm{Re}\,\mathrm{diag}\left(\mathbf{X}_p^H(\boldsymbol{\theta})\,\mathbf{X}(\boldsymbol{\theta})\right) = \mathbf{0} \tag{7.47}$$

for $p=1,\ldots,P$ where $\mathbf{X}(\boldsymbol{\theta})$ and $\mathbf{X}_p(\boldsymbol{\theta})$ were defined in (7.35). If the last equation holds true, $\Psi(\mathbf{K})$ and $\Gamma(\mathbf{K})$ are exactly zero in view of (7.42) and (7.46). This condition will be tested in the following sections to validate the Gaussian assumption in some relevant estimation problems in digital communications.

Notice that the last condition is more restrictive than the one presented in (7.25). Actually, it is straightforward to realize that (7.47) is satisfied if (7.25) is held because, in that case, $\mathbf{B}_p^H(\boldsymbol{\theta}) = \mathbf{A}^H(\boldsymbol{\theta})\,\mathbf{N}^{-1}\partial\mathbf{A}(\boldsymbol{\theta})/\partial\theta_p = \mathbf{0}$ and hence $\mathbf{X}_p(\boldsymbol{\theta})=\mathbf{0}$ (7.35).

[2]Notice that $\Gamma(\mathbf{K})$ is actually the second term in (7.23).

7.4.4 Second-Order Estimation in Digital Communications

In this section, simple asymptotic closed-form expressions are obtained for any estimation problem in which multiple replicas of the same waveform (pulse) are periodically received. The received waveform is parameterized by a finite set of parameters $\boldsymbol{\theta}$ and will be referred to as $g(t;\boldsymbol{\theta})$ in this section. Assuming that a continuous stream of symbols is received, the structure of $\mathbf{A}(\boldsymbol{\theta})$ corresponds to the one represented in Fig. 6.1 (right-hand side) in Section 6.1.2.

This framework encompasses most estimation problems in digital communications, among others the synchronization problems described in Section 6.1, the problem of blind channel identification in Section 6.4 and, the problem of time-of-arrival estimation studied in Section 6.3. Although the problem of frequency estimation does not fall into this category because the phase of the received waveform is time-varying, it is proved in Appendix 7.J that quadratic NDA techniques are only aware of the carrier phase variation within the received pulse duration.

In this section, the aymptotic value of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ (7.33) is determined for N_s going to infinity. In that case, the size of these $K \times K$ square matrices also increases proportionally as $N_s \to \infty$ because

$$K = N_s + L - 1$$

with L the pulse duration in symbol periods. However, although the size of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ tends to infinity, the central rows and columns of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ contain delayed versions of the following autocorrelation and cross-correlation functions[3]

$$\begin{aligned} R[k] &\triangleq \int_{-\infty}^{\infty} g(t)g^*(t+kT)dt \\ R_p[k] &\triangleq \int_{-\infty}^{\infty} g(t)g_p^*(t+kT)dt \\ R_{p,q}[k] &\triangleq \int_{-\infty}^{\infty} g_p(t)g_q^*(t+kT)dt, \end{aligned}$$

where $g_p(t;\boldsymbol{\theta}) \triangleq \partial g(t;\boldsymbol{\theta})/\partial\theta_p$ stands for derivative of $g(t;\boldsymbol{\theta})$ with respect to the p-th parameter θ_p. In the sequel, the dependence on $\boldsymbol{\theta}$ will be omitted for the sake of brevity.

Henceforth, only the central rows and columns of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ will be considered bearing in mind that the "edge effect" is negligible in the asymptotic case ($N_s \to \infty$) or in case of TDMA signals (Section 6.1.2). This analysis is inspired in the asymptotic study carried out in [Rib01b] for the CML timing estimator[4]. In [Rib01b], it is shown that the multiplication of these matrices yields another matrix whose central columns and rows are the *convolution* ($*$)

[3]For simplicity the noise is assumed uncorrelated, i.e., $\mathbf{N} = \mathbf{I}_M$. Otherwise, the same expressions are valid for the *whitened* waveform $\eta(t) * g(t;\boldsymbol{\theta})$ where $\eta(mT_s)$ is the central column of $\mathbf{N}^{-1/2}$.

[4]Likewise, the same reasoning was adopted in [Kay93b, Sec. 7.9] to get asymptotic expressions for the Newton-Raphson and scoring recursions in the context of maximum likelihood estimation.

of the central columns and rows of the original matrices. This allows computing $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$, $\mathbf{B}_{gml}(\boldsymbol{\theta})$ and $\mathbf{B}_{bque}(\boldsymbol{\theta})$ as follows:

$$\begin{aligned}\left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q} &= 2\sigma_w^{-4}\,\mathrm{Re}\,\mathrm{Tr}\left(\mathbf{X}_p\left(\boldsymbol{\theta}\right)\mathbf{X}_q\left(\boldsymbol{\theta}\right)+\mathbf{X}\left(\boldsymbol{\theta}\right)\mathbf{X}_{p,q}\left(\boldsymbol{\theta}\right)\right)\\ &= 2N_s\sigma_w^{-4}\,\mathrm{Re}\left(X_p\left[k\right]*X_q\left[k\right]+X\left[k\right]*X_{p,q}\left[k\right]\right)\big|_{k=0}+o\left(N_s\right)\end{aligned}$$

where the central rows and columns of $\mathbf{X}\left(\boldsymbol{\theta}\right)$, $\mathbf{X}_p\left(\boldsymbol{\theta}\right)$ and $\mathbf{X}_{p,q}\left(\boldsymbol{\theta}\right)$ are given by

$$\begin{aligned}X[k] &\triangleq R[k]-R[k]*\left(R[k]+\sigma_w^2\delta[k]\right)^{-1}*R[k]\\ X_p[k] &\triangleq R_p[k]-R_p[k]*\left(R[k]+\sigma_w^2\delta[k]\right)^{-1}*R[k]\\ X_{p,q}[k] &\triangleq R_{p,q}[k]-R_p[k]*\left(R[k]+\delta[k]\right)^{-1}*R_q^*[-k].\end{aligned}\tag{7.48}$$

In the above equations, the inverse operator $(\cdot)^{-1}$ stands for the deconvolution, i.e., $a^{-1}[k]$ is the sequence holding $a[k]*a^{-1}[k]=\delta[k]$. As it is the usual practice, this deconvolution is solved in the frequency domain. Using standard Fourier calculus, it is found that

$$\left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q} = 2N_s\sigma_w^{-4}\,\mathrm{Re}\int_{-0.5}^{0.5} S_{X_p}(f)S_{X_q}(f)+S_X(f)S_{X_{p,q}}(f)df+o\left(N_s\right)$$

where the Fourier transforms of $X[k]$, $X_p[k]$ and $X_{p,q}[k]$ are given next in terms of the Fourier transforms of $R[k]$, $R_p[k]$ and $R_{p,q}[k]$:

$$\begin{aligned}S_X(f) &\triangleq \mathfrak{F}\left\{X[k]\right\}=S(f)-\frac{S^2(f)}{S(f)+\sigma_w^2}=\frac{\sigma_w^2S(f)}{S(f)+\sigma_w^2}\\ S_{X_p}(f) &\triangleq \mathfrak{F}\left\{X_p[k]\right\}=S_p\left(f\right)-\frac{S_p\left(f\right)S\left(f\right)}{S(f)+\sigma_w^2}=\frac{\sigma_w^2S_p\left(f\right)}{S(f)+\sigma_w^2}\\ S_{X_{p,q}}(f) &\triangleq \mathfrak{F}\left\{X_{p,q}[k]\right\}=S_{p,q}(f)-\frac{S_p(f)S_q^*\left(f\right)}{S(f)+\sigma_w^2}.\end{aligned}\tag{7.49}$$

Focusing uniquely on circular complex alphabets (e.g., MPSK and QAM), the term $\Psi\left(\mathbf{K}\right)$ appearing in the GML covariance matrix (7.37) is asymptotically ($N_s\rightarrow\infty$) given by

$$\left[\Psi\left(\mathbf{K}\right)\right]_{p,q}=\sigma_w^{-4}\left(\rho-2\right)\mathrm{Tr}\left(\mathbf{Y}_p\left(\boldsymbol{\theta}\right)\odot\mathbf{Y}_q\left(\boldsymbol{\theta}\right)\right)=N_s\sigma_w^{-4}\left(\rho-2\right)Y_p^2[0]+o\left(N_s\right)$$

where

$$Y_p[k]\triangleq X[k]*X_p[k]+X_p^*[-k]*X[k]=2\,\mathrm{Re}\left\{X_p^*[-k]*X[k]\right\}.$$

On the other hand, the term $\Gamma(\mathbf{K})$ (7.46) appearing in the BQUE covariance matrix (7.43)

is asympotically equal to

$$\begin{aligned}
[\Gamma(\mathbf{K})]_{p,q} &= -\sigma_w^{-4}\,\mathrm{diag}^H(\mathbf{Y}_p(\boldsymbol{\theta}))\left(\mathbf{X}^*(\boldsymbol{\theta})\odot\mathbf{X}(\boldsymbol{\theta})+\sigma_w^4(\rho-2)^{-1}\mathbf{I}_K\right)^{-1}\mathrm{diag}(\mathbf{Y}_q(\boldsymbol{\theta})) \\
&= -N_s\sigma_w^{-4}Y_p^2[0]\sum_{k=-\infty}^{\infty}\left(|X[k]|^2+\sigma_w^4(\rho-2)^{-1}\delta[k]\right)^{-1}+o(N_s) \\
&= -\sigma_w^{-4}N_s\left.\frac{Y_p^2[0]}{S(f)*S^*(-f)+\sigma_w^4/(\rho-2)}\right|_{f=0}+o(N_s) \\
&= -\sigma_w^{-4}N_s\frac{\int_{-0.5}^{0.5}S_{Y_p}^2(f)df}{\int_{-0.5}^{0.5}|S(f)|^2df+\sigma_w^4/(\rho-2)}+o(N_s)
\end{aligned}$$

with $S_{Y_p}(f)=\mathfrak{F}\{Y_p[k]\}$ the Fourier transform of $Y_p[k]$.

Regarding now the Gaussian condition in (7.47), the matrices $\Gamma(\mathbf{K})$ and $\Psi(\mathbf{K})$ are asymptotically null if

$$\begin{aligned}
Y_p[0] &= 2\,\mathrm{Re}\left\{X_p^*[-k]*X[k]\right\}\big|_{k=0}=2\,\mathrm{Re}\left\{\sum_n X_p^*[n]X[n]\right\} \\
&= 2\,\mathrm{Re}\left\{\int_{-0.5}^{0.5}S_{X_p}^*(f)S(f)df\right\}=2\sigma_w^4\,\mathrm{Re}\left\{\int_{-0.5}^{0.5}\frac{S_p^*(f)S(f)}{|S(f)+\sigma_w^2|^2}df\right\}
\end{aligned}$$

is equal to zero independently of the actual value of ρ and σ_w^2. The last expression has been formulated in the frequency domain using the Parceval's identity and the Fourier transforms of $X[k]$ and $X_p[k]$ in (7.49).

Notice that the energy spectrum $S(f)=\mathfrak{F}\{R[k]\}$ is always real because $R[k]$ has Hermitian symmetry, i.e., $R[k]=R^*[-k]$. Besides, $S(f)$ is even if $R[k]$ is real-valued, which implies that $g(t;\boldsymbol{\theta})$ is also real-valued. Therefore, there are three possible situations leading to $Y_p[0]=0$, and hence validating the Gaussian assumption:

1. $S_p(f)$ is *imaginary*, i.e., $\mathrm{Re}\{S_p(f)\}=0$. From the Fourier theory, $S_p(f)=\mathfrak{F}\{R_p[k]\}$ is imaginary if $R_p[k]$ is imaginary or, $R_p[k]$ is real but it has odd symmetry.

2. $S(f)$ is an *even* function whereas $S_p(f)$ is an *odd* function. The former condition is held for $g(t;\boldsymbol{\theta})$ real-valued. The latter condition is held if and only if the cross-correlation $R_p[k]$ is also odd.

3. $S(f)$ is an *odd* function whereas $S_p(f)$ is an *even* function. The former condition is held when the received waveform $g(t;\boldsymbol{\theta})$ is imaginary and even. The latter condition is held if the cross-correlation $R_p[k]$ is also even.

It is found that the last conditions usually apply in frequency and timing[5] synchronization. In frequency synchronization, the cross-correlation $R_p[k]$ is imaginary (condition 1). On the

[5]The same conclusion applies to the related problem of time-of-arrival estimation in radiolocation applications.

other hand, in timing synchronization, the cross-correlation $R_p[k]$ is a real-valued odd function because the transmitted pulse is usually a real symmetric function in digital communications (condition 2)[6]. Nonetheless, the Gaussian assumption generally fails in the problem of blind channel identification (Section 6.4) because the received waveform $g(t;\boldsymbol{\theta})$ is distorted by the *complex* channel impulse response and, hence, the complex cross-correlation $R_p[k]$ does not exhibit any symmetry. For example, if the channel is multiplicative, the received waveform is a scaled version of the transmitted pulse and it is easy to show that $R_p[k]$ is proportional to $R[k]$.

7.4.5 Second-Order Estimation in Array Signal Processing

In array signal processing, the spatio-temporal observation can be written as

$$\mathbf{y} = \text{vec}\left(\mathbf{A}_s(\boldsymbol{\theta})(\mathbf{A}_t\mathbf{X})^T\right) + \mathbf{w} \tag{7.50}$$

where

$$[\mathbf{A}_s(\boldsymbol{\theta})]_p = \exp\left(j\pi\theta_p\widetilde{\mathbf{d}}_M\right)$$

is the spatial signature of the p-th user impinging on a $\lambda/2$-spaced linear array from the direction $\theta_p \in [-1,1)$ with

$$\begin{aligned}\widetilde{\mathbf{d}}_M &\triangleq \mathbf{d}_M - \frac{M-1}{2}\\ \mathbf{d}_M &= [0,\ldots M-1]^T.\end{aligned}$$

In (7.50), the modulation matrix $\mathbf{A}_t \in \mathbb{R}^{N_s\times K}$ contains the shaping pulse $p(t)$, $[\mathbf{X}]_k$ are the received symbols from to the k-th user and, $\mathbf{w}$ the spatio-temporal Gaussian noise vector. Notice that the array is calibrated to have unitary response when the signal comes from the broadside ($\theta_p = 0$). However, the same results would be obtained adopting any other calibration.

The observation vector $\mathbf{y}$ can be arranged in the standard form,

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

using that $\text{vec}\left(\mathbf{ABC}^T\right) = (\mathbf{A}\otimes\mathbf{C})\,\text{vec}(\mathbf{B})$. Then, we have

$$\begin{aligned}\mathbf{A}(\boldsymbol{\theta}) &= \mathbf{A}_t \otimes \mathbf{A}_s(\boldsymbol{\theta})\\ \mathbf{x} &= \text{vec}\left(\mathbf{X}^T\right).\end{aligned}$$

Based on the general expressions deduced in Section 7.4, the asymptotic value of $\mathbf{B}(\boldsymbol{\theta})$, $\mathbf{B}_p(\boldsymbol{\theta})$ and $\mathbf{B}_{p,q}(\boldsymbol{\theta})$ in (7.33) is now obtained for the above spatio-temporal signal model. It is

[6]This implies that the same pulse shaping is used in the in-phase and quadrature components.

straightforward to show that

$$\begin{aligned}\mathbf{B}(\boldsymbol{\theta}) &= \mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t \otimes \mathbf{A}_s^H(\boldsymbol{\theta}) \mathbf{N}_s^{-1} \mathbf{A}_s(\boldsymbol{\theta}) \\ \mathbf{B}_p(\boldsymbol{\theta}) &= \mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t \otimes \frac{\partial \mathbf{A}_s^H(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{N}_s^{-1} \mathbf{A}_s(\boldsymbol{\theta}) \\ \mathbf{B}_{p,q}(\boldsymbol{\theta}) &= \mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t \otimes \frac{\partial \mathbf{A}_s^H(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{N}_s^{-1} \frac{\partial \mathbf{A}_s(\boldsymbol{\theta})}{\partial \theta_q}\end{aligned}$$

assuming that the temporal and spatial components of the noise are decoupled as $\mathbf{N} = \mathbf{N}_t \otimes \mathbf{N}_s$ and using that

$$\left[\frac{\partial \mathbf{A}_s(\boldsymbol{\theta})}{\partial \theta_p}\right]_i = \begin{cases} j\pi \widetilde{\mathbf{d}}_M \odot \exp\left(j\pi\theta_p \widetilde{\mathbf{d}}_M\right) & i = p \\ \mathbf{0} & i \neq p \end{cases}.$$

Moreover, assuming for simplicity that the noise is spatially uncorrelated (i.e., $\mathbf{N}_s = \mathbf{I}_M$), we have that the *spatial* cross-correlation for the P users is determined by the following matrices:

$$\begin{aligned}[\mathcal{B}(\boldsymbol{\theta})]_{i,k} &\triangleq \left[\mathbf{A}_s^H(\boldsymbol{\theta}) \mathbf{N}_s^{-1} \mathbf{A}_s(\boldsymbol{\theta})\right]_{i,k} = F_M(\theta_i - \theta_k) \\ [\mathcal{B}_p(\boldsymbol{\theta})]_{i,k} &\triangleq \left[\frac{\partial \mathbf{A}_s^H(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{N}_s^{-1} \mathbf{A}_s(\boldsymbol{\theta})\right]_{i,k} = \left.\frac{dF_M(f)}{df}\right|_{f=\theta_i-\theta_k} \delta(i,p) \\ [\mathcal{B}_{p,q}(\boldsymbol{\theta})]_{i,k} &\triangleq \left[\frac{\partial \mathbf{A}_s^H(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{N}_s^{-1} \frac{\partial \mathbf{A}_s(\boldsymbol{\theta})}{\partial \theta_q}\right]_{i,k} \\ &= \left.\frac{d^2 F_M(f)}{df^2}\right|_{f=\theta_i-\theta_k} \delta(i,p)\,\delta(k,q)\end{aligned} \tag{7.51}$$

where $\delta(i,j)$ is the Kronecker delta and $F_M(f)$ is the following *sinc* function,

$$F_M(f) \triangleq \begin{cases} M & f = 0 \\ \frac{\sin(\pi M f/2)}{\sin(\pi f/2)} & f \neq 0 \end{cases}.$$

Notice that $\mathcal{B}_p(\boldsymbol{\theta})$ and $\mathcal{B}_{p,q}(\boldsymbol{\theta})$ in (7.51) are derived resorting to the differential property of the Fourier transform, i.e.,

$$-j2\pi \mathfrak{F}\{nv[n]\} = \frac{dV(f)}{df}$$

where $V(f) = \mathfrak{F}\{v[n]\} = \sum_n v[n] e^{-j2\pi f n}$ is the Fourier transform of a given sequence $v[n]$.

In the studied space-time signal model, the observation $\mathbf{y}$ can be increased by augmenting either the number of antennas (M) or the number of snapshots (N_s), where M and N_s are the number of rows of $\mathbf{A}_s(\boldsymbol{\theta})$ and $\mathbf{A}_t$, respectively. The asymptotic performance of the GML and CML direction-of-arrival estimators have already been studied in [Sto89] [Sto90a][Vib95] when $M \to \infty$ and in [Sto89] [Sto90a][Ott92][Car94] when $N_s \to \infty$. In the following two sections, the aforementioned study is extended to the optimal second-order DOA estimator deduced in Chapter 4.

Infinite number of antennas

If the number of sensors is increased ($M \to \infty$), the asymptotic MSE matrix for the optimal and the GML estimators is computed in Appendix 7.K, having that

$$\begin{aligned}\mathbf{B}_{UCRB}(\boldsymbol{\theta}), \mathbf{B}_{gml}(\boldsymbol{\theta}), \mathbf{B}_{bque}(\boldsymbol{\theta}) &= \frac{6}{\pi^2 M^3} \frac{\sigma_w^2}{\mathrm{Tr}\left(\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t\right)} \mathbf{I}_P + o\left(M^{-3}\right) \\ &= \frac{6}{\pi^2 M^3 N_s E_s/N_0} \mathbf{I}_P + o\left(N_s^{-1} M^{-3}\right) \end{aligned} \tag{7.52}$$

with $\sigma_w^2 = N_0$ the noise double-sided spectral density and E_s the energy of the received symbols. In the last expression, we have taken into account that

$$\lim_{N_s \to \infty} \frac{1}{N_s} \mathrm{Tr}\left(\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t\right) = E_s.$$

This result was previously obtained in [Sto89] for the conditional model (Section 2.3). Moreover, it has been proved in [Sto90a, R8-R9] that the conditional and unconditional model yield efficient estimates when the number of sensors or the SNR goes to infinity. The analysis in Appendix 7.K focuses on the unconditional model and becomes an extension to the concise solution provided in [Sto89]. In this appendix, the asymptotic value of the off-diagonal entries of $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$ as well as the non-Gaussian terms in $\mathbf{B}_{gml}(\boldsymbol{\theta})$ and $\mathbf{B}_{bque}(\boldsymbol{\theta})$ is calculated, concluding that they are totally negligible if $M \to \infty$.

Also, notice that (7.52) coincides with the modified CRB (MCRB) for the problem of carrier frequency-offset estimation [Men97, Eq. 2.4.23][Rif74], which is known to be attained by means of data-aided (DA) methods. In both cases, the estimator tries to infere the frequency of an infinitely long sinusoid either in the space domain (DOA) or in the time domain (DA frequency synchronization). Nonetheless, in array signal processing, the array size is implicityly limited by the narrowband and far-field assumptions [Vib95].

Infinite number of snapshots

When the number of observed symbols N_s is large, it is shown in [Ott92][Car94] that most second-order DOA estimators in the literature –based on both the conditional and unconditional model– are asymptotically robust. This means that the covariance matrix of the estimation error is independent of the sources statistical distribution provided that $N_s \to \infty$. This statement implies that the higher-order term $\mathbf{X}_{gml}(\mathbf{K})$ (7.8) is negligible for $N_s \to \infty$ whatever the content of matrix $\mathbf{K}$. However, it was shown in Section 6.5 that the knowledge of $\mathbf{K}$ can be exploited to improve significantly the estimator accuracy when *multiple constant-modulus* sources transmit towards the array from near directions. Actually, this result was already pointed out in [Ott92] where the authors stated that *'[...] a Gaussian distribution of the emitter signals represents the*

worst case. Any other distribution would tipically give better estimates, provided the appropiate ML method is used.'

In Appendix 7.L, the asymptotic expressions of $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$, $\mathbf{B}_{gml}(\boldsymbol{\theta})$ and $\mathbf{B}_{bque}(\boldsymbol{\theta})$ are obtained when the number of received symbols N_s tends to infinity whereas the number of antennas M is kept constant. In that case, it is shown that $\Gamma(\mathbf{K})$ is exactly zero in the single user case (Appendix 7.L). On the other hand, if there are multiple users transmitting towards the array ($P > 1$), both $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ and $\Gamma(\mathbf{K})$ are proportional to the number of snapshots and, therefore, $\Gamma(\mathbf{K})$ does not disappear as $N_s \to \infty$. In addition, it is found that, at high SNR, the second term $\Gamma(\mathbf{K})$ is proportional to σ_w^{-2} *if and only if* $\rho = 1$. If so, the contribution of $\Gamma(\mathbf{K})$ remains as the SNR is increased. This important result is formalized in the next equations. If E_s/N_0 and N_s go to infinity, it is proved in Appendix 7.Lthat

$$\left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q}, \left[\mathbf{B}_{gml}^{-1}(\boldsymbol{\theta})\right]_{p,q} = 2N_s\frac{E_s}{N_0}\operatorname{Re}\operatorname{Tr}\left(\mathcal{B}_{p,q}(\boldsymbol{\theta}) - \mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta})\mathcal{B}_q^H(\boldsymbol{\theta})\right) + o(N_s)$$

$$\left[\mathbf{B}_{bque}^{-1}(\boldsymbol{\theta})\right]_{p,q} = \begin{cases} \left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q} + o(N_s) & \rho \neq 1 \\ \left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q} + [\Gamma(\mathbf{K})]_{p,q} + o(N_s) & \rho = 1 \end{cases}$$

with

$$[\Gamma(\mathbf{K})]_{p,q} = 2\xi N_s\frac{E_s}{N_0}\operatorname{Tr}\left(\mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta}) \odot Dg^{-1}\left[\mathcal{B}^{-1}(\boldsymbol{\theta})\right] \odot \mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta})\right)\delta(p,q) + o(N_s)$$

being the result a function of the cross-correlation of the P users signatures and their derivatives (7.51). On the other hand, the constant $\xi \leq 1$ is a function of the temporal correlation of the received signal (7.91) and is unitary in the uncorrelated case (Appendix 7.L).

7.5 Simulations

In this section, the asymptotic studies carried out in this chapter are validated via computer simulations.

7.5.1 SNR asymptotic results for the BQUE and GML estimators

To evaluate the asymptotic performance of the BQUE and GML small-error estimators when the SNR goes to zero and infinity, the problem of DOA estimation is adopted (see Section 6.5). The angle of arrival of two users is estimated with a linear array of four elements ($M = 4$). A single snapshot is taken at the matched filter output ($N_s = 1$). Assuming perfect timing synchronization and ISI-free received pulses, the estimator MSE becomes inversely proportional to the number of integrated snapshots. In Fig. 7.1 and Fig. 7.2, the sum of the variance of the

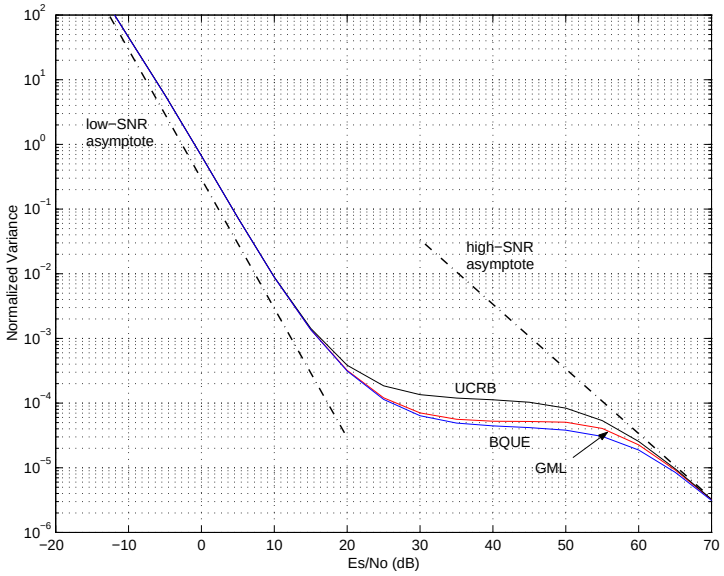

Figure 7.1: Normalized variance for the GML and BQUE DOA estimators in case of having two 16-QAM sources transmitting from ±0.5 degrees. The low- and high-SNR limits computed in this chapter are plotted as well.

two users, i.e.,

$$\begin{aligned} VAR_{bque} &= \mathrm{Tr}\left(\mathbf{B}_{bque}\left(\boldsymbol{\theta}\right)\right) \\ VAR_{gml} &= \mathrm{Tr}\left(\mathbf{B}_{gml}\left(\boldsymbol{\theta}\right)\right), \end{aligned} \tag{7.53}$$

is evaluated as a function of the E_s/N_0 per user when the two users are located at $\pm 0.5^o$ from the broadside. For more details the reader is referred to Section 6.5.

In these figures, it is shown how the asymptotic expressions deduced in this chapter predict exactly the low and high SNR performance of the studied quadratic small-error estimators. In Fig. 7.1, the Gaussian assumption is shown to be optimal at low and high SNR whereas minor losses are observed in the middle of these extremes. It has been checked that the BQUE converges to the GML performance when the alphabet dimension is augmented (e.g., 64-QAM). On the other hand, if the constellation has constant modulus (e.g., MPSK or CPM), the Gaussian assumption is found to yield important losses when the SNR exceeds a given critical value or threshold determined by the array size (Fig. 7.2). The position of this SNR threshold is actually independent of the number of processed snapshots.

Additional simulations have been carried out for the CPM modulation obtaining the same curves than in Fig. 7.2. Therefore, it seems that the only relevant feature of the nuisance parameters for DOA estimation is their constant amplitude.

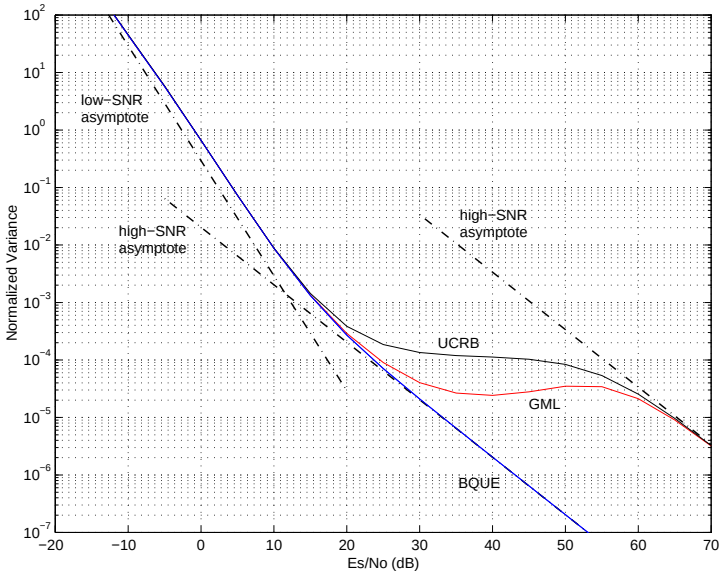

Figure 7.2: Normalized variance for the GML and BQUE DOA estimators in case of having two MPSK sources transmitting from ± 0.5 degrees. The low- and high-SNR limits computed in this chapter are plotted as well.

Regarding Fig. 7.2, another important remark is that the low- and high-SNR asymptotes can be combined to lower bound the performance of any second-order technique in case of constant-modulus alphabets. Finally, notice that the UCRB only predicts the *asymptotic* performance of the GML estimator. However, in both figures, the GML estimator outperforms the UCRB for intermediate SNRs.

7.5.2 SNR asymptotic results for the large-error estimators

In this section, the large-error frequency-offset estimators presented in Section 3.4 are simulated again. The low- and high-SNR asymptotic expressions deduced in this chapter are validated for the 16-QAM, MPSK and MSK modulations. In all the simulations, the rank of $\mathbf{G} = \mathbf{E} \odot \mathbf{A}\mathbf{A}^H$ is full (Appendix 3.D). A uniform prior with $\Delta = 0.4$ is considered. Although this prior is rather informative and the variance floor was not observed in Fig. 3.7 for the MSK modulation, its existence is evidenced in Fig. 7.3. In this figure, it is also shown how the Gaussian assumption leads to a higher variance floor at high SNR. Comparing Figs. 7.3, 7.4 and 7.5, the floor level depends on the modulation at hand. This statement is true for both the optimal estimator and the one deduced under the Gaussian assumption, although the latter is not represented in Figs. 7.4 and 7.5 for the sake of clarity.

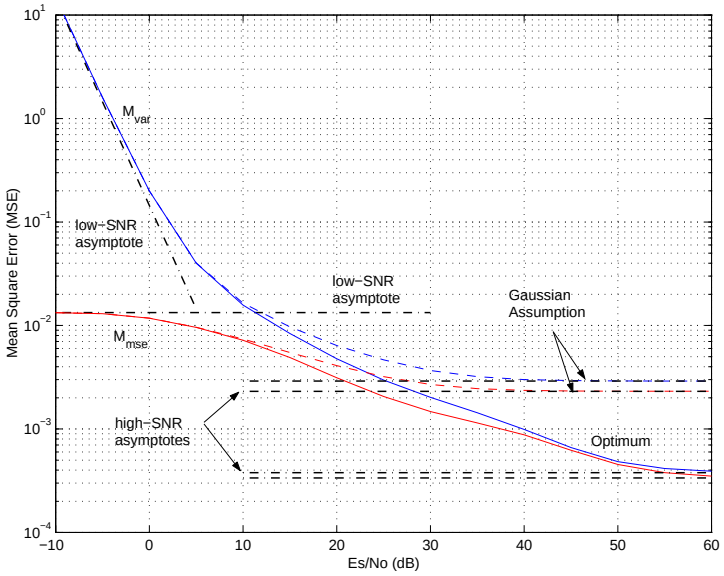

Figure 7.3: Mean square error for the MMSE and minimum variance frequency-offset estimators for the MSK modulation ($N_{ss} = 2$, $M = 8$, $K = 5$). The estimators based on the Gaussian assumption as well as all the low- and high-SNR limits are indicated.

7.5.3 Large sample asymptotic results for the BQUE and GML estimators

In this section, the large sample study in Section 7.4 is validated numerically. In the first two figures (Fig. 7.6 and 7.7), the *normalized* variance is computed for the optimal second-order timing and frequency estimators deduced in Chapter 4 under the small-error condition. The normalization consists in multiplying the estimator variance by the number of processed symbols, i.e., $N_s = M/N_{ss}$. The estimators variance is simulated for different data lengths and is compared to the asymptotic variance obtained from the large sample study ($N_s \to \infty$) in Section 7.4.4. The Gaussian assumption is optimal in all the simulations except in the timing synchronization problem (Fig. 7.6). In that case, the Gaussian assumption exhibits a higher variance floor (self-noise) when the noise subspace of matrix $\mathbf{A}(\boldsymbol{\theta})$ is null ($M \leq K$).

Regarding the DOA estimation problem, the large sample study presented in Section 7.4.5 is validated via simulation for the same scenario considered in Section 7.5.1. In the first simulations (Fig. 7.8 and 7.9), the estimator variance (7.53) is evaluated for different values of ρ considering an array of four antennas an a single snapshot[7]. The performance associated to a hypothetical super-Gaussian constellation with $\rho = 10$ is also depicted in Fig. 7.9, although all the alphabets

[7]Remember that estimator variance is inversely proportional to the number of processed snapshots whatever the value of ρ. Therefore, all the results and conclusions are still correct if $N_s \to \infty$.

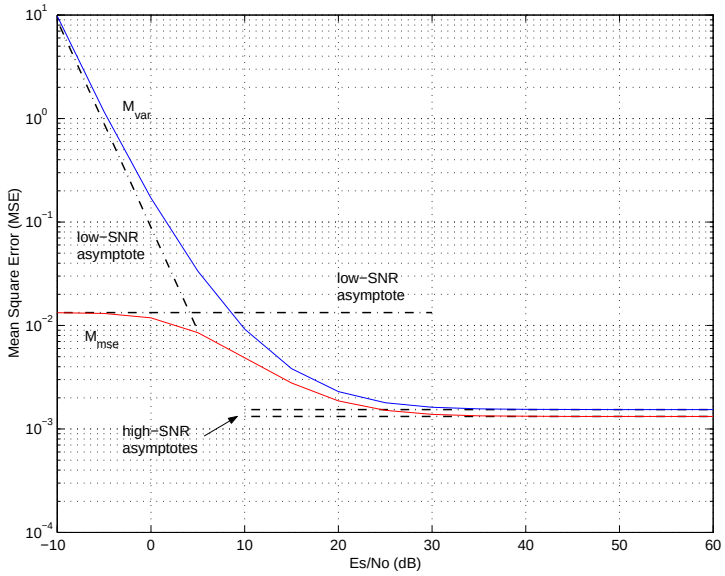

Figure 7.4: Mean square error for the MMSE and minimum variance frequency-offset estimators for the 16-QAM modulation with roll-off 0.75 ($N_{ss} = 2$, $M = 8$, $K = 8$). The low- and high-SNR limits computed in this chapter are plotted as well.

of interest in digital communications are sub-Gaussian ($\rho < 2$).

Regarding Fig. 7.8 and Fig. 7.9, one concludes that the asymptotic expression derived in (7.52) for $M \to \infty$,

$$\mathbf{B}_{UCRB}(\boldsymbol{\theta}), \mathbf{B}_{gml}(\boldsymbol{\theta}), \mathbf{B}_{bque}(\boldsymbol{\theta}) = \frac{6}{\pi^2 M^3 N_s E_s/N_0}\mathbf{I}_P + o(N_s^{-1}M^{-3}),$$

is attained for practical SNRs in case of having constant-amplitude nuisance parameters ($\rho = 1$) *even if the number of antennas is very small* ($M = 4$). Notice that the optimality at high SNR is verified irrespective of the users angular separation if one compares Fig. 7.8 and Fig. 7.9. Nonetheless, minor discrepancies are observed in Fig. 7.8 due to sinc-like beam pattern when M is finite (7.51).

On the other hand, if $\rho > 1$, the estimator performance at high SNR converge to the (Gaussian) UCRB, that corresponds to $\rho = 2$. It can be seen that the larger is ρ and the closer are the sources, the lower is the E_s/N_0 from which the convergence to the UCRB is manifested. Moreover, the closer are the users the more significant is the loss incurred by the Gaussian assumption in case of constant-modulus nuisance parameters.

These conclusions are manifested again when the estimator variance is evaluated as a function of M (Figs. 7.10-7.12). In that case, the UCRB attains the asymptotic limit (7.52) if the number of antennas goes to infinity ($M \to \infty$). On the other hand, when the nuisance parameters have

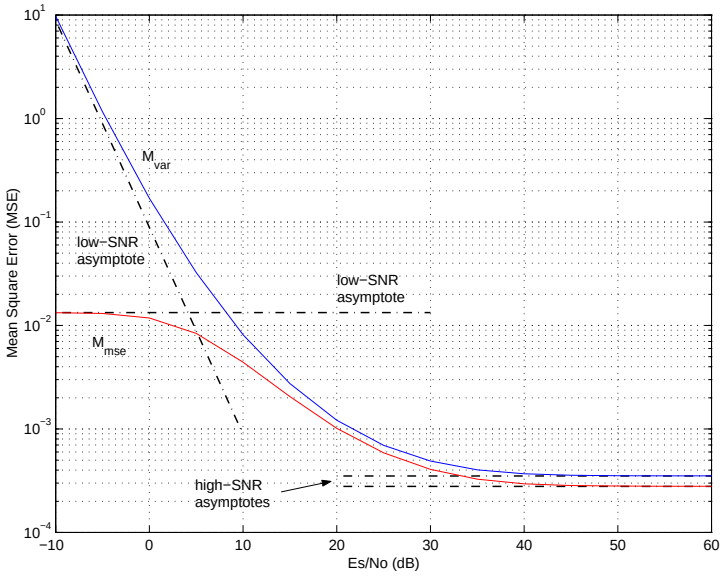

Figure 7.5: Mean square error for the MMSE and minimum variance frequency-offset estimators with MPSK symbols and roll-off 0.75 ($N_{ss} = 2$, $M = 8$, $K = 8$). The low- and high-SNR limits computed in this chapter are plotted as well.

constant modulus ($\rho = 1$), the optimal second-order estimator attains (7.52) *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 $\mathbf{B}_{bque}(\boldsymbol{\theta})$ departs from (7.52) is inversely proportional to the angular separation of the users. Specifically, this critical value occurs when $\Gamma(\mathbf{K})$ (7.46) attains its maximum value. In case of having two users, the critical value of M corresponds to the first maximum of (7.90) in Appendix 7.K that, asymptotically, takes place at $M = 0.5/\left|\theta_1 - \theta_2\right|$. For example, using equation (7.90), the referred threshold should take place at $M \simeq 20$ and $M \simeq 100$ in Fig. 7.11 and Fig. 7.12, respectively.

The GML performance coincides with the UCRB unless the angular separation is reduced (Fig. 7.12). When the number of antennas is less than 20, the GML outperforms the UCRB bound for both the MPSK and 16-QAM modulations. Indeed, the UCRB is severely degraded when the number of antennas is less than 10 whereas the variance of the BQUE and GML estimators is practically constant for $M < 10$ (Fig. 7.13).

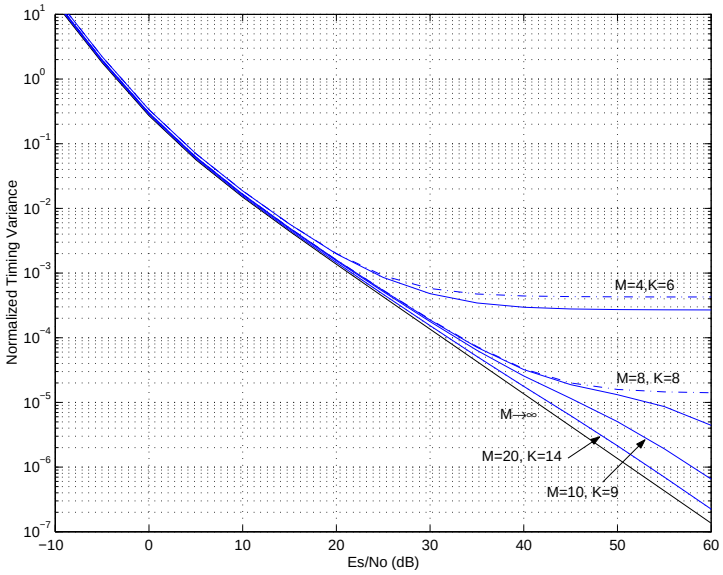

Figure 7.6: Normalized variance for the optimal second-order timing synchronizer in case of the MPSK modulation. The transmitted pulse is a square-root raised cosine with roll-off 0.75, truncated at $\pm 5T$. The observation timel (M) is augmented with $N_{ss} = 2$ constant. The dashed curves correspond to estimator based on the Gaussian assumption.

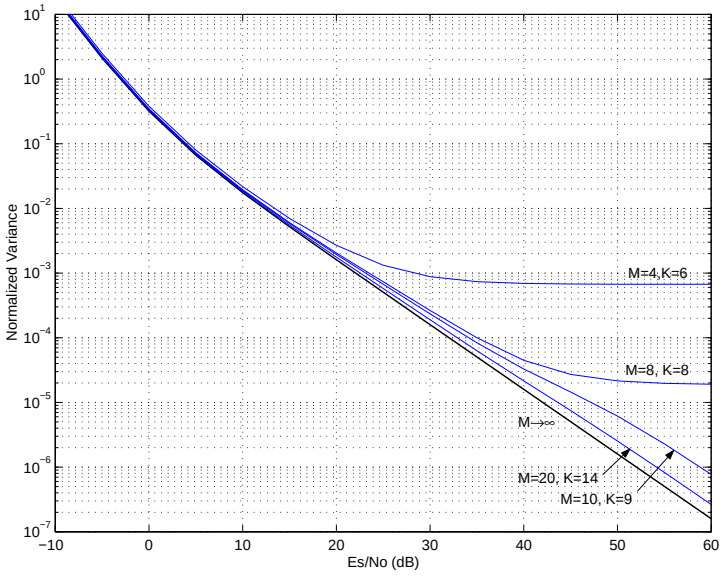

Figure 7.7: Normalized variance for the optimal second-order frequency-offset synchronizer in case of the MPSK modulation. The transmitted pulse is a square-root raised cosine with roll-off 0.75, truncated at $\pm 5T$. The observation interval (M) is augmented with $N_{ss} = 2$ constant.

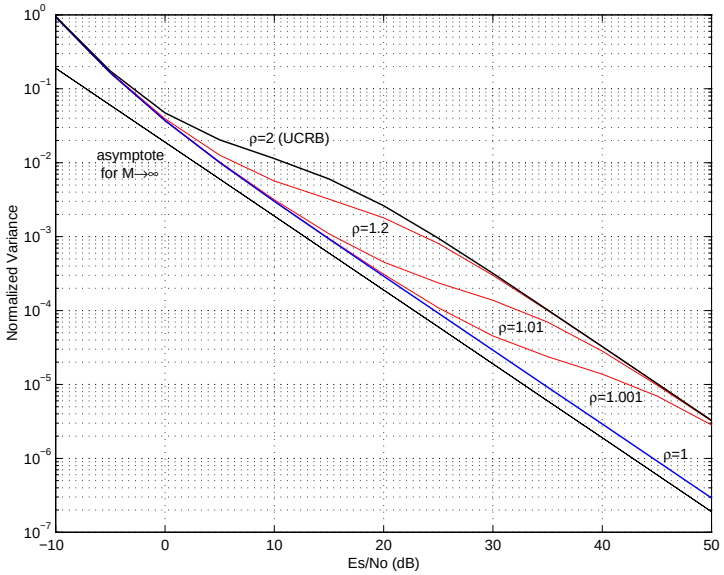

Figure 7.8: Normalized variance for the optimal second-order small-error DOA estimator for different values of ρ. The simulation parameters are $N_{ss} = 1$, Nyquist pulse shaping, $K = 1$, $M = 4$, two users transmitting from ± 5 degrees.

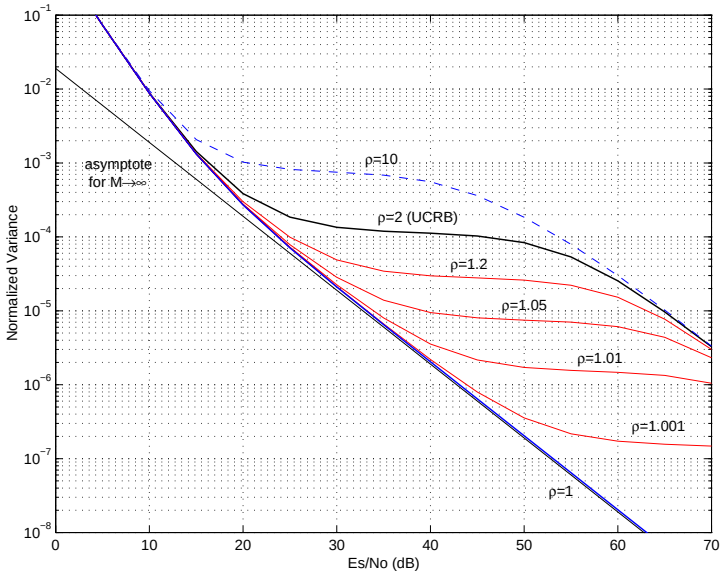

Figure 7.9: Normalized variance for the optimal second-order small-error DOA estimator for different values of ρ. The simulation parameters are $N_{ss} = 1$, Nyquist pulse shaping, $K = 1$, $M = 4$, two users transmitting from ± 0.5 degrees.

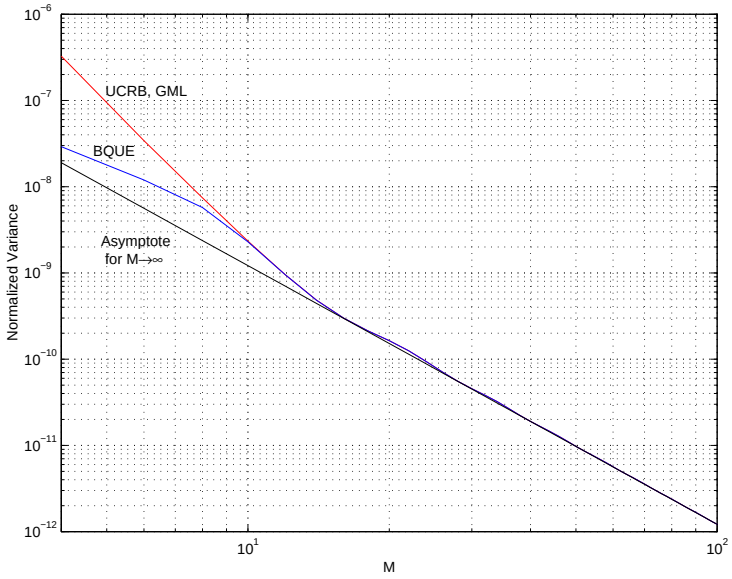

Figure 7.10: Normalized variance for the optimal second-order small-error DOA estimator as a function of M. The simulation parameters are $N_{ss} = 1$, Nyquist pulse shaping, $K = 1$, E_s/N_0=60dB, two MPSK users transmitting from ± 5 degrees.

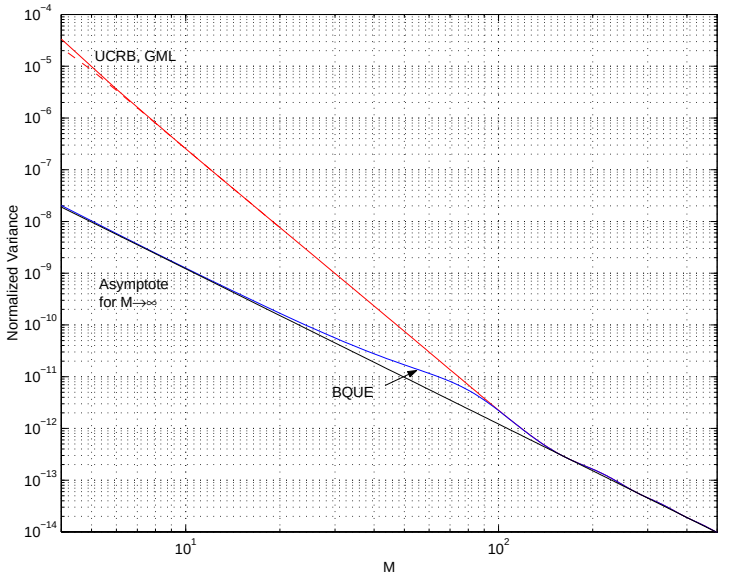

Figure 7.11: Normalized variance for the optimal second-order small-error DOA estimator as a function of M. The simulation parameters are $N_{ss} = 1$, Nyquist pulse shaping, $K = 1$, E_s/N_0=60dB, two MPSK users transmitting from ± 0.5 degrees.

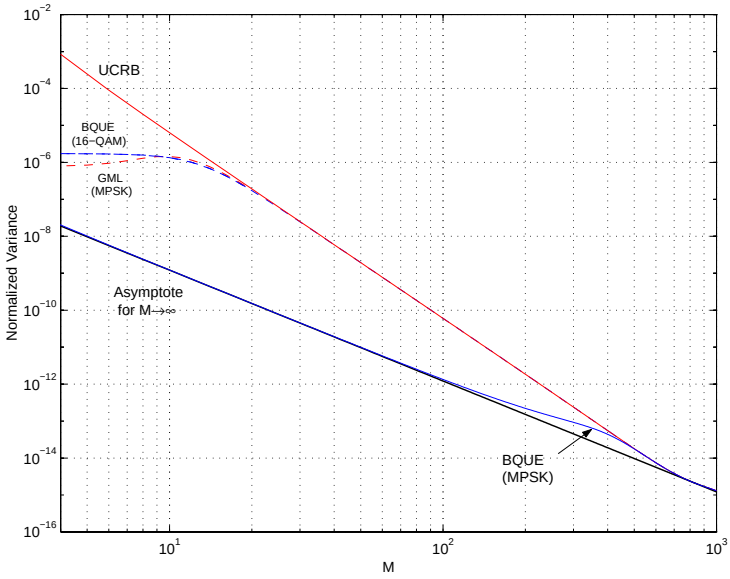

Figure 7.12: Normalized variance for the optimal second-order small-error DOA estimator as a function of M. The simulation parameters are $N_{ss} = 1$, Nyquist pulse shaping, $K = 1$, E_s/N_0=60dB, two MPSK (or 16-QAM) users transmitting from ±0.1 degrees.

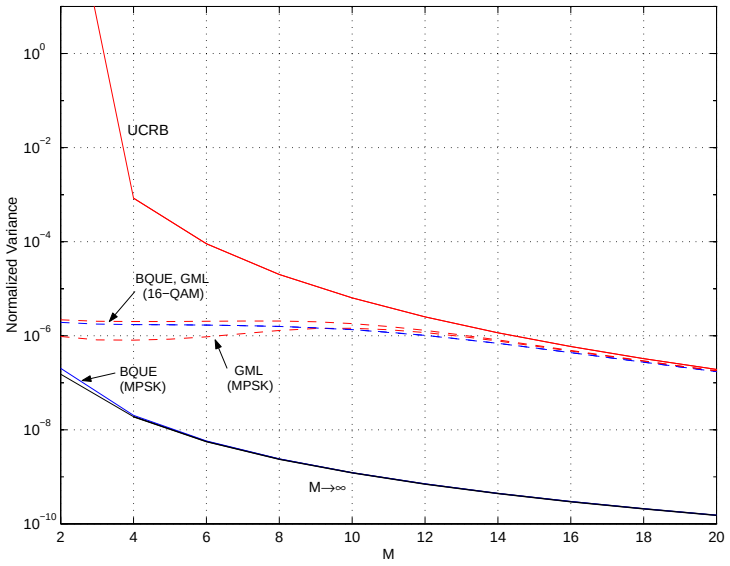

Figure 7.13: Zoom of the previous plot between $M = 2$ and $M = 20$. The simulation parameters are $N_{ss} = 1$, Nyquist pulse shaping, $K = 1$, E_s/N_0=60dB, two MPSK (or 16-QAM) users transmitting from ±0.1 degrees.

7.6 Conclusions

In the previous chapters, the Gaussian assumption was proved to yield significant losses at high SNR if the nuisance parameters had a constant modulus and the observation interval was reduced. In this chapter, the Gaussian assumption has been examined again when the observation interval is increased to infinity. From the Central Limit Theorem, it seems that the data statistics will be irrelevant in this asymptotic case. This intuition is validated in some important estimation problems such as digital synchronization and DOA estimation, if the number of antennas goes to infinity. However, the Gaussian assumption is shown to be suboptimal in some other scenarios. In particular, the Gaussian assumption fails when the estimator suffers from self-noise at high SNR. In that case, the fourth-order information about the nuisance parameters can be exploited to reduce the variance floor, mainly when the nuisance parameters have constant amplitude.

By considering the fourth-order information of the nuisance parameters, second-order DOA estimators are able to attain the asymptotic performance associated to an infinite number of antennas, even if the array is very short. On the other hand, the Gaussian assumption yields an important loss that is a function of the sources angular separation. Therefore, in array signal processing, we have concluded that the Gaussian assumption is only optimal if there is a single source or the number of antennas goes to infinity.

Finally, in the problem of blind channel identification, some improvement is also expected in the asymptotic case when the transmitted symbols are drawn from a constant-modulus alphabet (Section 6.4).

Appendix 7.A Low-SNR ML scoring implementation

It was obtained in Section 2.4.1 that the log-likelihood function in a low SNR scenario is given by

$$\ln f_{\mathbf{y}}\left(\mathbf{y};\boldsymbol{\theta}\right)=\mathrm{Tr}\left\{\mathbf{R}_w^{-1}\mathbf{A}\left(\boldsymbol{\theta}\right)\mathbf{A}^H\left(\boldsymbol{\theta}\right)\mathbf{R}_w^{-1}\left(\widehat{\mathbf{R}}-\mathbf{R}_w\right)\right\}+o\left(\sigma_w^{-2}\right)$$

that, resorting again to the $\mathrm{vec}\left(\cdot\right)$ operator, can be manipulated as follows

$$\begin{aligned}\ln f_{\mathbf{y}}\left(\mathbf{y};\boldsymbol{\theta}\right)&=\mathrm{vec}^H\left(\mathbf{R}_w^{-1}\mathbf{A}\left(\boldsymbol{\theta}\right)\mathbf{A}^H\left(\boldsymbol{\theta}\right)\mathbf{R}_w^{-1}\right)\left(\widehat{\mathbf{r}}-\mathbf{r}_w\right)+o\left(\sigma_w^{-2}\right)\\&=\mathrm{vec}^H\left(\mathbf{A}\left(\boldsymbol{\theta}\right)\mathbf{A}^H\left(\boldsymbol{\theta}\right)\right)\left(\mathbf{R}_w^{*}\otimes\mathbf{R}_w\right)^{-1}\left(\widehat{\mathbf{r}}-\mathbf{r}_w\right)+o\left(\sigma_w^{-2}\right)\\&=\sigma_w^{-4}\,\mathrm{vec}^H\left(\mathbf{A}\left(\boldsymbol{\theta}\right)\mathbf{A}^H\left(\boldsymbol{\theta}\right)\right)\mathcal{N}^{-1}\left(\widehat{\mathbf{r}}-\mathbf{r}_w\right)+o\left(\sigma_w^{-2}\right)\end{aligned}$$

where $\widehat{\mathbf{r}}$ and $\mathbf{r}_w$ are the vectorization of the Hermitian matrices $\widehat{\mathbf{R}}$ and $\mathbf{R}_w$, respectively, and the following relations have been applied

$$\begin{aligned}\mathrm{vec}^H(\mathbf{A})\,\mathrm{vec}(\mathbf{B})&=\mathrm{Tr}(\mathbf{A}^H\mathbf{B})=\mathrm{Tr}(\mathbf{B}\mathbf{A}^H)\\\mathrm{vec}\left(\mathbf{A}\mathbf{B}\mathbf{C}^H\right)&=\left(\mathbf{C}^*\otimes\mathbf{A}\right)\mathrm{vec}(\mathbf{B}).\\\mathbf{A}^H\otimes\mathbf{B}^H&=\left(\mathbf{A}\otimes\mathbf{B}\right)^H\\\mathbf{A}^{-1}\otimes\mathbf{B}^{-1}&=\left(\mathbf{A}\otimes\mathbf{B}\right)^{-1}\end{aligned}\tag{7.54}$$

The gradient of the asymptotic log-likelihood function is given by

$$\frac{\partial\ln f_{\mathbf{y}}\left(\mathbf{y};\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}=\sigma_w^{-4}\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{N}^{-1}\left(\widehat{\mathbf{r}}-\mathbf{r}_w\right)+o\left(\sigma_w^{-2}\right)$$

since $\partial\mathbf{R}\left(\boldsymbol{\theta}\right)/\partial\boldsymbol{\theta}=\partial\left[\mathbf{A}\left(\boldsymbol{\theta}\right)\mathbf{A}^H(\boldsymbol{\theta})\right]/\partial\boldsymbol{\theta}$. Next, the Fisher's information matrix is computed as the expected value of the Hessian matrix, obtaining the following asymptotic expression:

$$\begin{aligned}E_{\mathbf{y}}\left\{\frac{\partial^2\ln f_{\mathbf{y}}\left(\mathbf{y};\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right\}&=E_{\mathbf{y}}\left\{\left(\frac{\partial\ln f_{\mathbf{y}}\left(\mathbf{y};\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\ln f_{\mathbf{y}}\left(\mathbf{y};\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right)^H\right\}\\&=\sigma_w^{-8}\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{N}^{-1}\mathbf{Q}\left(\boldsymbol{\theta}\right)\mathcal{N}^{-1}\mathbf{D}_r(\boldsymbol{\theta})+o\left(\sigma_w^{-4}\right)\\&=\sigma_w^{-4}\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{N}^{-1}\mathbf{D}_r(\boldsymbol{\theta})+o\left(\sigma_w^{-4}\right)\end{aligned}$$

using that, at low SNR, the fourth-order matrix $\mathbf{Q}\left(\boldsymbol{\theta}\right)$ is given by

$$\mathbf{Q}\left(\boldsymbol{\theta}\right)=E\left\{\left(\widehat{\mathbf{r}}-\mathbf{r}_w\right)\left(\widehat{\mathbf{r}}-\mathbf{r}_w\right)^H\right\}=\sigma_w^4\mathcal{N}+o\left(\sigma_w^4\right).$$

Therefore, the following scoring recursion, which was presented in equation (2.29),

$$\begin{aligned}\widehat{\boldsymbol{\alpha}}_{k+1}&=\widehat{\boldsymbol{\alpha}}_k+\mathbf{M}^H(\widehat{\boldsymbol{\theta}}_k)\left(\widehat{\mathbf{r}}-\mathbf{r}_w\right)\\\mathbf{M}(\boldsymbol{\theta})&\triangleq\mathcal{N}^{-1}\mathbf{D}_r(\boldsymbol{\theta})\left(\mathbf{D}_r^H(\boldsymbol{\theta})\mathcal{N}^{-1}\mathbf{D}_r(\boldsymbol{\theta})\right)^{-1}\mathbf{D}_g^H(\boldsymbol{\theta})\end{aligned}$$

is known to attain the CRB at low SNR if the small-error condition is verified.

Appendix 7.B High-SNR limit of $\mathbf{R}^{-1}(\boldsymbol{\theta})$ and $\widetilde{\mathbf{R}}^{-1}(\boldsymbol{\theta})$

In this appendix, we consider that $\mathbf{A}(\boldsymbol{\theta})$ is full column rank. In that case, the asymptotic value of $\mathbf{R}^{-1}(\boldsymbol{\theta})$ can be easily obtained by means of the inversion lemma:

$$\begin{aligned}\mathbf{R}^{-1}(\boldsymbol{\theta}) &= \left[\mathbf{A}(\boldsymbol{\theta})\mathbf{A}^H(\boldsymbol{\theta}) + \sigma_w^2\mathbf{N}\right]^{-1} \\ &= \sigma_w^{-2}\mathbf{N}^{-1}\left(\mathbf{I}_M - \mathbf{A}(\boldsymbol{\theta})\left(\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta}) + \sigma_w^2\mathbf{I}_K\right)^{-1}\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\right). \end{aligned} \quad (7.55)$$

At high SNR, the inner inverse can be expanded in a Taylor series around $\sigma_w^2 = 0$, having that[8]

$$\begin{aligned}\left(\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta}) + \sigma_w^2\mathbf{I}_K\right)^{-1} = &\left[\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\right]^{-1} - \sigma_w^2\left[\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\right]^{-2} \\ &+\sigma_w^4\left[\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\right]^{-3} + O\left(\sigma_w^6\right).\end{aligned}$$

Finally, plugging these three terms into (7.55), the high-SNR limit of $\mathbf{R}^{-1}(\boldsymbol{\theta})$ is given by

$$\mathbf{R}^{-1}(\boldsymbol{\theta}) = \sigma_w^{-2}\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) + \mathbf{B}(\boldsymbol{\theta}) - \sigma_w^2\mathbf{B}(\boldsymbol{\theta})\mathbf{N}\mathbf{B}(\boldsymbol{\theta}) + O\left(\sigma_w^4\right) \quad (7.56)$$

where $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})$ and $\mathbf{B}(\boldsymbol{\theta})$ are defined in (7.13)-(7.15), and the following identity has been considered:

$$\left[\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\right]^{-1} = \mathbf{A}^{\#}(\boldsymbol{\theta})\mathbf{N}\left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H. \quad (7.57)$$

The key property of $\mathbf{R}^{-1}(\boldsymbol{\theta})$ is that, asymptotically, it holds that

$$\begin{aligned}\mathbf{A}^H(\boldsymbol{\theta})\mathbf{R}^{-1}(\boldsymbol{\theta}) &= \mathbf{A}^H(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta}) + O\left(\sigma_w^2\right) = \mathbf{A}^{\#}(\boldsymbol{\theta}) + O\left(\sigma_w^2\right) \\ \mathbf{R}^{-1}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) &= \mathbf{B}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) + O\left(\sigma_w^2\right) = \left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H + O\left(\sigma_w^2\right) \\ \mathbf{A}^H(\boldsymbol{\theta})\mathbf{R}^{-1}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) &= \mathbf{A}^H(\boldsymbol{\theta})\mathbf{B}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) + O\left(\sigma_w^2\right) = \mathbf{I}_K + O\left(\sigma_w^2\right)\end{aligned} \quad (7.58)$$

because, by definition,

$$\begin{aligned}\mathbf{A}^H(\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) &= \mathbf{0} \\ \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta}) &= \mathbf{0}.\end{aligned} \quad (7.59)$$

[8] The following relation has been considered to obtain the terms of the Taylor expansion:

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

To conclude this appendix, the asymptotic value of $\mathcal{R}^{-1}(\boldsymbol{\theta})$ is obtained from (7.56) as indicated now:

$$\begin{aligned}\mathcal{R}^{-1}(\boldsymbol{\theta}) &= \mathbf{R}^{*-1}(\boldsymbol{\theta}) \otimes \mathbf{R}^{-1}(\boldsymbol{\theta}) \\ &= \sigma_w^{-4}\left[\mathbf{P}_{\mathbf{A}}^{\perp *}(\boldsymbol{\theta}) \otimes \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\right] \\ &\quad +\sigma_w^{-2}\left[\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) + \mathbf{P}_{\mathbf{A}}^{\perp *}(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta})\right] \\ &\quad +\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta}) \\ &\quad -\sigma_w^2\left(\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta})\mathbf{N}\mathbf{B}(\boldsymbol{\theta}) + (\mathbf{B}(\boldsymbol{\theta})\mathbf{N}\mathbf{B}(\boldsymbol{\theta}))^* \otimes \mathbf{B}(\boldsymbol{\theta})\right) \\ &\quad +O\left(\sigma_w^4\right).\end{aligned} \tag{7.60}$$

The key property of $\mathcal{R}^{-1}(\boldsymbol{\theta})$ is that the term proportional to σ_w^{-4} is orthogonal to $\mathrm{vec}(\mathbf{A}(\boldsymbol{\theta})\mathbf{X})$, $\mathrm{vec}(\mathbf{X}\mathbf{A}^H(\boldsymbol{\theta}))$, $\mathbf{X} \otimes \mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{X}$ for any matrix $\mathbf{X}$ on account of (7.59). In particular, this is true for the matrix of derivatives $\mathbf{D}_r(\boldsymbol{\theta})$ in (7.4) and for $\mathcal{A}(\boldsymbol{\theta})$ in (7.2). On the other hand, the first term on σ_w^{-2} is orthogonal to $\mathrm{vec}(\mathbf{A}(\boldsymbol{\theta})\mathbf{X})$ and $\mathbf{X} \otimes \mathbf{A}(\boldsymbol{\theta})$ whereas the second one is orthogonal to $\mathrm{vec}(\mathbf{X}\mathbf{A}^H(\boldsymbol{\theta}))$ and $\mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{X}$, based again on (7.59).

The same properties in (7.58) can be stated for $\mathcal{R}^{-1}(\boldsymbol{\theta})$, having that

$$\begin{aligned}\mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta}) &= \mathcal{A}^H(\boldsymbol{\theta})\left[\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta})\right] + O\left(\sigma_w^2\right) = \mathcal{A}^{\#}(\boldsymbol{\theta}) + O\left(\sigma_w^2\right) \\ \mathcal{R}^{-1}(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta}) &= \left[\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta})\right]\mathcal{A}(\boldsymbol{\theta}) + O\left(\sigma_w^2\right) = \left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H + O\left(\sigma_w^2\right) \\ \mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta}) &= \mathbf{I}_{K^2} + O\left(\sigma_w^2\right)\end{aligned} \tag{7.61}$$

using the following definition of pseudoinverse:

$$\mathcal{A}^{\#}(\boldsymbol{\theta}) \triangleq \left(\mathcal{A}^H(\boldsymbol{\theta})\mathcal{N}^{-1}\mathcal{A}(\boldsymbol{\theta})\right)^{-1}\mathcal{A}^H(\boldsymbol{\theta})\mathcal{N}^{-1} = \left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^* \otimes \mathbf{A}^{\#}(\boldsymbol{\theta}).$$

All these properties will be used to simplify the high-SNR expressions in Section 7.3.

Appendix 7.C High-SNR limit of $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ ($\mathbf{K}$ full-rank)

Assuming that $\mathbf{K}$ is invertible, the inversion lemma allows expressing $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ as follows

$$\begin{aligned}\mathbf{Q}^{-1}(\boldsymbol{\theta}) &= \left(\mathcal{A}(\boldsymbol{\theta})\mathbf{K}\mathcal{A}^H(\boldsymbol{\theta}) + \mathcal{R}(\boldsymbol{\theta})\right)^{-1} \\ &= \mathcal{R}^{-1}(\boldsymbol{\theta})\left[\mathbf{I}_{M^2} - \mathcal{A}(\boldsymbol{\theta})\left(\mathbf{K}^{-1} + \mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta})\right)^{-1}\mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\right] \quad (7.62)\end{aligned}$$

Using (7.61) from Appendix 7.B, it follows that

$$\mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta}) = \mathbf{I}_{K^2} + O\left(\sigma_w^2\right)$$

and, therefore, the asymptotic value of $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ is straightforward if $\mathbf{K}^{-1} + \mathbf{I}_{K^2}$ is invertible. In that case, (7.62) becomes

$$\mathbf{Q}^{-1}(\boldsymbol{\theta}) = \mathcal{R}^{-1}(\boldsymbol{\theta}) - \left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H\left(\mathbf{K}^{-1} + \mathbf{I}\right)^{-1}\mathcal{A}^{\#}(\boldsymbol{\theta}) + O\left(\sigma_w^2\right)$$

using that $\mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta}) = \mathcal{A}^{\#}(\boldsymbol{\theta})$ (7.61). Notice that the term depending on $\mathbf{K}$ is negligible unless the dominant terms of $\mathcal{R}^{-1}(\boldsymbol{\theta})$ in (7.60) are null.

However, $\mathbf{K}^{-1} + \mathbf{I}_{K^2}$ becomes singular in case of CPM modulations and, therefore, the inner inverse in (7.62) is a little more involved. In that case, the terms in (7.60) depending on σ_w^2 must also be considered obtaining that

$$\mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta}) = \mathbf{I}_{K^2} - \sigma_w^2\mathbf{U}(\boldsymbol{\theta}) + O\left(\sigma_w^4\right)$$

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

$$\begin{aligned}\mathbf{U}(\boldsymbol{\theta}) &\triangleq \mathcal{A}^H(\boldsymbol{\theta})\left[\mathbf{B}^*(\boldsymbol{\theta})\otimes\mathbf{B}(\boldsymbol{\theta})\mathbf{N}\mathbf{B}(\boldsymbol{\theta}) + \left[\mathbf{B}(\boldsymbol{\theta})\mathbf{N}\mathbf{B}(\boldsymbol{\theta})\right]^*\otimes\mathbf{B}(\boldsymbol{\theta})\right]\mathcal{A}(\boldsymbol{\theta}) \\ &= \mathbf{I}_K\otimes\left[\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\right]^{-1} + \left[\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\right]^{*-1}\otimes\mathbf{I}_K, \quad (7.63)\end{aligned}$$

that is simplified applying the associative property of the Kronecker product,

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

and using then the results in (7.57) and (7.58).

Thus, the inverse in (7.62) can be solved computing the "economy-size" diagonalization of $\mathbf{K}^{-1} + \mathbf{I}_{K^2}$ as follows

$$\mathbf{K}^{-1} + \mathbf{I}_{K^2} = \mathbf{V}\left(\Sigma^{-1} + \mathbf{I}_{K^2}\right)\mathbf{V}^H$$

where Σ is the diagonal matrix containing the eigenvalues of $\mathbf{K}$ that are different from -1 and the columns of $\mathbf{V}$ are the associated eigenvectors. Then, the inversion lemma can be applied once more to obtain

$$\begin{aligned}\left[\mathbf{K}^{-1} + \mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta})\right]^{-1} &= \left[\mathbf{V}\left(\Sigma^{-1} + \mathbf{I}\right)\mathbf{V}^H - \sigma_w^2\mathbf{U}(\boldsymbol{\theta})\right]^{-1} + O(1) \\ &= -\sigma_w^{-2}\mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta}) + O(1)\end{aligned}$$

where the orthogonal projector $\mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta})$ onto the subspace spanned by $\mathbf{V}$ is defined as

$$\mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta}) \triangleq \mathbf{U}^{-1}(\boldsymbol{\theta}) \left[\mathbf{I} - \mathbf{V}\left(\mathbf{V}^H \mathbf{U}^{-1}(\boldsymbol{\theta})\,\mathbf{V}\right)^{-1} \mathbf{V}^H \mathbf{U}^{-1}(\boldsymbol{\theta})\right]. \tag{7.64}$$

As it was argumented for the projector $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})$ in (7.14), 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 that

$$\mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta})\,\mathbf{V} = \mathbf{0}$$
$$\mathbf{V}^H \mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta}) = \mathbf{0},$$

and, thus, $\mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta})$ is the 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, we obtain that

$$\mathbf{Q}^{-1}(\boldsymbol{\theta}) = \mathcal{R}^{-1}(\boldsymbol{\theta}) + \sigma_w^{-2}\left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H \mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta})\,\mathcal{A}^{\#}(\boldsymbol{\theta}) + O(1)$$

using that $\mathcal{A}^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta}) = \mathcal{A}^{\#}(\boldsymbol{\theta})$ (7.61).

In the following, the orthogonal projector $\mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta})$ will be referred to as $\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})$ in order to emphasize the dependence on the kurtosis matrix $\mathbf{K}$, i.e.,

$$\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta}) \triangleq \mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta}).$$

Appendix 7.D High-SNR limit of $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ ($\mathbf{K}$ singular)

If $\mathbf{K}$ is singular, as it happens when the nuisance parameters are drawn from a circular constellation (3.12), the inversion lemma cannot be applied directly and it is necessary to diagonalize previously the matrix $\mathbf{K}$ as indicated next:

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

where Σ_K is the diagonal matrix containing the non-zero eigenvalues of $\mathbf{K}$, and $\mathbf{V}_K$ the associated eigenvectors. Therefore, the study carried out in Appendix 7.C is still correct if $\mathbf{K}^{-1}$ and $\mathcal{A}(\boldsymbol{\theta})$ are substituted by Σ_K^{-1} and $\mathcal{A}(\boldsymbol{\theta})\mathbf{V}_K$, respectively. In that case, the inner inverse studied in detail in Appendix 7.C is given by

$$\left(\Sigma_K^{-1} + \mathbf{V}_K^H\mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta})\mathbf{V}_K\right)^{-1} \tag{7.65}$$

with

$$\mathbf{V}_K^H\mathcal{A}^H(\boldsymbol{\theta})\mathcal{R}^{-1}(\boldsymbol{\theta})\mathcal{A}(\boldsymbol{\theta})\mathbf{V}_K = \mathbf{I} - \sigma_w^2\mathcal{U}(\boldsymbol{\theta}) + O\left(\sigma_w^4\right) \tag{7.66}$$

and

$$\mathcal{U}(\boldsymbol{\theta}) \triangleq \mathbf{V}_K^H\left(\mathbf{I}_K \otimes \left(\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\right)^{-1} + \left(\mathbf{A}^H(\boldsymbol{\theta})\mathbf{N}^{-1}\mathbf{A}(\boldsymbol{\theta})\right)^{*-1} \otimes \mathbf{I}_K\right)\mathbf{V}_K. \tag{7.67}$$

According to this result, the same two scenarios of Appendix 7.C can be distinguished:

1. If $\Sigma_K^{-1} + \mathbf{I}$ is invertible, it follows that

$$\mathbf{Q}^{-1}(\boldsymbol{\theta}) = \mathcal{R}^{-1}(\boldsymbol{\theta}) - \left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H\mathbf{V}_K\left(\Sigma_K^{-1} + \mathbf{I}\right)^{-1}\mathbf{V}_K^H\mathcal{A}^{\#}(\boldsymbol{\theta}) + O\left(\sigma_w^2\right)$$

   and the second term becomes negligible at high SNR.

2. Otherwise, if $\Sigma_K^{-1} + \mathbf{I}$ is singular, the diagonal matrix $\Sigma_K^{-1} + \mathbf{I}$ has to be diagonalized as $\mathcal{V}\Sigma\mathcal{V}^H$ where Σ is the diagonal matrix containing the eigenvalues of $\mathbf{K}$ different from -1 and $\mathcal{V}$ are the vectors of the canonical basis $\{\mathbf{e}_k\}$ selecting the position of these eigenvalues in Σ_K. Formally, the k-th diagonal entry of Σ_K different from -1 is selected by means of the vector $\mathbf{e}_k$ defined as

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

   In that case, the term $-\sigma_w^2\mathcal{U}(\boldsymbol{\theta})$ in (7.66) must be considered in the computation of (7.65), yielding

$$\begin{aligned}\left(\Sigma_K^{-1} + \mathbf{I} - \sigma_w^2\mathcal{U}(\boldsymbol{\theta})\right)^{-1} &= \left(\mathcal{V}\Sigma\mathcal{V}^H - \sigma_w^2\mathcal{U}(\boldsymbol{\theta})\right)^{-1} \\ &= -\sigma_w^{-2}\mathbf{P}_{\mathcal{V}}^{\perp}(\boldsymbol{\theta}) + O(1)\end{aligned}$$

where $\mathbf{P}_{\mathcal{V}}^{\perp}(\boldsymbol{\theta})$ is the following orthogonal projector:

$$\mathbf{P}_{\mathcal{V}}^{\perp}(\boldsymbol{\theta}) \triangleq \mathcal{U}^{-1}(\boldsymbol{\theta}) \left[\mathbf{I} - \mathcal{V}\left(\mathcal{V}^H \mathcal{U}^{-1}(\boldsymbol{\theta})\,\mathcal{V}\right)^{-1} \mathcal{V}^H \mathcal{U}^{-1}(\boldsymbol{\theta})\right]$$

with $\mathcal{U}(\boldsymbol{\theta})$ defined in (7.67).

Finally, we obtain that

$$\mathbf{Q}^{-1}(\boldsymbol{\theta}) = \mathcal{R}^{-1}(\boldsymbol{\theta}) + \sigma_w^{-2}\left[\mathcal{A}^{\#}(\boldsymbol{\theta})\right]^H \mathbf{V}_K \mathbf{P}_{\mathcal{V}}^{\perp}(\boldsymbol{\theta})\,\mathbf{V}_K^H \mathcal{A}^{\#}(\boldsymbol{\theta}) + O(1)$$

using again that $\mathcal{A}^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta}) = \mathcal{A}^{\#}(\boldsymbol{\theta})$ (7.61).

The matrix $\mathbf{V}_K \mathbf{P}_{\mathcal{V}}^{\perp}(\boldsymbol{\theta})\,\mathbf{V}_K^H$ is also a projector onto the subspace generated by the eigenvectors of $\mathbf{K}$ associated to the eigenvalue -1. This projection is carried out in two steps. First, the matrices $\mathbf{V}_K$ and $\mathbf{V}_K^H$ are projecting onto the subspace associated to the eigenvalues of $\mathbf{K}$ different form 0. Afterwards, $\mathbf{P}_{\mathcal{V}}^{\perp}(\boldsymbol{\theta})$ is projecting onto the subspace associated to those eigenvalues different from -1.

In the following, the orthogonal projector $\mathbf{V}_K \mathbf{P}_{\mathcal{V}}^{\perp}(\boldsymbol{\theta})\,\mathbf{V}_K^H$ will be referred to as $\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta})$ in order to emphasize the dependence on the kurtosis matrix $\mathbf{K}$, i.e.,

$$\mathbf{P}_{\mathbf{K}}^{\perp}(\boldsymbol{\theta}) \triangleq \mathbf{V}_K \mathbf{P}_{\mathcal{V}}^{\perp}(\boldsymbol{\theta})\,\mathbf{V}_K^H.$$

Appendix 7.E High-SNR results with $\mathbf{A}(\boldsymbol{\theta})$ singular

Depending on the rank of $\mathbf{A}(\boldsymbol{\theta}) \in \mathbb{C}^{M\times K}$, two singular situations can be distinguished:

1. If the rank of $\mathbf{A}(\boldsymbol{\theta})$ is equal to M with $M \leq K$, the high-SNR limit of $\mathbf{R}^{-1}(\boldsymbol{\theta})$ and $\mathcal{R}^{-1}(\boldsymbol{\theta})$ is independent of the noise variance and is simply given by

$$\mathbf{R}^{-1}(\boldsymbol{\theta}) = \left[\mathbf{A}(\boldsymbol{\theta})\,\mathbf{A}^H(\boldsymbol{\theta})\right]^{-1} + o(1)$$
$$\mathcal{R}^{-1}(\boldsymbol{\theta}) = \left[\mathcal{A}(\boldsymbol{\theta})\,\mathcal{A}^H(\boldsymbol{\theta})\right]^{-1} + o(1)$$

whereas the asymptotic value of $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ is determined by the rank of the $K^2 \times K^2$ matrix $\mathbf{I}_{K^2} + \mathbf{K}$. If the rank of $\mathbf{I}_{K^2} + \mathbf{K}$ is greater or equal to M^2, the inverse of $\mathbf{Q}(\boldsymbol{\theta})$ is the following constant matrix:

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

In that case, all the MSE matrices in (7.5) and (7.6) suffer a serious floor at high SNR. This situation arises when there are less observations than nuisance parameters (i.e., $M \leq K$) and the noise subspace becomes null.

On the other hand, if the rank of $\mathbf{I}_{K^2} + \mathbf{K}$ is less than M^2 (assuming that $M \leq K$), the use of the fourth-order information avoids the variance floor at high SNR because in that case $\mathcal{A}(\boldsymbol{\theta})\,(\mathbf{I}_{K^2} + \mathbf{K})\,\mathcal{A}^H(\boldsymbol{\theta})$ is not invertible and the terms of $\mathcal{R}^{-1}(\boldsymbol{\theta})$ (7.12) proportional to σ_w^2 has to be considered, as done in Appendix 7.C and Appendix 7.D. This situation is only possible if the nuisance parameters have constant modulus as, for example, the MPSK and CPM constellations. In the MPSK case, the rank of $\mathbf{I}_{K^2}+\mathbf{K}$ is exactly K^2-K because

$$\mathbf{K} = (\rho - 2)\,\mathrm{diag}\left(\mathrm{vec}\left(\mathbf{I}_K\right)\right).$$

In the CPM case, the rank reduction is still more significant.

2. If the rank of $\mathbf{A}(\boldsymbol{\theta})$ is lower than $min(M, K)$, the covariance matrix $\mathbf{R}(\boldsymbol{\theta})$ must be diagonalized as follows

$$\mathbf{R}(\boldsymbol{\theta}) = \mathbf{V}_A(\boldsymbol{\theta})\,\Sigma_A(\boldsymbol{\theta})\,\mathbf{V}_A^H(\boldsymbol{\theta}) + \sigma_w^2\mathbf{N} \tag{7.68}$$

where $\Sigma_A(\boldsymbol{\theta})$ is the diagonal matrix having the positive eigenvalues of $\mathbf{A}(\boldsymbol{\theta})\,\mathbf{A}^H(\boldsymbol{\theta})$ and, $\mathbf{V}_A(\boldsymbol{\theta})$ are the associated eigenvectors. Therefore, the inverse of $\mathbf{R}(\boldsymbol{\theta})$ is obtained after applying the inversion lemma to (7.68), obtaining that

$$\mathbf{R}^{-1}(\boldsymbol{\theta}) = \sigma_w^{-2}\mathbf{N}^{-1}\left[\mathbf{I}_{M^2} - \mathbf{V}_A(\boldsymbol{\theta})\left(\mathbf{V}_A^H(\boldsymbol{\theta})\,\mathbf{N}^{-1}\mathbf{V}_A(\boldsymbol{\theta}) + \sigma_w^2\Sigma_A^{-1}(\boldsymbol{\theta})\right)^{-1}\mathbf{V}_A^H(\boldsymbol{\theta})\,\mathbf{N}^{-1}\right].$$

Then, similar results to those in Appendix 7.B are obtained with these substitutions:

$$\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \longrightarrow \mathbf{N}^{-1}\left[\mathbf{I}_M - \mathbf{V}_A(\boldsymbol{\theta})\,\mathbf{V}_A^{\#}(\boldsymbol{\theta})\right]$$
$$\mathbf{B}(\boldsymbol{\theta}) \longrightarrow \left[\mathbf{V}_A^{\#}(\boldsymbol{\theta})\right]^H \Sigma_A^{-1}(\boldsymbol{\theta})\,\mathbf{V}_A^{\#}(\boldsymbol{\theta})$$

with $\mathbf{V}_A^{\#}(\boldsymbol{\theta}) \triangleq \left(\mathbf{V}_A^H(\boldsymbol{\theta})\,\mathbf{N}^{-1}\mathbf{V}_A(\boldsymbol{\theta})\right)^{-1}\mathbf{V}_A^H(\boldsymbol{\theta})\,\mathbf{N}^{-1}$. This second situation is observed when some columns of $\mathbf{A}(\boldsymbol{\theta})$, if $M \geq K$, or some rows of $\mathbf{A}(\boldsymbol{\theta})$, if $M \leq K$, are linearly dependent. This is actually the case of the partial response CPM signals simulated in this thesis (e.g., 2REC, 3REC and GMSK).

Appendix 7.F High-SNR UCRB

Using the asymptotic results in (7.12), it follows that

$$\sigma_w^2 \mathbf{D}_r^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) \mathbf{D}_r(\boldsymbol{\theta}) = \mathbf{D}_r^H(\boldsymbol{\theta}) \left[\mathbf{B}^*(\boldsymbol{\theta}) \otimes \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\right] \mathbf{D}_r(\boldsymbol{\theta}) + \mathbf{D}_r^H(\boldsymbol{\theta}) \left[\mathbf{P}_{\mathbf{A}}^{\perp *}(\boldsymbol{\theta}) \otimes \mathbf{B}(\boldsymbol{\theta})\right] \mathbf{D}_r(\boldsymbol{\theta}) + o(1).$$

Then, the entries of this matrix can be simplified as indicated next:

$$\sigma_w^2 \left[\mathbf{D}_r^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) \mathbf{D}_r(\boldsymbol{\theta})\right]_{p,q} = \mathrm{Tr}\left(\frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_q} \mathbf{B}(\boldsymbol{\theta}) + \frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{B}(\boldsymbol{\theta}) \frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_q} \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\right) + o(1), \tag{7.69}$$

bearing in mind that $[\mathbf{D}_r(\boldsymbol{\theta})]_p = \mathrm{vec}(\partial \mathbf{R}(\boldsymbol{\theta}) / \partial \theta_p)$ and using the properties in (7.54). The final expression is simplified because all the matrices in (7.69) are Hermitian. Therefore, if $\partial \mathbf{R}(\boldsymbol{\theta}) / \partial \theta_p$ is decomposed as

$$\frac{\partial \mathbf{R}(\boldsymbol{\theta})}{\partial \theta_p} = \frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{A}^H(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta}) \frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_p},$$

and all the terms including $\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})$ are removed using (7.59), it follows that

$$\sigma_w^2 \left[\mathbf{D}_r^H(\boldsymbol{\theta}) \mathcal{R}^{-1}(\boldsymbol{\theta}) \mathbf{D}_r(\boldsymbol{\theta})\right]_{p,q} = \mathrm{Tr}\left(\frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_p} \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_q}\right) + \mathrm{Tr}\left(\frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_q} \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) \frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_p}\right) + o(1)$$

using that $\mathbf{A}^H(\boldsymbol{\theta}) \mathbf{B}(\boldsymbol{\theta}) \mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}_K$ (7.58). Finally, the matix $\mathcal{B}_1(\boldsymbol{\theta})$ in (7.18) is obtained observing that the last two terms are complex conjugated.

Appendix 7.G High-SNR UCRB variance floor

Using the asymptotic results in (7.12), it follows that

$$\mathbf{D}_r^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta}) = \mathbf{D}_r^H(\boldsymbol{\theta})\left[\mathbf{B}^*(\boldsymbol{\theta})\otimes\mathbf{B}(\boldsymbol{\theta})\right]\mathbf{D}_r(\boldsymbol{\theta}) + o(1).$$

Then, the entries of this matrix can be simplified as indicated next:

$$\left[\mathbf{D}_r^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})\right]_{p,q} = \mathrm{Tr}\left(\frac{\partial\mathbf{R}(\boldsymbol{\theta})}{\partial\theta_p}\mathbf{B}(\boldsymbol{\theta})\,\frac{\partial\mathbf{R}(\boldsymbol{\theta})}{\partial\theta_q}\mathbf{B}(\boldsymbol{\theta})\right) + o(1),$$

bearing in mind that $[\mathbf{D}_r(\boldsymbol{\theta})]_p = \mathrm{vec}\left(\partial\mathbf{R}(\boldsymbol{\theta})/\partial\theta_p\right)$ and using the properties listed in (7.54). The final expression is simplified because all the matrices in the last equation are Hermitian. Thus, if $\partial\mathbf{R}(\boldsymbol{\theta})/\partial\theta_p$ is decomposed as

$$\frac{\partial\mathbf{R}(\boldsymbol{\theta})}{\partial\theta_p} = \frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_p}\mathbf{A}^H(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta})\,\frac{\partial\mathbf{A}^H(\boldsymbol{\theta})}{\partial\theta_p},$$

and the relations in (7.58) are applied, it follows that

$$\begin{aligned}\left[\mathbf{D}_r^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})\right]_{p,q} = \mathrm{Tr}\Bigg(&\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_p}\mathbf{A}^{\#}(\boldsymbol{\theta})\,\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_q}\mathbf{A}^{\#}(\boldsymbol{\theta})\\ &+\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_p}\frac{\partial\mathbf{A}^H(\boldsymbol{\theta})}{\partial\theta_q}\left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H\mathbf{A}^{\#}(\boldsymbol{\theta})\\ &+\frac{\partial\mathbf{A}^H(\boldsymbol{\theta})}{\partial\theta_p}\left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H\frac{\partial\mathbf{A}^H(\boldsymbol{\theta})}{\partial\theta_q}\left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H\\ &+\frac{\partial\mathbf{A}^H(\boldsymbol{\theta})}{\partial\theta_p}\left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H\mathbf{A}^{\#}(\boldsymbol{\theta})\,\frac{\partial\mathbf{A}(\boldsymbol{\theta})}{\partial\theta_q}\Bigg) + o(1)\end{aligned}$$

taking into that $\mathbf{B}(\boldsymbol{\theta}) \triangleq \left[\mathbf{A}^{\#}(\boldsymbol{\theta})\right]^H\mathbf{A}^{\#}(\boldsymbol{\theta})$. Finally, the matrix $\mathcal{B}_2(\boldsymbol{\theta})$ in (7.20) is obtained observing that the third and fourth terms are the complex conjugated versions of the first and second terms[9].

[9]Notice that $Tr(\mathbf{AB}) = Tr(\mathbf{BA})$.

Appendix 7.H High-SNR study in feedforward second-order estimation

Looking at the MSE matrices Σ_{mse}, Σ_{var}, Σ'_{mse} and Σ'_{var} in (7.5) and (7.6), the high-SNR limit of $\mathcal{R}^{-1}$, $\mathbf{Q}^{-1}$, $\left(\mathcal{R}+\widetilde{\mathbf{Q}}\right)^{-1}$ and $\left(\mathbf{Q}+\widetilde{\mathbf{Q}}\right)^{-1}$ has to be computed. In all these four cases, the following inversion problem must be solved:

$$\left(\mathbf{T}+\sigma_w^2\mathbf{U}+\sigma_w^4\mathcal{N}\right)^{-1}$$

where the expression of $\mathbf{T}$ depends on the inverse that is being solved (7.32) and

$$\mathbf{U} \triangleq [\mathbf{G}^* \otimes \mathbf{N} + \mathbf{N}^* \otimes \mathbf{G}] \tag{7.70}$$

$$\mathbf{G} \triangleq E_{\boldsymbol{\theta}} \left\{ \mathbf{A}\left(\boldsymbol{\theta}\right) \mathbf{A}^H\left(\boldsymbol{\theta}\right) \right\}. \tag{7.71}$$

The Bayesian expectation is found to augment the rank of the involved matrices. This effect is actually negative since it reduces the dimension of the noise subspace. In the limit, if the constant term $\mathbf{T}$ became full-rank, the estimators would exhibit the typical variance floor at high SNR. However, in the sequel, we will assume that $\mathbf{T}$ *is always rank defficient*.

It is worth noting that the kurtosis matrix $\mathbf{K}$ appearing in $\mathbf{Q}$ (7.26), reduces the rank of $\mathbf{T}$ and, therefore, the dimension of the noise subspace is increased. This aspect was also addressed in the first point of Appendix 7.E

In order to evaluate the above inverse, the "economy-size" diagonalization of $\mathbf{T} = \mathbf{V}_T \Sigma_T \mathbf{V}_T^H$ is calculated and the auxiliary matrix $\mathbf{X} \triangleq \mathbf{U} + \sigma_w^2 \mathcal{N}$ is introduced. Then, the inversion lemma is invoked as it was done in Appendix 7.B, obtaining

$$\begin{aligned}\left(\mathbf{T}+\sigma_w^2\mathbf{U}+\sigma_w^4\mathcal{N}\right)^{-1} &= \left(\mathbf{V}_T\Sigma_T\mathbf{V}_T^H+\sigma_w^2\mathbf{X}\right)^{-1} \\ &= \sigma_w^{-2}\mathbf{P}_{\mathbf{T}}^{\perp}+\mathbf{B}_T+O\left(\sigma_w^2\right)\end{aligned} \tag{7.72}$$

where $\mathbf{B}_T$ is the following matrix:

$$\mathbf{B}_T \triangleq \left[\mathbf{V}_T^{\#}\right]^H \Sigma_T^{-1} \mathbf{V}_T^{\#} \tag{7.73}$$

with

$$\begin{aligned}\mathbf{V}_T^{\#} &\triangleq \left(\mathbf{V}_T^H\mathbf{X}^{-1}\mathbf{V}_T\right)^{-1}\mathbf{V}_T^H\mathbf{X}^{-1} \\ \mathbf{P}_{\mathbf{T}}^{\perp} &\triangleq \mathbf{X}^{-1}\left(\mathbf{I}_{M^2}-\mathbf{V}_T\mathbf{V}_T^{\#}\right)\end{aligned}$$

the generalization of the pseudoinverse and the projection matrix onto the noise subspace of $\mathbf{T}$, respectively.

In most problems, the rank of $\mathbf{G} \triangleq E_{\boldsymbol{\theta}} \left\{ \mathbf{A}(\boldsymbol{\theta}) \mathbf{A}^H(\boldsymbol{\theta}) \right\}$ grows rapidly due to the Bayesian expectation and, eventually, matrix $\mathbf{G}$ becomes full-rank. In that case, taking into account (7.70), matrix $\mathbf{U}$ is also invertible so that

$$\lim_{\sigma_w^2 \to 0} \mathbf{X}^{-1} = \mathbf{U}^{-1},$$

bearing in mind that $\mathbf{X} \triangleq \mathbf{U} + \sigma_w^2 \mathcal{N}$.

When this happens (i.e., $\mathbf{G}$ is full rank), if the first term $\sigma_w^{-2} \mathbf{P}_{\mathbf{T}}^{\perp}$ survives when (7.72) is multiplied by $\widetilde{\mathbf{Q}}$ or $\mathbf{S} = \widetilde{\mathbf{Q}}\mathbf{M}$ in (7.5)-(7.6), it is posssible to have self-noise free estimates. Otherwise, if $\widetilde{\mathbf{Q}} \in span(\mathbf{T})$, the estimator exhibits the typical variance floor because the surviving term $\mathbf{B}_T$ in (7.72) is constant at high SNR.

When the Gaussian assumption is adopted, it is shown in Appendix 7.I that

$$\begin{aligned} \widetilde{\mathbf{Q}} = & E_{\boldsymbol{\theta}} \left\{ \mathcal{A}(\boldsymbol{\theta}) \operatorname{vec}(\mathbf{I}_K) \operatorname{vec}^H(\mathbf{I}_K) \mathcal{A}^H(\boldsymbol{\theta}) \right\} \\ & - E_{\boldsymbol{\theta}} \left\{ \mathcal{A}(\boldsymbol{\theta}) \right\} \operatorname{vec}(\mathbf{I}_K) \operatorname{vec}^H(\mathbf{I}_K) E_{\boldsymbol{\theta}}^H \left\{ \mathcal{A}(\boldsymbol{\theta}) \right\} \end{aligned} \tag{7.74}$$

always lies in the subpace generated by

$$\begin{aligned} \mathbf{T}_3 &\triangleq E_{\boldsymbol{\theta}} \left\{ \mathcal{A}(\boldsymbol{\theta}) \mathcal{A}^H(\boldsymbol{\theta}) \right\} + \widetilde{\mathbf{Q}} \\ \mathbf{T}_4 &\triangleq E_{\boldsymbol{\theta}} \left\{ \mathcal{A}(\boldsymbol{\theta}) \mathcal{A}^H(\boldsymbol{\theta}) \right\}, \end{aligned}$$

which are the matrices $\mathbf{T}$ appearing in the MMSE and minimum variance estimators deduced under the Gaussian assumption (7.6). This result is independent of the actual parameterization and the nuisance parameters distribution. Consequently, if $\mathbf{G}$ is full rank, the Gaussian assumption always suffers from self-noise at high SNR (7.29), and the level of the variance floor is determined by $\mathbf{X}_{var}(\mathbf{K})$ and $\mathbf{X}_{mse}(\mathbf{K})$ (7.30).

Regarding the optimal estimators in (7.5), the cumulant matrix $\mathbf{K}$ is able to reduce the rank of

$$\begin{aligned} \mathbf{T}_1 &\triangleq E_{\boldsymbol{\theta}} \left\{ \mathcal{A}(\boldsymbol{\theta}) (\mathbf{I}_{K^2} + \mathbf{K}) \mathcal{A}^H(\boldsymbol{\theta}) \right\} + \widetilde{\mathbf{Q}} \\ \mathbf{T}_2 &\triangleq E_{\boldsymbol{\theta}} \left\{ \mathcal{A}(\boldsymbol{\theta}) (\mathbf{I}_{K^2} + \mathbf{K}) \mathcal{A}^H(\boldsymbol{\theta}) \right\} \end{aligned}$$

if the nuisance parameters have constant modulus. Unfortunately, this reduction is usually insufficient to move $\widetilde{\mathbf{Q}}$ out of the span of $\mathbf{T}_1$ or $\mathbf{T}_2$ (7.32). In that case, the optimal large-error estimators in (7.5) also exhibit a variance floor at high SNR (7.29). The level of this variance floor depends again on the kurtosis matrix $\mathbf{K}$.

Appendix 7.I High-SNR MSE floor under the Gaussian assumption

In this appendix, it is proved that $\widetilde{\mathbf{Q}}$ does not increases the rank of these two matrices (7.32):

$$\begin{aligned}\mathbf{T}_4 &\triangleq E_{\boldsymbol{\theta}}\left\{\mathcal{A}\left(\boldsymbol{\theta}\right)\mathcal{A}^H\left(\boldsymbol{\theta}\right)\right\}\\ \mathbf{T}_3 &\triangleq E_{\boldsymbol{\theta}}\left\{\mathcal{A}\left(\boldsymbol{\theta}\right)\mathcal{A}^H\left(\boldsymbol{\theta}\right)\right\}+\widetilde{\mathbf{Q}}.\end{aligned}$$

This implies that

$$\begin{aligned}rank\left(\mathbf{T}_4\right) &= rank\left(\mathbf{T}_4+\widetilde{\mathbf{Q}}\right)\\ rank\left(\mathbf{T}_3\right) &= rank\left(\mathbf{T}_3+\widetilde{\mathbf{Q}}\right).\end{aligned}$$

Regarding the first statement, it is found that

$$\begin{aligned}\widetilde{\mathbf{Q}} = &E_{\boldsymbol{\theta}}\left\{\mathcal{A}\left(\boldsymbol{\theta}\right)\mathrm{vec}\left(\mathbf{I}_K\right)\mathrm{vec}^H\left(\mathbf{I}_K\right)\mathcal{A}^H\left(\boldsymbol{\theta}\right)\right\}\\ &-E_{\boldsymbol{\theta}}\left\{\mathcal{A}\left(\boldsymbol{\theta}\right)\right\}\mathrm{vec}\left(\mathbf{I}_K\right)\mathrm{vec}^H\left(\mathbf{I}_K\right)E_{\boldsymbol{\theta}}^H\left\{\mathcal{A}\left(\boldsymbol{\theta}\right)\right\}\end{aligned}$$

is the sum of infinitessimal terms like this:

$$\begin{aligned}&\alpha_1\mathcal{A}\left(\boldsymbol{\theta}_1\right)\mathrm{vec}\left(\mathbf{I}_K\right)\mathrm{vec}^H\left(\mathbf{I}_K\right)\mathcal{A}^H\left(\boldsymbol{\theta}_1\right)+\alpha_2\mathcal{A}\left(\boldsymbol{\theta}_2\right)\mathrm{vec}\left(\mathbf{I}_K\right)\mathrm{vec}^H\left(\mathbf{I}_K\right)\mathcal{A}^H\left(\boldsymbol{\theta}_2\right)\\ &-\alpha_3\mathcal{A}\left(\boldsymbol{\theta}_1\right)\mathrm{vec}\left(\mathbf{I}_K\right)\mathrm{vec}^H\left(\mathbf{I}_K\right)\mathcal{A}^H\left(\boldsymbol{\theta}_2\right)-\alpha_3\mathcal{A}\left(\boldsymbol{\theta}_2\right)\mathrm{vec}\left(\mathbf{I}_K\right)\mathrm{vec}^H\left(\mathbf{I}_K\right)\mathcal{A}^H\left(\boldsymbol{\theta}_1\right), \quad (7.75)\end{aligned}$$

corresponding to two arbitrary values of $\boldsymbol{\theta}$, namely $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, with

$$\begin{aligned}\alpha_1 &= f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_1\right) > 0\\ \alpha_2 &= f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_2\right) > 0\\ \alpha_3 &= f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_1\right) f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}_2\right) > 0\end{aligned}$$

the associated probability densities.

It can be shown that (7.75) is contained in the span of the following matrix:

$$\alpha_1\mathcal{A}\left(\boldsymbol{\theta}_1\right)\mathrm{vec}\left(\mathbf{I}_K\right)\mathrm{vec}^H\left(\mathbf{I}_K\right)\mathcal{A}^H\left(\boldsymbol{\theta}_1\right)+\alpha_2\mathcal{A}\left(\boldsymbol{\theta}_2\right)\mathrm{vec}\left(\mathbf{I}_K\right)\mathrm{vec}^H\left(\mathbf{I}_K\right)\mathcal{A}^H\left(\boldsymbol{\theta}_2\right).$$

Then, if $\mathbf{T}_4 \triangleq E_{\boldsymbol{\theta}}\left\{\mathcal{A}\left(\boldsymbol{\theta}\right)\mathcal{A}^H\left(\boldsymbol{\theta}\right)\right\}$ is decomposed in the same way, $\mathbf{T}_4$ becomes the sum of infinitessimal terms such as

$$\alpha_1\mathcal{A}\left(\boldsymbol{\theta}_1\right)\mathbf{I}_{K^2}\mathcal{A}^H\left(\boldsymbol{\theta}_1\right)+\alpha_2\mathcal{A}\left(\boldsymbol{\theta}_2\right)\mathbf{I}_{K^2}\mathcal{A}^H\left(\boldsymbol{\theta}_2\right).$$

Therefore, $\widetilde{\mathbf{Q}}$ must lie in the span of $\mathbf{T}_4$ on account of the following relationship:

$$\mathbf{A}\,\mathrm{vec}\left(\mathbf{B}\right)\mathrm{vec}^H\left(\mathbf{B}\right)\mathbf{C} \in span\left\{\mathbf{A}\left(\mathbf{B}\otimes\mathbf{B}\right)\mathbf{C}\right\}$$

that is hold for arbitrary matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$.

Finally, if $\widetilde{\mathbf{Q}} \in span\left(\mathbf{T}_4\right)$, it belongs necessarily to the span of $\mathbf{T}_3 = \mathbf{T}_4 + \widetilde{\mathbf{Q}}$.

Appendix 7.J Performance limits in second-order frequency estimation

When the received signal exhibits a frequency offset equal to ν/T, the pulse received at time kT is given by

$$g\left(mT_s - kT;\nu\right) = g_0\left(mT_s - kT;\nu\right) e^{j2\pi\nu k}$$

where T and T_s are the symbol and sample period, respectively, and

$$g_0\left(mT_s;\nu\right) \triangleq p\left(mT_s\right) e^{j2\pi\nu m/N_{ss}}$$

stands for the pulse $p(t)$ received at $t = 0$. The derivative of $g\left(mT_s - kT;\nu\right)$ is given by

$$\frac{\partial g\left(mT_s - kT;\nu\right)}{\partial\nu} = \left[\frac{\partial g_0\left(mT_s - kT;\nu\right)}{\partial\nu} + g_0\left(mT_s - kT;\nu\right)\left(j2\pi k\right)\right] e^{j2\pi\nu k}.$$

Let $\mathbf{A}\left(\nu\right)$ and $\mathbf{A}_0\left(\nu\right)$ stand for the matrices whose columns are delayed replicas of $g\left(mT_s;\nu\right)$ and $g_0\left(mT_s;\nu\right)$, respectively. It can be shown that these two matrices –and their derivatives– are related in the following manner:

$$\begin{aligned}\mathbf{A}\left(\nu\right) &= \mathbf{A}_0\left(\nu\right)\mathrm{diag}\left(\exp\left(j2\pi\mathbf{d}_K\nu\right)\right)\\ \frac{\partial\mathbf{A}\left(\nu\right)}{\partial\nu} &= \left[\frac{\partial\mathbf{A}_0\left(\nu\right)}{\partial\nu} + \mathbf{A}_0\left(\nu\right)\mathrm{diag}\left(j2\pi\mathbf{d}_K\right)\right]\mathrm{diag}\left(\exp\left(j2\pi\mathbf{d}_K\nu\right)\right)\end{aligned}$$

where

$$\mathbf{d}_K = \left[0, 1, \ldots, K-1\right]^T$$

is the K-long vector accounting for the intersymbol phase slope. On the other hand, the *stationary* matrix $\mathbf{A}_0\left(\nu\right)$ only accounts for the phase variation during the observation interval.

Thus, in the frequency estimation problem, $\mathbf{X}_1\left(\boldsymbol{\theta}\right)$ and $\mathbf{X}_{1,1}\left(\boldsymbol{\theta}\right)$ (7.35) have some additional terms depending on $\mathbf{d}_K$ that are listed next:

$$\begin{aligned}\mathbf{X}\left(\nu\right) &= \sigma_w^2\mathbf{A}^H\left(\nu\right)\mathbf{R}^{-1}\left(\nu\right)\mathbf{A}\left(\nu\right) = \mathcal{X}\left(\nu\right)\odot\mathbf{E}^*\left(\nu\right)\\ \mathbf{X}_1\left(\nu\right) &= \sigma_w^2\frac{\partial\mathbf{A}^H\left(\nu\right)}{\partial\nu}\mathbf{R}^{-1}\left(\nu\right)\mathbf{A}\left(\nu\right) = \left[\mathcal{X}_1\left(\nu\right) - \sigma_w^2\,\mathrm{diag}\left(j2\pi\mathbf{d}_K\nu\right)\mathcal{X}\left(\nu\right)\right]\odot\mathbf{E}^*\left(\nu\right)\\ \mathbf{X}_{1,1}\left(\nu\right) &= \sigma_w^2\frac{\partial\mathbf{A}^H\left(\nu\right)}{\partial\nu}\mathbf{R}^{-1}\left(\nu\right)\frac{\partial\mathbf{A}\left(\nu\right)}{\partial\nu} = \left[\mathcal{X}_{1,1}\left(\nu\right) - \sigma_w^2\,\mathrm{diag}\left(j2\pi\mathbf{d}_K\nu\right)\mathcal{X}_1^H\left(\nu\right)\right.\\ &\quad\left.+\mathcal{X}_1\left(\nu\right)\mathrm{diag}\left(j2\pi\mathbf{d}_K\nu\right) - \mathrm{diag}\left(j2\pi\mathbf{d}_K\nu\right)\mathcal{X}\left(\nu\right)\mathrm{diag}\left(j2\pi\mathbf{d}_K\nu\right)\right]\odot\mathbf{E}^*\left(\nu\right)\end{aligned}$$

where

$$\begin{aligned}\mathcal{X}\left(\nu\right) &\triangleq \sigma_w^{-2}\mathbf{A}_0^H\left(\nu\right)\mathbf{R}^{-1}\left(\nu\right)\mathbf{A}_0\left(\nu\right)\\ \mathcal{X}_1\left(\nu\right) &\triangleq \sigma_w^{-2}\frac{\partial\mathbf{A}_0^H\left(\nu\right)}{\partial\nu}\mathbf{R}^{-1}\left(\nu\right)\mathbf{A}_0\left(\nu\right)\\ \mathcal{X}_{1,1}\left(\nu\right) &\triangleq \sigma_w^{-2}\frac{\partial\mathbf{A}_0^H\left(\nu\right)}{\partial\nu}\mathbf{R}^{-1}\left(\nu\right)\frac{\partial\mathbf{A}_0\left(\nu\right)}{\partial\nu}\end{aligned}$$

are the functions $\mathbf{X}(\boldsymbol{\theta})$, $\mathbf{X}_1(\boldsymbol{\theta})$ and $\mathbf{X}_{1,1}(\boldsymbol{\theta})$ associated to the stationary matrix $\mathbf{A}_0(\nu)$, and $[\mathbf{E}(\nu)]_{i,k} \triangleq \exp(j2\pi(i-k)\nu)$ was introduced in Appendix 3.D.

It can be shown that the new terms on $\mathbf{d}_K$ as well as the factor $\mathbf{E}^*(\nu)$ vanish when $\mathbf{X}(\boldsymbol{\theta})$, $\mathbf{X}_1(\boldsymbol{\theta})$ and $\mathbf{X}_{1,1}(\boldsymbol{\theta})$ are plugged into $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$ (7.36), $\Psi(\mathbf{K})$ (7.38) and $\Gamma(\mathbf{K})$ (7.45). The terms on $\mathbf{d}_K$ are imaginary and they are eliminated when the real part is extracted in $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$, $\Psi(\mathbf{K})$ and $\Gamma(\mathbf{K})$. On the other hand, only the diagonal entries of $\mathbf{E}^*(\nu)$ –which are all equal to 1–are involved in (7.36), (7.38) and (7.44).

The conclusion of this appendix is that, despite the received signal is not stationary in the frequency estimation problem, the asymptotic study can be addressed considering uniquely the *stationary* matrix $\mathbf{A}_0(\nu)$ and its derivatives.

Appendix 7.K Asymptotic study for $M \to \infty$

In this appendix, the asymptotic limit of $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$, $\mathbf{B}_{gml}(\boldsymbol{\theta})$ and $\mathbf{B}_{bque}(\boldsymbol{\theta})$ is derived when the number of antennas goes to infinity. While the asymptotic study of $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$ and $\mathbf{B}_{gml}(\boldsymbol{\theta})$ was already addressed in the literature, the asymptotic study of the optimal second-order DOA estimator is carried out in this appendix for the first time. The most important conclusion is that the term $\Gamma(\mathbf{K})$ that appears in $\mathbf{B}_{bque}(\boldsymbol{\theta})$ is always negligible in front of $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ independently of the nuisance parameters distribution. Indeed, $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ is found to grow as M^3 whereas $\Gamma(\mathbf{K})$ cannot grow faster than M. Moreover, if the nuisance parameters are circular, $\Gamma(\mathbf{K})$ goes to zero as M^{-1} or faster. An interesting conclusion is that the convergence order can be increased in one order when the parameters are drawn from a constant-modulus alphabet. However, this increase is not sufficient in the studied problem because the dominant term, $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$, goes to infinity faster.

When the number of sensors goes to infinity ($M \to \infty$), the *spatial* cross-correlation matrices in (7.51) have the following asymptotic expressions:

$$\begin{aligned}
\mathcal{B}(\boldsymbol{\theta}) &= \mathbf{A}_s^H(\boldsymbol{\theta})\mathbf{N}_s^{-1}\mathbf{A}_s(\boldsymbol{\theta}) = M\mathbf{I}_P + O(1) \\
\mathcal{B}_p(\boldsymbol{\theta}) &= \frac{\partial \mathbf{A}_s^H(\boldsymbol{\theta})}{\partial \theta_p}\mathbf{N}_s^{-1}\mathbf{A}_s(\boldsymbol{\theta}) = \mathcal{B}_p^{\infty}(\boldsymbol{\theta}) + O(1) \\
\mathcal{B}_{p,q}(\boldsymbol{\theta}) &= \frac{\partial \mathbf{A}_s^H(\boldsymbol{\theta})}{\partial \theta_p}\mathbf{N}_s^{-1}\frac{\partial \mathbf{A}_s(\boldsymbol{\theta})}{\partial \theta_q} = \begin{cases} \mathcal{B}_{p,q}^{\infty}(\boldsymbol{\theta}) + o(M^2) & p \neq q \\ \mathcal{B}_{p,q}^{\infty}(\boldsymbol{\theta}) + o(M^3) & p = q \end{cases}
\end{aligned}$$

with

$$[\mathcal{B}(\boldsymbol{\theta})]_{i,k} \triangleq \begin{cases} M & \theta_i = \theta_k \\ \frac{\sin(\pi M(\theta_i-\theta_k)/2)}{\sin(\pi(\theta_i-\theta_k)/2)} & \text{otherwise} \end{cases} \tag{7.76}$$

$$\left[\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\right]_{i,k} \triangleq \begin{cases} 0 & i \neq p, \theta_p = \theta_k \\ \pm\frac{\pi}{2}\frac{\cos(\pi(\theta_p-\theta_k)/2)}{\sin^2(\pi(\theta_p-\theta_k)/2)} & i = p, \theta_p - \theta_k = 1/M, 3/M, ... \\ \frac{\pi}{2}M\frac{\cos(\pi M(\theta_p-\theta_k)/2)}{\sin(\pi(\theta_p-\theta_k)/2)} & \text{otherwise} \end{cases} \tag{7.77}$$

$$\left[\mathcal{B}_{p,q}^{\infty}(\boldsymbol{\theta})\right]_{i,k} \triangleq \begin{cases} 0 & i \neq p \text{ or } k \neq q \\ \frac{\pi^2}{12}M^3 & i = p, k = q, \theta_p = \theta_q \\ \pm\frac{\pi^2}{2}M\frac{\cos(\pi(\theta_p-\theta_q)/2)}{\sin^2(\pi(\theta_p-\theta_q)/2)} & i = p, k = q, \theta_p - \theta_q = 2/M, 4/M, ... \\ \frac{\pi^2}{4}M^2\frac{\sin(\pi M(\theta_p-\theta_q)/2)}{\sin(\pi(\theta_p-\theta_q)/2)} & \text{otherwise} \end{cases}. \tag{7.78}$$

In order to find the asymptotic value of $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$, $\mathbf{B}_{gml}(\boldsymbol{\theta})$ and $\mathbf{B}_{bque}(\boldsymbol{\theta})$, it is necessary to obtain the limit of $\mathbf{X}(\boldsymbol{\theta})$, $\mathbf{X}_p(\boldsymbol{\theta})$ and $\mathbf{X}_{p,q}(\boldsymbol{\theta})$ (7.35) as $M \to \infty$. Before doing so, we have to evaluate the inverse appearing in $\mathbf{X}(\boldsymbol{\theta})$, $\mathbf{X}_p(\boldsymbol{\theta})$ and $\mathbf{X}_{p,q}(\boldsymbol{\theta})$ when the number of antennas goes to infinity. Taking into account that the diagonal entries of

$$\mathbf{B}(\boldsymbol{\theta}) = \mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t \otimes \mathcal{B}(\boldsymbol{\theta})$$

are proportional to M (7.76), it follows that

$$\left(\mathbf{B}\left(\boldsymbol{\theta}\right)+\sigma_w^2\mathbf{I}_{KP}\right)^{-1}=M^{-1}\left(M^{-1}\mathbf{B}\left(\boldsymbol{\theta}\right)+\frac{\sigma_w^2}{M}\mathbf{I}_{KP}\right)^{-1}=\mathbf{B}^{-1}\left(\boldsymbol{\theta}\right)+o\left(M^{-1}\right) \tag{7.79}$$

where the last expression is verified for $\sigma_w^2/M \to 0$. If the resulting inverse is now plugged into (7.35), $\mathbf{X}\left(\boldsymbol{\theta}\right)$ and $\mathbf{X}_p(\boldsymbol{\theta})$ become zero when σ_w^2/M goes to zero[10]. Hence, (7.79) should be expanded in a Taylor series around $\sigma_w^2/M=0$ in order to determine its order of convergence, obtaining

$$\left(\mathbf{B}\left(\boldsymbol{\theta}\right)+\sigma_w^2\mathbf{I}_{KP}\right)^{-1}=\mathbf{B}^{-1}\left(\boldsymbol{\theta}\right)-\sigma_w^2\mathbf{B}^{-2}\left(\boldsymbol{\theta}\right)+\sigma_w^4\mathbf{B}^{-3}\left(\boldsymbol{\theta}\right)+o\left(M^{-3}\right).$$

Plugging now the Taylor series into (7.35), it follows that

$$\begin{aligned}
\mathbf{X}\left(\boldsymbol{\theta}\right)&=\sigma_w^2\mathbf{I}_{KP}-\sigma_w^4\mathbf{B}^{-1}\left(\boldsymbol{\theta}\right)+O\left(M^{-2}\right)\\
&=\sigma_w^2\mathbf{I}_{KP}-\sigma_w^4\left(\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t\right)^{-1}\otimes\mathcal{B}^{-1}\left(\boldsymbol{\theta}\right)+O\left(M^{-2}\right) &(7.80)\\
\mathbf{X}_p(\boldsymbol{\theta})&=\sigma_w^2\mathbf{B}_p\left(\boldsymbol{\theta}\right)\mathbf{B}^{-1}\left(\boldsymbol{\theta}\right)+o\left(1\right)\\
&=\sigma_w^2\mathbf{I}_K\otimes\mathcal{B}_p^{\infty}\left(\boldsymbol{\theta}\right)\mathcal{B}^{-1}\left(\boldsymbol{\theta}\right)+o\left(1\right) &(7.81)\\
\mathbf{X}_{p,q}(\boldsymbol{\theta})&=\begin{cases}\mathbf{B}_{p,q}\left(\boldsymbol{\theta}\right)+o(M^2)=\left(\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t\right)\otimes\mathcal{B}_{p,q}^{\infty}\left(\boldsymbol{\theta}\right)+o(M^2) & p\neq q\\ \mathbf{B}_{p,q}\left(\boldsymbol{\theta}\right)+o(M^3)=\left(\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t\right)\otimes\mathcal{B}_{p,q}^{\infty}\left(\boldsymbol{\theta}\right)+o(M^3) & p=q\end{cases} &(7.82)
\end{aligned}$$

where the inverse of $\mathcal{B}\left(\boldsymbol{\theta}\right)$ has the following asymptotic value:

$$\mathcal{B}^{-1}\left(\boldsymbol{\theta}\right)=M^{-1}\mathbf{I}_P+O\left(M^{-2}\right). \tag{7.83}$$

(Gaussian) Unconditional Cramér-Rao Bound

Using the above results, it can be shown that the *diagonal entries* of $\mathbf{B}_{UCRB}^{-1}\left(\boldsymbol{\theta}\right)$ (7.36) have the following asymptotic value

$$\begin{aligned}
\left[\mathbf{B}_{UCRB}^{-1}\left(\boldsymbol{\theta}\right)\right]_{p,p}&=\left[\mathbf{D}_r^H\left(\boldsymbol{\theta}\right)\mathcal{R}^{-1}\left(\boldsymbol{\theta}\right)\mathbf{D}_r\left(\boldsymbol{\theta}\right)\right]_{p,p}=2\sigma_w^{-4}\,\mathrm{Re}\,\mathrm{Tr}\left(\mathbf{X}\left(\boldsymbol{\theta}\right)\mathbf{X}_{p,p}\left(\boldsymbol{\theta}\right)\right)+o\left(M^3\right)\\
&=2\sigma_w^{-2}\,\mathrm{Re}\,\mathrm{Tr}\left(\mathbf{X}_{p,p}\left(\boldsymbol{\theta}\right)\right)+o\left(M^3\right)\\
&=2\sigma_w^{-2}\,\mathrm{Re}\,\mathrm{Tr}\left(\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t\otimes\mathcal{B}_{p,p}^{\infty}\left(\boldsymbol{\theta}\right)\right)+o\left(M^3\right)\\
&=\frac{\pi^2\sigma_w^{-2}}{6}M^3\,\mathrm{Re}\,\mathrm{Tr}\left(\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t\right)+o\left(M^3\right) &(7.84)
\end{aligned}$$

whereas the *off-diagonal entries* converge to a constant when the number of antennas is aug-

[10]Notice that this condition will be also satisfied at high SNR.

mented. After some tediuos calculations, it can be shown that

$$\begin{aligned}\left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q} &= \left[\mathbf{D}_r^H(\boldsymbol{\theta})\,\mathcal{R}^{-1}(\boldsymbol{\theta})\,\mathbf{D}_r(\boldsymbol{\theta})\right]_{p,q} \\ &= 2\sigma_w^{-4}\,\mathrm{Re}\,\mathrm{Tr}\left(\mathbf{X}_p(\boldsymbol{\theta})\,\mathbf{X}_q(\boldsymbol{\theta}) + \mathbf{X}(\boldsymbol{\theta})\,\mathbf{X}_{p,q}(\boldsymbol{\theta})\right) \\ &= 2\,\mathrm{Tr}\left(\mathbf{I}_K \otimes \left(M^{-2}\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\,\mathcal{B}_q^{\infty}(\boldsymbol{\theta}) - \mathcal{B}^{-1}(\boldsymbol{\theta})\,\mathcal{B}_{p,q}^{\infty}(\boldsymbol{\theta})\right)\right) + o(1) \\ &= 2K\left(M^{-2}\left[\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\right]_{p,q}\left[\mathcal{B}_q^{\infty}(\boldsymbol{\theta})\right]_{q,p} - \left[\mathcal{B}_{p,q}^{\infty}(\boldsymbol{\theta})\right]_{p,q}\left[\mathcal{B}^{-1}(\boldsymbol{\theta})\right]_{q,p}\right) + o(1) \\ &= 2KM^{-2}\left(\left[\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\right]_{p,q}\left[\mathcal{B}_q^{\infty}(\boldsymbol{\theta})\right]_{q,p} + \left[\mathcal{B}_{p,q}^{\infty}(\boldsymbol{\theta})\right]_{p,q}\left[\mathcal{B}(\boldsymbol{\theta})\right]_{q,p}\right) + o(1) \\ &= -\frac{K\pi^2}{2}\frac{\cos\left(\pi M\left(\theta_p-\theta_q\right)\right)}{\sin^2\left(\pi\left(\theta_p-\theta_q\right)/2\right)} + o(1)\end{aligned}$$

assuming in the last equation that $\theta_p - \theta_q$ is not multiple of $1/2M$ with probability one.[11]

Notice that the off-diagonal entries of $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ are constant because the term proportional to M is zero since the diagonal entries of $\mathcal{B}_{p,q}^{\infty}(\boldsymbol{\theta})$ are null for $p \neq q$ (7.78). Therefore, in order to evaluate the trace of $\mathcal{B}^{-1}(\boldsymbol{\theta})\,\mathcal{B}_{p,q}^{\infty}(\boldsymbol{\theta})$, the off-diagonal entries of $\mathcal{B}^{-1}(\boldsymbol{\theta})$ in (7.83) must be taken into account. Thus, $\mathcal{B}^{-1}(\boldsymbol{\theta})$ needs to be expanded in a Taylor series around $M^{-1} = 0$, obtaining

$$\begin{aligned}\mathcal{B}^{-1}(\boldsymbol{\theta}) &= M^{-1}\left(\mathbf{I}_P + M^{-1}\left[\mathcal{B}(\boldsymbol{\theta}) - M\mathbf{I}_P\right]\right)^{-1} \simeq M^{-1}\mathbf{I}_P - M^{-2}\left[\mathcal{B}(\boldsymbol{\theta}) - M\mathbf{I}_P\right] + o\left(M^{-1}\right) \\ &= 2M^{-1}\mathbf{I}_P - M^{-2}\mathcal{B}(\boldsymbol{\theta}) + o\left(M^{-1}\right) \qquad (7.85)\end{aligned}$$

and, using (7.76), we have

$$\left[\mathcal{B}^{-1}(\boldsymbol{\theta})\right]_{q,p} = -M^{-2}\left[\mathcal{B}(\boldsymbol{\theta})\right]_{q,p} + o\left(M^{-2}\right) = M^{-2}\frac{\sin\left(\pi M\left(\theta_p-\theta_q\right)/2\right)}{\sin\left(\pi\left(\theta_p-\theta_q\right)/2\right)} + o\left(M^{-2}\right)$$

Finally, the term $\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\,\mathcal{B}_q^{\infty}(\boldsymbol{\theta})$ is computed taking into account that

$$M^{-1}\left[\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\right]_{p,q} = \frac{\pi}{2}\frac{\cos\left(\pi M\left(\theta_p-\theta_k\right)/2\right)}{\sin\left(\pi\left(\theta_p-\theta_k\right)/2\right)}$$

using (7.77).

Gaussian Maximum Likelihood

In this section, the asymptotic study of $\mathbf{B}_{gml}(\boldsymbol{\theta}) = \mathbf{B}_{UCRB}(\boldsymbol{\theta}) + \mathbf{X}_{gml}(\mathbf{K})$ when $M \to \infty$ is addressed concluding that the second term $\mathbf{X}_{gml}(\mathbf{K})$ is negligible in front of $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$. Therefore, the GML estimator is proved to be robust to the sources' distribution when the number of antennas goes to infinity.

[11] Notice that, if $\theta_p - \theta_q$ were multiple of $1/M$, the final expression could be calculated considering these particular cases of $\mathcal{B}_p(\boldsymbol{\theta})$ and $\mathcal{B}_{p,q}(\boldsymbol{\theta})$ in (7.77)-(7.78). Anyway, the off-diagonal entries are found to become asymptotically constant unless $\theta_p - \theta_q = 0.5/M, 1.5/M, \ldots$ In that case, the constant term is equal to zero and, thus, the convergence order of $\left[\mathbf{B}_{UCRB}^{-1}\right]_{p,q}$ becomes $O(M^{-1})$.

To begin with, let us remind the expression of $\mathbf{X}_{gml}(\mathbf{K})$ in (7.37):

$$\begin{aligned}\mathbf{X}_{gml}(\mathbf{K}) &= \mathbf{B}_{UCRB}(\boldsymbol{\theta})\Psi(\mathbf{K})\,\mathbf{B}_{UCRB}(\boldsymbol{\theta}) \\ [\Psi(\mathbf{K})]_{p,q} &= \sigma_w^{-4}\,\mathrm{vec}^H(\mathbf{Y}_p(\boldsymbol{\theta}))\,\mathbf{K}\,\mathrm{vec}(\mathbf{Y}_q(\boldsymbol{\theta}))\,.\end{aligned}$$

Then, the asymptotic value of $\mathbf{Y}_p(\boldsymbol{\theta})$ is obtained from (7.39), concluding that

$$\begin{aligned}\mathbf{Y}_p(\boldsymbol{\theta}) &= \mathbf{X}(\boldsymbol{\theta})\,\mathbf{X}_p(\boldsymbol{\theta}) + \mathbf{X}_p^H(\boldsymbol{\theta})\,\mathbf{X}(\boldsymbol{\theta}) = \sigma_w^4\mathbf{B}_p(\boldsymbol{\theta})\,\mathbf{B}^{-1}(\boldsymbol{\theta}) + \sigma_w^4\mathbf{B}^{-1}(\boldsymbol{\theta})\,\mathbf{B}_p^H(\boldsymbol{\theta}) + o(1) \\ &= \sigma_w^4\mathbf{I}_K \otimes \left[\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\,\mathcal{B}^{-1}(\boldsymbol{\theta}) + \left(\mathcal{B}^{-1}(\boldsymbol{\theta})\,\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\right)^H\right] + o(1) \\ &= \sigma_w^4\mathbf{I}_K \otimes \left[M^{-1}\mathcal{B}_p^{\infty}(\boldsymbol{\theta}) + \left(M^{-1}\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\right)^H\right] + o(1) \qquad (7.86)\end{aligned}$$

and, taking into account that $\mathcal{B}_p^{\infty}(\boldsymbol{\theta})$ is proportional to M (7.77), the convergence order of $\mathbf{Y}_p(\boldsymbol{\theta})$ is $O(1)$, *at most*. In that case, $\mathbf{X}_{gml}(\mathbf{K})$ decays as $O\left(M^{-6}\right)$ whereas $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$ decreases as $O\left(M^{-3}\right)$ (7.84) when the number of sensors goes to infinity.

Focusing now on those circular alphabets considered in (7.42), it is found that $\mathbf{X}_{gml}(\mathbf{K})$ decays as $O\left(M^{-8}\right)$ because $\Psi(\mathbf{K})$ becomes proportional to M^{-2}, as indicated next:[12]

$$\begin{aligned}[\Psi(\mathbf{K})]_{p,q} &= \sigma_w^{-4}(\rho-2)\,\mathrm{diag}^H(\mathbf{Y}_p(\boldsymbol{\theta}))\,\mathrm{diag}(\mathbf{Y}_q(\boldsymbol{\theta})) \\ &= \sigma_w^{-4}(\rho-2)\,\mathrm{Tr}\left(\mathbf{Y}_p(\boldsymbol{\theta}) \odot \mathbf{Y}_q(\boldsymbol{\theta})\right) \\ &= 4K(\rho-2)\,\mathrm{Tr}\left(\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\,\mathcal{B}^{-1}(\boldsymbol{\theta}) \odot \mathcal{B}_q^{\infty}(\boldsymbol{\theta})\,\mathcal{B}^{-1}(\boldsymbol{\theta})\right) + o\left(M^{-2}\right) \\ &= 4K(\rho-2)\sum_{i\neq p,q}\left[\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\right]_{p,i}\left[\mathcal{B}^{-1}(\boldsymbol{\theta})\right]_{i,p} \odot \left[\mathcal{B}_q^{\infty}(\boldsymbol{\theta})\right]_{q,i}\left[\mathcal{B}^{-1}(\boldsymbol{\theta})\right]_{i,q} + o\left(M^{-2}\right) \\ &= \frac{4K(\rho-2)}{M^2}\sum_{i\neq p,q}\left[\mathcal{B}_p^{\infty}(\boldsymbol{\theta})\right]_{p,i}\left[\mathcal{B}(\boldsymbol{\theta})\right]_{i,p} \odot \left[\mathcal{B}_q^{\infty}(\boldsymbol{\theta})\right]_{q,i}\left[\mathcal{B}(\boldsymbol{\theta})\right]_{i,q} + o\left(M^{-2}\right) \\ &= \begin{cases}\frac{K\pi^2}{4M^2}(\rho-2)\left(\sum\limits_{p\neq q}\frac{\sin(\pi M(\theta_p-\theta_q))}{\sin^2(\pi(\theta_p-\theta_q)/2)}\right)^2 + o\left(M^{-2}\right) & p=q \\ 0 & p\neq q\end{cases} \qquad (7.87)\end{aligned}$$

where the off-diagonal elements of $\mathcal{B}^{-1}(\boldsymbol{\theta})$ in (7.85) are considered again because the diagonal of $\mathcal{B}_p^{\infty}(\boldsymbol{\theta})$ is zero (7.77). Remember that K is the number of columns of matrix $\mathbf{A}_t$ or, in other words, the number of nuisance parameters per user.

Best Quadratic Unbiased Estimator

Thus far, the performance of the GML estimator is shown to be independent of the nuisance parameters distribution when the number of sensors goes to infinity. Next, the BQUE estimator is shown to converge asymptotically to the (Gaussian) UCRB when $M \to \infty$. Specifically, if the nuisance parameters have constant modulus, the non-Gaussian term $\Gamma(\mathbf{K})$ could be proportional

[12] All the matrices are real-valued and the Re{} operator is omitted for simplicity.

to M as the number of antennas is augmented. However, this is not possible if the nuisance parameters are circular. In that case, $\Gamma(\mathbf{K})$ goes to zero as M^{-1}. On the other hand, if the modulus of the nuisance parameters is not constant, $\Gamma(\mathbf{K})$ might be constant but it decays as M^{-2} if the nuisance parameters are circular.

To support this conclusion, we begin by recovering the general expression of $\Gamma(\mathbf{K})$ from (7.45):

$$[\Gamma(\mathbf{K})]_{p,q} \triangleq -\sigma_w^{-4} \operatorname{vec}^H(\mathbf{Y}_p(\boldsymbol{\theta}))\mathbf{V}_K\left(\mathbf{V}_K^H[\mathbf{X}^*(\boldsymbol{\theta})\otimes\mathbf{X}(\boldsymbol{\theta})]\mathbf{V}_K+\sigma_w^4\Sigma_K^{-1}\right)^{-1}\mathbf{V}_K^H\operatorname{vec}(\mathbf{Y}_p(\boldsymbol{\theta}))$$

where $\mathbf{K}=\mathbf{V}_K\Sigma_K\mathbf{V}_K^H$ is the "economy-size" diagonalization of $\mathbf{K}$. Then, bearing in mind that $\mathbf{Y}_p(\boldsymbol{\theta})$ (7.86) is constant in the best case, the asymptotic order of $\Gamma(\mathbf{K})$ is determined by

$$\left(\mathbf{V}_K^H[\mathbf{X}^*(\boldsymbol{\theta})\otimes\mathbf{X}(\boldsymbol{\theta})]\mathbf{V}_K+\sigma_w^4\Sigma_K^{-1}\right)^{-1},$$

that converges to a constant if all the eigenvalues of the kurtosis matrix $\mathbf{K}$ are different from -1. This condition on the eigenvalues of $\mathbf{K}$ is equivalent to the aforementioned constant-modulus condition. In that case, bearing in mind that $\mathbf{X}(\boldsymbol{\theta})=\sigma_w^{-2}\mathbf{I}_{KP}+o(1)$, it is straightforward to obtain

$$\left(\mathbf{V}_K^H[\mathbf{X}^*(\boldsymbol{\theta})\otimes\mathbf{X}(\boldsymbol{\theta})]\mathbf{V}_K+\sigma_w^4\Sigma_K^{-1}\right)^{-1}=\sigma_w^{-4}\left(\mathbf{I}+\Sigma_K^{-1}\right)^{-1}+o(1)$$

and, therefore, $\Gamma(\mathbf{K})$ converges to a constant as $M\to\infty$.

On the other hand, if some eigenvalues of $\mathbf{K}$ are equal to -1, the inverse of $\mathbf{I}+\Sigma_K^{-1}$ does not exist and the second component of $\mathbf{X}(\boldsymbol{\theta})$ (7.80) must be considered. Thus, it follows that

$$\left(\mathbf{V}_K^H[\mathbf{X}^*(\boldsymbol{\theta})\otimes\mathbf{X}(\boldsymbol{\theta})]\mathbf{V}_K+\sigma_w^4\Sigma_K^{-1}\right)^{-1}=\sigma_w^4\left(\mathbf{I}+\Sigma_K^{-1}\right)-2\sigma_w^6\mathbf{U}(\boldsymbol{\theta})+o\left(M^{-1}\right)$$

where the second term,

$$\mathbf{U}(\boldsymbol{\theta})\triangleq\mathbf{V}_K^H\mathbf{B}^{-1}(\boldsymbol{\theta})\mathbf{V}_K=M^{-1}\mathbf{V}_K^H\left[\left(\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t\right)^{-1}\otimes\mathbf{I}_P\right]\mathbf{V}_K+o\left(M^{-1}\right),\tag{7.88}$$

is proportional to M^{-1}. At this point, the inversion lemma should be applied to compute the above inverse, as it was done in Section 7.3. By doing so, the inversion would yield a term proportional to M and, therefore, the non-Gaussian term $\Gamma(\mathbf{K})$ will become proportional to M, as well.

To illustrate this general conclusion, the previous analysis is particularized in case of having circular nuisance parameters. In that case, the non-Gaussian term $\Gamma(\mathbf{K})$ is given in (7.46):

$$[\Gamma(\mathbf{K})]_{p,q}\triangleq-\sigma_w^{-4}\operatorname{diag}^H(\mathbf{Y}_p(\boldsymbol{\theta}))\left(\mathbf{X}^*(\boldsymbol{\theta})\odot\mathbf{X}(\boldsymbol{\theta})+\sigma_w^4(\rho-2)^{-1}\mathbf{I}_{KP}\right)^{-1}\operatorname{diag}(\mathbf{Y}_q(\boldsymbol{\theta})).$$

where

$$\mathbf{X}^*(\boldsymbol{\theta})\odot\mathbf{X}(\boldsymbol{\theta})+\sigma_w^4(\rho-2)^{-1}\mathbf{I}_{KP}=\sigma_w^4\frac{\rho-1}{\rho-2}\mathbf{I}_{KP}-2\sigma_w^6\mathbf{U}(\boldsymbol{\theta})+o\left(M^{-1}\right)\tag{7.89}$$

and $\mathbf{U}(\boldsymbol{\theta})$ is the matrix introduced in (7.88), that can be written as

$$\begin{aligned}\mathbf{U}(\boldsymbol{\theta}) &= M^{-1}\left[\mathbf{I}_{KP} \odot \left(\left(\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t\right)^{-1} \otimes \mathbf{I}_P\right)\right] + O\left(M^{-2}\right) \\ &= M^{-1} Dg\left[\left(\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t\right)^{-1}\right] \otimes \mathbf{I}_P + O\left(M^{-2}\right)\end{aligned}$$

being $Dg[\mathbf{A}]$ the diagonal matrix built from the diagonal of $\mathbf{A}$.

Therefore, if the fourth- to second-order ratio ρ is not unitary, $\Gamma(\mathbf{K})$ is given by

$$\begin{aligned}[\Gamma(\mathbf{K})]_{p,q} &= \sigma_w^{-8} \frac{2-\rho}{\rho-1} \operatorname{diag}^H(\mathbf{Y}_p(\boldsymbol{\theta})) \operatorname{diag}(\mathbf{Y}_q(\boldsymbol{\theta})) \\ &= \frac{4K}{M^4} \frac{2-\rho}{\rho-1} \operatorname{Tr}\left(\mathcal{B}_p^{\infty}(\boldsymbol{\theta}) \mathcal{B}^{-1}(\boldsymbol{\theta}) \odot \mathcal{B}_q^{\infty}(\boldsymbol{\theta}) \mathcal{B}^{-1}(\boldsymbol{\theta})\right) + o\left(M^{-2}\right) \\ &= \begin{cases} \frac{K\pi^2}{4M^2} \frac{2-\rho}{\rho-1} \left(\sum\limits_{p \neq q} \frac{\sin(\pi M(\theta_p - \theta_q))}{\sin^2(\pi(\theta_p - \theta_q)/2)}\right)^2 + o\left(M^{-2}\right) & p = q \\ 0 & p \neq q \end{cases},\end{aligned}$$

repeating the calculations in (7.87). Notice that this term goes to zero as $O\left(M^{-2}\right)$ and, therefore, it is absolutely negligible when compared to $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ (7.84).

On the other hand, if we deal with a constant modulus alphabet with $\rho = 1$, the constant term in (7.89) is zero and the next term must be considered, yielding

$$\begin{aligned}[\Gamma(\mathbf{K})]_{p,q} &= 0.5\sigma_w^{-10} \operatorname{diag}^H(\mathbf{Y}_p(\boldsymbol{\theta})) \mathbf{U}^{-1}(\boldsymbol{\theta}) \operatorname{diag}(\mathbf{Y}_q(\boldsymbol{\theta})) \\ &= \frac{2\sigma_w^{-2} \xi K E_s}{M^3} \operatorname{Tr}\left(\mathcal{B}_p^{\infty}(\boldsymbol{\theta}) \mathcal{B}^{-1}(\boldsymbol{\theta}) \odot \mathcal{B}_q^{\infty}(\boldsymbol{\theta}) \mathcal{B}^{-1}(\boldsymbol{\theta})\right) + o\left(M^{-1}\right) \\ &= \begin{cases} \sigma_w^{-2} \frac{\xi K E_s \pi^2}{8M} \left(\sum\limits_{p \neq q} \frac{\sin(\pi M(\theta_p - \theta_q))}{\sin^2(\pi(\theta_p - \theta_q)/2)}\right)^2 + o\left(M^{-1}\right) & p = q \\ 0 & p \neq q \end{cases}\end{aligned} \tag{7.90}$$

where

$$E_s \triangleq \frac{1}{K} \operatorname{Tr}\left(\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t\right)$$

is the energy of the received symbols (for K sufficiently large) and

$$\xi \triangleq \frac{\operatorname{Tr}\left(Dg^{-1}\left[\left(\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t\right)^{-1}\right]\right)}{\operatorname{Tr}\left(\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t\right)} \leq 1 \tag{7.91}$$

is a coefficient determined by the snapshots correlation. In particular, ξ is unitary if the snapshots are uncorrelated because, in that case, $\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t = E_s \mathbf{I}_K$. Therefore, ξK can be understood as the effective observation time. The index ξ is therefore the only information about the temporal waveform that is retained in the asymptotic performance of the optimal second-order estimator.

Finally, regarding (7.90), we can state that the term $\Gamma(\mathbf{K})$ decays as $O\left(M^{-1}\right)$ and, therefore, it is asymptotically negligible in front of $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ (7.84).

Putting together all these partial results, if follows that the estimator performance is independent of the number of interfering users because the off-diagonal terms of $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$, $\mathbf{B}_{gml}(\boldsymbol{\theta})$ and $\mathbf{B}_{bque}(\boldsymbol{\theta})$ are negligible. Furthermore, the non-Gaussian information is negligible when the number of antennas goes to infinity because $\Gamma(\mathbf{K}) \ll \mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ and $\mathbf{X}_{gml}(\mathbf{K}) \ll \mathbf{B}_{UCRB}(\boldsymbol{\theta})$. Regarding the asymptotic value of $\Gamma(\mathbf{K})$, it can be seen how a positive term proportional to σ_w^{-2} appears when the nuisance parameters have a constant amplitude. However, this term is actually proportional to M^{-1} and, therefore, $\Gamma(\mathbf{K})$ is absolutely negligible in front of $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ (7.84) whatever the actual SNR.

Appendix 7.L Asymptotic study for $N_s \to \infty$

The asymptotic study considering an arbitrary temporal correlation $\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t$ and a finite number of sensors is rather involved because of the inverse appearing in $\mathbf{X}(\boldsymbol{\theta})$, $\mathbf{X}_p(\boldsymbol{\theta})$ and $\mathbf{X}_{p,q}(\boldsymbol{\theta})$ (7.35). To circumvent this obstacle, two directions have been adopted. In the first approach, the asymptotic study ($N_s \to \infty$) is carried out considering that the SNR goes to infinity without any assumption about $\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t$. The objective is to prove that the non-Gaussian term $\Gamma(\mathbf{K})$ (7.46) remains significant even if the observed time is infinite. An important conclusion is that, asymptotically, the estimator performance is independent of the temporal structure of the received signals, at least at high SNR. Bearing this result in mind, in the second part of this appendix, the same asymptotic study is done assuming that the received snapshots are uncorrelated, i.e., $\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t = E_s \mathbf{I}_K$. This scenario is actually the one simulated in Section 6.5 considering that the received symbols are detected without ISI at the matched filter output.

Large sample study for high SNR and arbitrary temporal correlation

To begin with, let us consider that the SNR is very high, i.e, $\sigma_w^2 \to 0$. In that case, the inverse in $\mathbf{X}(\boldsymbol{\theta})$, $\mathbf{X}_p(\boldsymbol{\theta})$ and $\mathbf{X}_{p,q}(\boldsymbol{\theta})$ (7.35) can be evaluated as we did when the number of antennas was infinite (7.80)-(7.82), obtaining

$$\left(\mathbf{B}(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}_{KP}\right)^{-1} = \mathbf{B}^{-1}(\boldsymbol{\theta}) - \sigma_w^2 \mathbf{B}^{-2}(\boldsymbol{\theta}) + \sigma_w^4 \mathbf{B}^{-3}(\boldsymbol{\theta}) + o\left(M^{-3}\right).$$

Then, plugging this result into (7.35), we get

$$\begin{aligned}
\mathbf{X}(\boldsymbol{\theta}) &= \sigma_w^2 \mathbf{I}_{KP} - \sigma_w^4 \mathbf{B}^{-1}(\boldsymbol{\theta}) + o\left(\sigma_w^4\right) \\
&= \sigma_w^2 \mathbf{I}_{KP} - \sigma_w^4 \left(\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t\right)^{-1} \otimes \mathcal{B}^{-1}(\boldsymbol{\theta}) + o\left(\sigma_w^4\right) \\
\mathbf{X}_p(\boldsymbol{\theta}) &= \sigma_w^2 \mathbf{B}_p(\boldsymbol{\theta}) \mathbf{B}^{-1}(\boldsymbol{\theta}) + o\left(\sigma_w^2\right) \\
&= \sigma_w^2 \mathbf{I}_K \otimes \mathcal{B}_p(\boldsymbol{\theta}) \mathcal{B}^{-1}(\boldsymbol{\theta}) + o\left(\sigma_w^2\right) \\
\mathbf{X}_{p,q}(\boldsymbol{\theta}) &= \mathbf{B}_{p,q}(\boldsymbol{\theta}) + \mathbf{B}_p(\boldsymbol{\theta}) \mathbf{B}^{-1}(\boldsymbol{\theta}) \mathbf{B}_q^H(\boldsymbol{\theta}) + o(1) \\
&= \mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t \otimes \left(\mathcal{B}_{p,q}(\boldsymbol{\theta}) - \mathcal{B}_p(\boldsymbol{\theta}) \mathcal{B}^{-1}(\boldsymbol{\theta}) \mathcal{B}_q^H(\boldsymbol{\theta})\right) + o(1)
\end{aligned}$$

where $\mathcal{B}(\boldsymbol{\theta})$, $\mathcal{B}_p(\boldsymbol{\theta})$ and $\mathcal{B}_{p,q}(\boldsymbol{\theta})$ are the spatial correlation matrices *for M finite* (7.51).

Based on the above *high-SNR* expressions, the (Gaussian) UCRB,

$$\left[\mathbf{B}_{UCRB}^{-1}\right]_{p,q} = 2\sigma_w^{-4} \operatorname{Re} \operatorname{Tr}\left(\mathbf{X}_p(\boldsymbol{\theta}) \mathbf{X}_q(\boldsymbol{\theta}) + \mathbf{X}(\boldsymbol{\theta}) \mathbf{X}_{p,q}(\boldsymbol{\theta})\right),$$

as well as the non-Gaussian terms $\Psi(\mathbf{K})$ and $\Gamma(\mathbf{K})$ introduced in (7.38) and (7.45),

$$\begin{aligned}
\left[\Psi(\mathbf{K})\right]_{p,q} &= \sigma_w^{-4} \operatorname{vec}^H\left(\mathbf{Y}_p(\boldsymbol{\theta})\right) \mathbf{V}_K \Sigma_K \mathbf{V}_K^H \operatorname{vec}\left(\mathbf{Y}_q(\boldsymbol{\theta})\right) \\
\left[\Gamma(\mathbf{K})\right]_{p,q} &= -\sigma_w^{-4} \operatorname{vec}^H\left(\mathbf{Y}_p(\boldsymbol{\theta})\right) \mathbf{V}_K \left(\mathbf{V}_K^H \left[\mathbf{X}^*(\boldsymbol{\theta}) \otimes \mathbf{X}(\boldsymbol{\theta})\right] \mathbf{V}_K + \sigma_w^4 \Sigma_K^{-1}\right)^{-1} \mathbf{V}_K^H \operatorname{vec}^H\left(\mathbf{Y}_p(\boldsymbol{\theta})\right),
\end{aligned}$$

can be evaluated when the number of received symbols goes to infinity ($N_s \to \infty$). In the last equations, $\mathbf{V}_K \Sigma_K \mathbf{V}_K^H$ is the "economy-size" diagonalization of the kurtosis matrix $\mathbf{K}$ and the high-SNR limit of $\mathbf{Y}_p(\boldsymbol{\theta})$ is given by

$$\begin{aligned}\mathbf{Y}_p(\boldsymbol{\theta}) &= \mathbf{X}(\boldsymbol{\theta})\mathbf{X}_P(\boldsymbol{\theta}) + \mathbf{X}_p^H(\boldsymbol{\theta})\mathbf{X}(\boldsymbol{\theta}) \\ &= \sigma_w^4\left(\mathbf{I}_K \otimes \mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta}) + \mathcal{B}^{-1}(\boldsymbol{\theta})\mathcal{B}_p^H(\boldsymbol{\theta})\right) + o\left(\sigma_w^4\right).\end{aligned}$$

At this point, the formulation in Appendix 7.K can be reproduced to produce asymptotic expressions for $N_s \to \infty$. Thus, in this appendix, similar asymptotic expressions to those in Appendix 7.K are deduced for $\mathcal{B}(\boldsymbol{\theta})$, $\mathcal{B}_p(\boldsymbol{\theta})$ and $\mathcal{B}_{p,q}(\boldsymbol{\theta})$ the spatial correlation matrices of the studied *finite* sensor array (7.51).

In Appendix 7.K, the asymptotic form of $\Psi(\mathbf{K})$ and $\Gamma(\mathbf{K})$ is derived as a function of $\mathbf{K}$ and, afterwards, the obtained expressions are simplified in case of having circular nuisance parameters. Next, assuming again *circular nuisance parameters*, the limit of $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$, $\mathbf{B}_{gml}(\boldsymbol{\theta})$ and $\mathbf{B}_{bque}(\boldsymbol{\theta})$ is calculated as the number of received symbols N_s goes to infinity[13]. Thus, starting from the above high-SNR limits of $\mathbf{X}(\boldsymbol{\theta})$, $\mathbf{X}_p(\boldsymbol{\theta})$, $\mathbf{X}_{p,q}(\boldsymbol{\theta})$ and $\mathbf{Y}_p(\boldsymbol{\theta})$, we arrive at

$$\left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q} = 2N_s\frac{E_s}{\sigma_w^2}\,\mathrm{Re}\,\mathrm{Tr}\left(\mathcal{B}_{p,q}(\boldsymbol{\theta}) - \mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta})\mathcal{B}_q^H(\boldsymbol{\theta})\right) + o(N_s)$$

$$\left[\Psi(\mathbf{K})\right]_{p,q} = \begin{cases} 4N_s(\rho-2)\,\mathrm{Tr}\left(\mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta}) \odot \mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta})\right) + o(N_s) & p = q \\ 0 & p \neq q \end{cases}$$

$$\left[\Gamma(\mathbf{K})\right]_{p,q} = \begin{cases} 4N_s\frac{2-\rho}{\rho-1}\,\mathrm{Tr}\left(\mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta}) \odot \mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta})\right) + o(N_s) & \rho \neq 1, p = q \\ 2\xi N_s\frac{E_s}{\sigma_w^2}\,\mathrm{Tr}\left(\mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta}) \odot Dg^{-1}\left[\mathcal{B}^{-1}(\boldsymbol{\theta})\right] \odot \mathcal{B}_p(\boldsymbol{\theta})\mathcal{B}^{-1}(\boldsymbol{\theta})\right) + o(N_s) & \rho = 1, p = q \\ 0 & p \neq q \end{cases}$$

taking into account that, if the number of received symbols N_s goes to infinity, the central rows and columns of $\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t$ are delayed versions of the pulse autocorrelation $R[k]$ (Section 7.4.4) and, therefore,

$$\lim_{N_s\to\infty}\frac{1}{N_s}\,\mathrm{Tr}\left(\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t\right) = R[0] = E_s$$

Besides, the coefficient ξ (7.91) can be manipulated using the spectral analysis in Section 7.4.4, yielding

$$\lim_{N_s\to\infty}\xi = \lim_{N_s\to\infty}\frac{\mathrm{Tr}\left(Dg^{-1}\left[\left(\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t\right)^{-1}\right]\right)}{\mathrm{Tr}\left(\mathbf{A}_t^H\mathbf{N}_t^{-1}\mathbf{A}_t\right)} = \frac{1}{\int_0^1 E_s/S(f)\,df}$$

[13]Notice that, if N_s goes to infinity,the number of observed symbols $K = N_s + L - 1$ is asymptotically equal to N_s.

with $S(f) = \mathfrak{F}\{R[k]\}$.

Following the same reasoning in Appendix 7.K, the asymptotic expression of $[\Gamma(\mathbf{K})]_{p,q}$ for $\rho = 1$ is obtained from (7.46) by expanding the argument of the inverse as follows:

$$\mathbf{X}^*(\boldsymbol{\theta}) \odot \mathbf{X}(\boldsymbol{\theta}) + \sigma_w^4 (\rho - 2)^{-1} \mathbf{I}_{KP} = \sigma_w^4 \frac{\rho - 1}{\rho - 2} \mathbf{I}_{KP} - 2\sigma_w^6 \mathbf{U}(\boldsymbol{\theta}) + o\left(\sigma_w^6\right)$$

where

$$\mathbf{U}(\boldsymbol{\theta}) = Dg\left[\left(\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t\right)^{-1}\right] \otimes Dg\left[\mathcal{B}^{-1}(\boldsymbol{\theta})\right]$$

is the surviving term when $\rho = 1$.

Notice that the asymptotic results in this appendix are equivalent to those obtained in the high-SNR study of Section 7.3.3 if we deal with circular nuisance parameters. The first conclusion is that $\mathbf{X}_{gml}(\boldsymbol{\theta}) = \mathbf{B}_{UCRB}(\boldsymbol{\theta}) \Psi(\mathbf{K}) \mathbf{B}_{UCRB}(\boldsymbol{\theta})$ (7.35) is negligible at high SNR because it is proportional to σ_w^4 whereas $\mathbf{B}_{UCRB}(\boldsymbol{\theta})$ is only proportional to σ_w^2. The second conclusion is that the second term $\Gamma(\mathbf{K})$ has the same dependence on N_s and σ_w^{-2} than $\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})$ in case of a constant-modulus alphabet and, therefore, $\Gamma(\mathbf{K})$ is *not* negligible even if $N_s \to \infty$.

However, the last conclusion is only verified in the multiuser case, i.e., $P > 1$. In the single user case, $\mathcal{B}(\boldsymbol{\theta})$, $\mathcal{B}_p(\boldsymbol{\theta})$ and $\mathcal{B}_{p,q}(\boldsymbol{\theta})$ are the following scalars:

$$\begin{aligned} \mathcal{B}(\boldsymbol{\theta}) &= M \\ \mathcal{B}_p(\boldsymbol{\theta}) &= 0 \\ \mathcal{B}_{p,q}(\boldsymbol{\theta}) &= \frac{\pi^2 \left(M^2 - 1\right) M}{12}. \end{aligned}$$

and, therefore, the non-Gaussian terms $\Psi(\mathbf{K})$ and $\Gamma(\mathbf{K})$ are zero *for any SNR* because $\mathcal{B}_p(\boldsymbol{\theta}) = 0$. Thus, in the single user case, the asymptotic performance of second-order bearing estimators is given by

$$\begin{aligned} \mathbf{B}_{UCRB}(\boldsymbol{\theta}), \mathbf{B}_{bque}(\boldsymbol{\theta}), \mathbf{B}_{gml}(\boldsymbol{\theta}) &= \frac{6}{\pi^2} \frac{\sigma_w^2}{N_s E_s} \frac{M + \sigma_w^2}{(M^2 - 1) M^2} + o\left(N_s^{-1}\right) \\ &= \frac{6}{\pi^2 N_s E_s / N_0} \frac{M + \frac{1}{E_s/N_0}}{(M^2 - 1) M^2} + o\left(N_s^{-1}\right) \end{aligned}$$

where $\sigma_w^2 = N_0$ is the double-sided spectral density of the AWG noise.

The above result is valid for any value of σ_w^2 or M. Moreover, this expression converges to the bound in (7.52) when the number of antennas holds that

$$M \gg \max\left\{(E_s/N_0)^{-1}, 1\right\},$$

which is equivalent to $M \gg 1$ in the context of digital communications.

Large sample study for uncorrelated snapshots and arbitrary SNR.

Next, the performance of the GML and BQUE estimators is evaluated considering an arbitrary SNR and uncorrelated snapshots, i.e., $\mathbf{A}_t^H \mathbf{N}_t^{-1} \mathbf{A}_t = E_s \mathbf{I}_K$. In this scenario, it is straightforward to show that the (Gaussian) UCRB (7.36) is inversely proportional to the number of snapshots $K = N_s$, *even if* N_s *is finite*. Actually, we have

$$\begin{aligned}\left[\mathbf{B}_{UCRB}^{-1}(\boldsymbol{\theta})\right]_{p,q} &= 2N_s E_s \sigma_w^{-4} \operatorname{Re}\operatorname{Tr}\left(\mathcal{X}_p(\boldsymbol{\theta})\,\mathcal{X}_q(\boldsymbol{\theta}) + \mathcal{X}(\boldsymbol{\theta})\,\mathcal{X}_{p,q}(\boldsymbol{\theta})\right) \\ &= 2N_s E_s \sigma_w^{-4}\left(\left[\mathcal{X}(\boldsymbol{\theta})\right]_{q,p}\left[\mathcal{X}_{p,q}(\boldsymbol{\theta})\right]_{p,q} + \left[\mathcal{X}_p(\boldsymbol{\theta})\right]_{p,q}\left[\mathcal{X}_q(\boldsymbol{\theta})\right]_{q,p}\right)\end{aligned}$$

where $\mathcal{X}(\boldsymbol{\theta})$, $\mathcal{X}_p(\boldsymbol{\theta})$ and $\mathcal{X}_{p,q}(\boldsymbol{\theta})$ are the spatial components of $\mathbf{X}(\boldsymbol{\theta})$, $\mathbf{X}_p(\boldsymbol{\theta})$ and $\mathbf{X}_{p,q}(\boldsymbol{\theta})$, that is,

$$\begin{aligned}\mathcal{X}(\boldsymbol{\theta}) &\triangleq \mathcal{B}(\boldsymbol{\theta}) - \mathcal{B}(\boldsymbol{\theta})\left(\mathcal{B}(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}_P\right)^{-1}\mathcal{B}(\boldsymbol{\theta}) \\ \mathcal{X}_p(\boldsymbol{\theta}) &\triangleq \mathcal{B}_p(\boldsymbol{\theta}) - \mathcal{B}_p(\boldsymbol{\theta})\left(\mathcal{B}(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}_P\right)^{-1}\mathcal{B}(\boldsymbol{\theta}) \\ \mathcal{X}_{p,q}(\boldsymbol{\theta}) &\triangleq \mathcal{B}_{p,q}(\boldsymbol{\theta}) - \mathcal{B}_p(\boldsymbol{\theta})\left(\mathcal{B}(\boldsymbol{\theta}) + \sigma_w^2 \mathbf{I}_P\right)^{-1}\mathcal{B}_q^H(\boldsymbol{\theta}).\end{aligned}$$

Furthermore, the non-Gaussian terms $\Psi(\mathbf{K})$ (7.42) and $\Gamma(\mathbf{K})$ (7.46) are also proportional to N_s and, consequently, they do not vanish as more snapshots are processed. Simple manipulations yield the following expressions in case of circular nuisance parameters:

$$\begin{aligned}\left[\Psi(\mathbf{K})\right]_{p,q} &= 4N_s E_s \sigma_w^{-4}(\rho - 2)\operatorname{Tr}\left(\mathcal{X}(\boldsymbol{\theta})\,\mathcal{X}_p(\boldsymbol{\theta}) \odot \mathcal{X}(\boldsymbol{\theta})\,\mathcal{X}_p(\boldsymbol{\theta})\right) \\ \left[\Gamma(\mathbf{K})\right]_{p,q} &= -4N_s E_s \sigma_w^{-4}\operatorname{diag}^H\left(\mathcal{X}(\boldsymbol{\theta})\,\mathcal{X}_p(\boldsymbol{\theta})\right)\left(\mathcal{X}^*(\boldsymbol{\theta}) \odot \mathcal{X}(\boldsymbol{\theta}) + \sigma_w^4(\rho - 2)^{-1}\mathbf{I}_P\right)^{-1} \\ &\quad \operatorname{diag}\left(\mathcal{X}(\boldsymbol{\theta})\,\mathcal{X}_q(\boldsymbol{\theta})\right).\end{aligned}$$

Finally, notice that $\mathcal{B}_p(\boldsymbol{\theta})$ is still null in the single user case and, therefore, $\Psi(\mathbf{K})$ and $\Gamma(\mathbf{K})$ are also zero because $\mathcal{X}_p(\boldsymbol{\theta}) = 0$.