Chapter 4

Optimal Second-Order Small-Error Estimation

In the last chapter, second-order estimators were designed by achieving a trade-off between bias and variance. The MMSE and minimum variance estimators were obtained averaging all the possible values of the parameter of interest. This approach has a few drawbacks that are summarized next. First of all, second-order estimators are usually *biased* even if the observation time is increased indefinitely. This fact precludes the existence of consistent quadratic estimators in a majority of nonlinear estimation problems. Moreover, the randomness of the nuisance parameters generally causes a serious variance floor at high SNR for *finite* data records and, therefore, self-noise free estimates are only possible asymptotically in case of infinite data samples.

In this chapter and the next one, the above problems are faced following two different but complementary approaches. In both cases, a *closed-loop or feed-back* scheme is adopted in which the estimator output is fed back in order to re-design the estimator coefficients and estimate once more the parameters of interest. The closed-loop implementation allows approaching succesively to the true parameter until the estimator attains –after convergence– the so-called *small-error regime* in which the estimator operates in the neighborhood of the true solution $\boldsymbol{\theta}_o$. Contrarily, the estimators studied in the previous chapter were based on an *open-loop or feedforward* architecture in which the parameter was extracted in a "single iterate" from the observed vector.

Based on this closed-loop architecture, two different approaches are considered in this chapter following the arguments in Section 2.5. On the one hand, the design of iterative methods is considered in which the observed vector $\mathbf{y}$ is repeatedly processed until attaining the small-error regime. With this aim, the gradient-based algorithms presented in Section 2.5 are implemented. The contribution of this chapter is the deduction of the *optimal second-order gradient*, and the corresponding Hessian, in case of arbitrarily distributed nuisance parameters. Throughout this chapter, we will assume that the length of the observed vector $\mathbf{y}$ is sufficient for exceeding the

SNR threshold and, thus, working in the small-error regime. Otherwise, the algorithm might converge to a spurious solution, usually referred to as *outlier* (Section 2.3.2).

On the other hand, the design of closed-loop estimators (Section 2.5.1) and trackers (Section 2.5.2) is also addressed in this chapter. As explained in Chapter 2, closed-loop estimators process the observation vector *sequentially*. The sequential implementation allows a significant reduction in terms of complexity and is unavoidable in case of dealing with a continuous transmission system in which the observation is infinite. It is shown in Section 2.5.1 that the closed-loop architecture yields efficient estimates if the observation is appropriately fragmented and all the parameters have been acquired correctly. Another important feature of closed-loop schemes is their capability of *tracking* the parameter evolution in time-variant scenarios as explained in Section 2.5.2.

As it was explained in Section 2.5.1, closed-loop estimators are composed of a discriminator and a loop filter. The discriminator is actually a *small-error* estimator dedicated to detect parameter deviations from the current estimate of $\boldsymbol{\theta}$. On the other hand, the loop filter is responsible for filtering the noisy estimates from the discriminator and predicting the parameter evolution in time-varying scenarios. The contribution of this chapter is the deduction of the *optimal second-order discriminator* assuming that the closed loop has attained the steady-state and, thus, it is working in the small-error regime. The actual distribution of the nuisance parameters is considered in order to cope with the self-noise in an optimal way.

The optimal second-order discriminator is obtained focusing uniquely on the steady-state performance and ignoring absolutely the acquisition and tracking behaviour. To complement this approach, the optimal second-order tracker is sought in Chapter 5 based on the Kalman filter theory. In that case, the discriminator and the loop filter are *jointly, adaptively* designed to optimize both the acquisition and steady-state performance.

To summarize this introduction, the small-error regime can be achieved by means of iterative or closed-loop algorithms. Once the small-error regime is achieved, second-order estimators are known to be *unbiased* since the estimator mean response $E\{\widehat{\boldsymbol{\alpha}}\}$ is approximately linear on the parameter $\boldsymbol{\alpha}$, irrespectively of the actual parameterization. Besides, in this small-error situation, second-order estimators become efficient for Gaussian nuisance parameters or in low-SNR scenarios. In the following sections, optimal second-order estimators are designed for the small-error regime and, afterwards, the resulting estimators are applied to the same estimation problem dealt with in Section 3.4; blind frequency offset estimation from digitally-modulated signals. More results can be found in Chapter 6 for the problems of NDA timing synchronization (Section 6.1), NDA carrier phase synchronization (Section 6.2), time-of-arrival estimation in multipath channels (Section 6.3), blind channel identification (Section 6.4) and, angle-of-arrival estimation (Section 6.5).

4.1 Small-Error Assumption

In the last chapter, the variability of $\boldsymbol{\theta}$ was considered by means of the prior $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. This chapter deals with the asymptotic case in which this variability is very small ($\boldsymbol{\theta} \simeq \boldsymbol{\theta}_o$). In this small-error regime, the prior $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is concentrated around the true parameter $\boldsymbol{\theta}_o$. Then, the formulation presented in the last chapter can be particularized for a very informative prior $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ holding that $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) < \varepsilon$ for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_o$ with ε arbitrarily small. Accordingly, the prior can be appropriately modelled as a Dirac's delta centered at $\boldsymbol{\theta} = \boldsymbol{\theta}_o$, that is, $f_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta} - \boldsymbol{\theta}_o)$.

Assuming that the estimator works in the small-error regime, the expected value of those complex matrices appearing in Section 3.2 and Section 3.3 can be approximated by means of their Taylor expansion at $\boldsymbol{\theta} \simeq \boldsymbol{\theta}_o$. Thus, if $\mathcal{F}(\boldsymbol{\theta})$ is a generic complex matrix depending on the vector of parameters $\boldsymbol{\theta}$, its mean value in the neighborhood of $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ can be approximated as follows:

$$E_{\boldsymbol{\theta}}\{\mathcal{F}(\boldsymbol{\theta})\} \simeq \mathcal{F}(\boldsymbol{\theta}_o) + \frac{1}{2}\sum_{p,q=1}^{P} \left.\frac{\partial^2 \mathcal{F}(\boldsymbol{\theta})}{\partial\theta_p \partial\theta_q}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} [\mathbf{C}_{\boldsymbol{\theta}}]_{p,q} \tag{4.1}$$

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

$$\mathbf{C}_{\boldsymbol{\theta}} \triangleq E_{\boldsymbol{\theta}}\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_o)(\boldsymbol{\theta} - \boldsymbol{\theta}_o)^H\right\}. \tag{4.2}$$

In Appendix 4.A, the vectors and matrices $\mathbf{r}$, $\mathbf{g}$, $\widetilde{\mathbf{Q}}$, $\mathbf{S}$ and $\mathbf{Q}$ (Section 3.1) are approximated in the small-error using (4.1), obtaining that

$$\mathbf{r} \simeq \mathbf{r}(\boldsymbol{\theta}_o) \triangleq \mathbf{r}_o \tag{4.3}$$

$$\mathbf{g} \simeq \mathbf{g}(\boldsymbol{\theta}_o) \tag{4.4}$$

$$\widetilde{\mathbf{Q}} \simeq \mathbf{D}_r \mathbf{C}_{\boldsymbol{\theta}} \mathbf{D}_r^H \tag{4.5}$$

$$\mathbf{S} \simeq \mathbf{D}_r \mathbf{C}_{\boldsymbol{\theta}} \mathbf{D}_g^H \tag{4.6}$$

$$\mathbf{Q} \simeq \mathbf{Q}(\boldsymbol{\theta}_o) \triangleq \mathbf{Q}_o \tag{4.7}$$

where

$$\mathbf{D}_r \triangleq \left.\frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \tag{4.8}$$

$$\mathbf{D}_g \triangleq \left.\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \tag{4.9}$$

Finally, under the small-error assumption, the prior is concentrated in $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ so that $\mathbf{C}_{\boldsymbol{\theta}}$ (4.2) collapses at this point becoming proportional to a given matrix $\mathbf{C}_{\boldsymbol{\theta}}^0$ defined as

$$\mathbf{C}_{\boldsymbol{\theta}}^0 \triangleq \lim_{\Delta \to 0} \frac{1}{\Delta} \mathbf{C}_{\boldsymbol{\theta}} \tag{4.10}$$

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

4.2 Second-Order Minimum Variance Estimator

In the small-error regime, the MMSE solution deduced in Section 3.2 makes no sense because it becomes dominated by the prior in such a way that $\widehat{\boldsymbol{\alpha}} = \mathbf{g}(\boldsymbol{\theta}_o)$ with $\mathbf{M}_{mse} = \mathbf{0}$. Thus, the MMSE solution must be constrained in some way to avoid the trivial solution. Once more, the minimum bias constraint is imposed to guarantee minimum bias around the true solution $\boldsymbol{\theta}_o$ and, then, the second-order minimum variance estimator in Section 3.3 is formulated again for the small-error regime. The important point is that the bias contribution can be totally eliminated in the small-error case (i.e., $\Delta \to 0$). Actually, a perfect matching between the estimator mean response

$$\overline{\boldsymbol{\alpha}}(\boldsymbol{\theta}) = \mathbf{g}(\boldsymbol{\theta}_o) + \mathbf{M}^H(\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r}_o)$$

and the target response $\mathbf{g}(\boldsymbol{\theta})$ is possible. The necessary and sufficient condition to have unbiased estimates ($BIAS^2 = 0$) is the equality of the derivatives of $\overline{\boldsymbol{\alpha}}(\boldsymbol{\theta})$ and $\mathbf{g}(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ (Appendix 4.B):

$$\mathbf{D}_r^H \mathbf{M} = \mathbf{D}_g^H \tag{4.11}$$

For the time being, the target response $\mathbf{g}(\boldsymbol{\theta})$ is supposed to verify the above equality for at least one matrix $\mathbf{M}$. Therefore, solving again the minimization problem in (3.29) under the constraints on $\mathbf{b}$ and $\mathbf{M}$ obtained in (3.16) and (4.11), the optimal *small-error* estimator is given by

$$\widehat{\boldsymbol{\alpha}} = \mathbf{g}(\boldsymbol{\theta}_o) + \mathbf{D}_g\left(\mathbf{D}_r^H\mathbf{Q}_o^{-1}\mathbf{D}_r\right)^{\#}\mathbf{D}_r^H\mathbf{Q}_o^{-1}(\widehat{\mathbf{r}} - \mathbf{r}_o) \tag{4.12}$$

where $\mathbf{r}_o$ and $\mathbf{Q}_o$ were defined in (4.3) and (4.7) and, the Moore-Penrose pseudoinverse is maintained to cover those cases in which $\mathbf{D}_r$ is singular. Thus, the estimation error covariance matrix is given by[1]

$$\mathbf{B}_{BQUE}(\boldsymbol{\theta}_o) \triangleq E\left\{(\widehat{\boldsymbol{\alpha}} - \mathbf{g}(\boldsymbol{\theta}_o))(\widehat{\boldsymbol{\alpha}} - \mathbf{g}(\boldsymbol{\theta}_o))^H\right\} = \mathbf{D}_g\left(\mathbf{D}_r^H\mathbf{Q}_o^{-1}\mathbf{D}_r\right)^{\#}\mathbf{D}_g^H \tag{4.13}$$

and the overall variance defined in (3.34) is calculated as the trace of $\mathbf{B}_{BQUE}(\boldsymbol{\theta}_o)$, i.e.,

$$VAR_{\min} = VAR(\boldsymbol{\theta}_o) = \mathrm{Tr}\left\{\mathbf{B}_{BQUE}(\boldsymbol{\theta}_o)\right\}.$$

Regarding the obtained solution, it is remarkable that the estimator covariance matrix in (4.13) has the same structure than the CRB in Section 2.6.1 where now

$$\mathbf{J}_2 \triangleq \mathbf{D}_r^H\mathbf{Q}_o^{-1}\mathbf{D}_r \tag{4.14}$$

[1]The following estimator was named in [Vil01a] the “Best Quadratic Unbiased Estimator” (BQUE) since it can be understood as a logical extension of the well-known “Best Linear Unbiased Estimator” (BLUE) [Kay93b, Ch.6] in case of dealing with a quadratic observation, i.e., $\widehat{\mathbf{r}} = vec\left(\widehat{\mathbf{R}}\right)$ (3.3).

plays the same role than the Fisher information matrix (FIM) for the family of *second-order* estimators considered in this dissertation. Therefore, (4.13) can be seen as the particularization of the Crámer-Rao bound to second-order estimation techniques. In section 2.6.1, the matrix $\mathbf{J}_2$ is shown to coincide with the FIM of the problem when the SNR is asymptotically low (Section 2.4.1) and/or the nuisance parameters are Gaussian (Section 2.4.3).

In general, it can be affirmed that

$$E\left\{(\widehat{\boldsymbol{\alpha}}-\mathbf{g}(\boldsymbol{\theta}_o))(\widehat{\boldsymbol{\alpha}}-\mathbf{g}(\boldsymbol{\theta}_o))^H\right\} \geq \mathbf{B}_{BQUE}(\boldsymbol{\theta}_o) \geq \mathbf{B}_{CRB}(\boldsymbol{\theta}_o) \quad \forall \boldsymbol{\theta}_o \tag{4.15}$$

for any unbiased estimator $\widehat{\boldsymbol{\alpha}}$ *based on the sample covariance matrix* $\widehat{\mathbf{R}} = \mathbf{y}\mathbf{y}^H$ where $\mathbf{B}_{CRB}(\boldsymbol{\theta})$ is the CRB of $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$ (Section 2.6.1). As stated before, the second inequality in (4.15) becomes an identity if the SNR tends to zero and/or the nuisance parameters are Gaussian random variables.

4.3 Second-Order Identifiability

This section is devoted to the analysis of the minimum-bias constraints introduced in (4.11). Using basic results on linear algebra, the solution of the system of equations in (4.11) offer three different possibilities [Mag98, Sec. 2.9], which are enumerated next:

1. $\mathbf{D}_r \in \mathbb{C}^{M^2 \times P}$ *is full column rank*. In that case, (4.11) is always consistent independently of the content of $\mathbf{D}_g \in \mathbb{R}^{Q \times P}$. Assuming that $\mathbf{D}_r$ is a tall matrix (i.e., $M^2 > P$), the solution of (4.11) is not unique since (4.11) becomes underdetermined. Actually, the bias minimization is only consuming QP degrees of freedom from $\mathbf{M} \in \mathbb{C}^{M^2 \times Q}$, whereas the remaining degrees of freedom, $\left(M^2 - P\right)Q$ are dedicated to minimize the estimator variance.

   In Appendix 4.B, it is shown that $\boldsymbol{\alpha} \simeq \mathbf{g}(\boldsymbol{\theta}_o) + \mathbf{D}_g(\boldsymbol{\theta} - \boldsymbol{\theta}_o)$ in the small-error regime with $\mathbf{D}_g = \mathbf{M}^H \mathbf{D}_r$ (4.11). This means that the rank of $\mathbf{M}^H \mathbf{D}_r$ determines the dimension of the subspace that contains the values of $\boldsymbol{\alpha} \in \mathbb{R}^Q$ that can be estimated in the small-error regime from the sample covariance matrix without any ambiguity. As the rank of $\mathbf{D}_r$ is P, the rank of $\mathbf{M}^H \mathbf{D}_r$ is equal to $min(P, Q)$ and, thus, $\boldsymbol{\alpha} \in \mathbb{R}^Q$ is locally *identifiable* from the sample covariance matrix assuming that $Q \leq P$.

2. $\mathbf{D}_r \in \mathbb{C}^{M^2 \times P}$ *is singular and* $\mathbf{D}_g^H \in span\left(\mathbf{D}_r^H\right)$. In that case, (4.11) is consistent if and only if $\mathbf{D}_g^H \in \mathbb{R}^{P \times Q}$ lies in the subspace generated by the rows of $\mathbf{D}_r$. Then, if $R < P$ is the column rank of $\mathbf{D}_r$, only QR constraints, out of the total QP constraints in (4.11), can be imposed.

In that case, the rank of $\mathbf{M}^H\mathbf{D}_r$ is the minimum of R and Q. Therefore, the parameter $\boldsymbol{\alpha} \in \mathbb{R}^Q$ is locally *identifiable* from the sample covariance matrix if and only if $\boldsymbol{\alpha}$ belongs to the subspace generated by $\mathbf{g}(\boldsymbol{\theta}_o) + \mathbf{D}_g(\boldsymbol{\theta} - \boldsymbol{\theta}_o)$, where the rank of $\mathbf{D}_g = \mathbf{M}^H\mathbf{D}_r$ is equal to $min(R, Q)$.

3. $\mathbf{D}_r$ *is singular but* $\mathbf{D}_g^H \notin$ *span* $\left(\mathbf{D}_r^H\right)$. In that case, (4.11) has no solution and, therefore, there is not exist any unbiased second-order estimator of $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$, even in the small-error regime. In [Sto01], Stoica and Marzetta proved that a finite variance estimator does not exist if (4.11) is not satisfied. Alternatively, the same conclusion can be drawn following the geometrical interpretation derived in [McW93].

In that case, the designer has to proceed as done in (3.29) to obtain the best approximation of $\mathbf{g}(\boldsymbol{\theta})$ holding the minimum bias constraints in (3.27). Thus, substituting (4.5)-(4.6) into (3.27), the minimum-bias constraints are given by

$$\mathbf{D}_r\mathbf{C}_{\boldsymbol{\theta}}^0\mathbf{D}_r^H\mathbf{M} = \mathbf{D}_r\mathbf{C}_{\boldsymbol{\theta}}^0\mathbf{D}_g^H \tag{4.16}$$

where the a priori covariance matrix $\mathbf{C}_{\boldsymbol{\theta}}^0$ (4.10) is used to carry out the matching proposed in (3.27). Otherwise, if $\mathbf{D}_r$ is full-rank, $\mathbf{C}_{\boldsymbol{\theta}}^0$ is not profitable, showing that Bayesian estimators cannot improve deterministic ones when the small-error assumption applies.

Focusing on the second case, there are two circumstances reducing the rank of $\mathbf{D}_r$:

- *The parameterization is not appropriate.* In the following three situations, the estimation problem is not correctly defined and $\mathbf{D}_r$ becomes singular. *Example 1*: the number of parameters Q is greater than the size of the sample covariance matrix M^2. *Example 2*: the Q parameters are not linearly independent and, therefore, the model is "overparameterized". *Example 3*: the sample covariance matrix, $\widehat{\mathbf{R}} = \mathbf{y}\mathbf{y}^H$, is insensitive to the phase of $\mathbf{y}$ in second-order estimation.[2]

- *The estimator has a finite resolution.* The estimator is unable to resolve two parameters of the same nature if they are very similar. For example, this problem arises in multiuser estimation problems as, for example, the problem of angle-of-arrival estimation in array signal processing (Section 6.5). It is worth noting that this situation, contrary to the ambiguities related before, cannot be predicted beforehand so it is not possible to guarantee (4.11) all the time. Therefore, the constraints in (4.16) should be used instead of those in (4.11) and the general estimator in (3.33) must be adopted using now the small-error matrices in (4.3)-(4.7).

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

However, from the designer viewpoint, the use of (4.16) may be problematic because the estimator would reduce automatically the rank of $\mathbf{D}_g^H = \mathbf{D}_r^H \mathbf{M}$ when entering into a singular situation (e.g., if two users cross each other as studied in Section 6.5), changing the value of $\mathbf{D}_g$. In the next section, this problem is overcome by setting free the value of the cross derivatives of (4.11).

4.4 Generalized Second-Order Constrained Estimators

Thus far, the estimator is designed to have an unbiased mean response when working under the small-error regime. Let us consider first that $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$ is a vector of Q independent parameters holding that $\mathbf{D}_g^H$ *is diagonal*. In that case, the diagonal entries of $\mathbf{D}_r^H \mathbf{M}$ are related to the estimator bias in the neighborhood of $\boldsymbol{\theta} = \boldsymbol{\theta}_o$ whereas the cross-terms reflect the coupling between parameters or, in other words, the *interparameter interference* (IPI). The classical unbiased solution forces $\mathbf{D}_r^H \mathbf{M} = \mathbf{D}_g^H$ (4.11) in order to yield unbiased estimates without IPI. However, strictly speaking, unbiased estimators are only required to constrain the value of the diagonal entries, that is,

$$\operatorname{diag}\left(\mathbf{D}_r^H \mathbf{M}\right) = \operatorname{diag}\left(\mathbf{D}_g^H\right), \tag{4.17}$$

since the IPI contribution is zero-mean in the small-error regime and, therefore, can only increase the estimator variance. Moreover, in noisy scenarios, the IPI-free condition usually causes noise-enhancement whereas, if the cross-terms in (4.11) are kept free, the estimator makes automatically a trade-off among noise, self-noise and IPI in order to minimize the overall variance.

Therefore, in case of independent parameters for which $\mathbf{D}_g$ is diagonal, the proposed second-order *unbiased* estimator is given by

$$\widehat{\boldsymbol{\alpha}} = \mathbf{g}(\boldsymbol{\theta}_o) + \mathbf{D}_g \operatorname{Dg}^{-1}(\mathbf{J}_2)\, \mathbf{D}_r^H \mathbf{Q}_o^{-1} (\widehat{\mathbf{r}} - \mathbf{r}_o) \tag{4.18}$$

where $\mathbf{J}_2 \triangleq \mathbf{D}_r^H \mathbf{Q}_o^{-1} \mathbf{D}_r$ is the second-order FIM introduced in (4.14).

However, the P parameters in $\boldsymbol{\theta}$ may appear coupled in $\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$. In that case, the matrix of derivatives $\mathbf{D}_g^H$ *is not diagonal* and the significance of the out-of-diagonal entries of $\mathbf{D}_g^H$ changes radically. Assuming that $\mathbf{D}_g^H$ is a full matrix (all the elements different from 0), any *unbiased* estimator of $\boldsymbol{\alpha}$ is required to fulfill (4.11) leading to the original small-error solution in (4.12). In general, if $\mathbf{D}_g^H$ is sparse, only the constraints in (4.11) corresponding to non-zero elements of $\mathbf{D}_g^H$ have to be imposed to obtain unbiased estimators of $\boldsymbol{\alpha}$.

In Section 6.5, the alternative solution obtained in (4.18) is evaluated and compared to the classical unbiased solution in (4.12) for the problem of tracking the angle-of-arrival of multiple digitally-modulated sources in the context of array signal processing.

4.5 A Case Study: Frequency Estimation

In this section, the small-error estimator proposed in (4.12) is simulated for the frequency estimation problem addressed in Section 3.4. Additional results are given in Chapter 6 for timing estimation and other relevant estimation problems. The results in this section show that the optimal second-order frequency estimator is unbiased and self-noise free. The Gaussian assumption is examined when the transmitted signal is digitally modulated showing that it is generally appropriate in the studied uniparametric problem. In addition, some singular cases are identified for the CPM modulation in which the Gaussian assumption is not able to cancel out the self-noise at high SNR. Later, in Section 6.5, the interest of including the digital information about the symbols is emphasized for the related problem of bearing estimation of multiple digitally-modulated sources.

Based on the signal model introduced in Section 3.4 for the problem of frequency estimation, the matrix of derivatives $\mathbf{D}_r$ is simply the following column vector:

$$\mathbf{d}_r \triangleq \mathrm{vec}\left(\left. \frac{\partial \mathbf{R}(\nu)}{\partial \nu} \right|_{\nu=\nu_o} \right) = \mathrm{vec}\left(\left. \frac{\partial \mathbf{E}(\nu/N_{ss})}{\partial \nu} \right|_{\nu=\nu_o} \odot \mathbf{A}\mathbf{A}^H \right)$$

where ν_o is the actual value of the parameter, and $[\mathbf{E}(\lambda)]_{i,k} = \exp\left(j2\pi\lambda\left(i-k\right)\right)$ are the elements of the Toeplitz matrix introduced in Appendix 3.D. The derivative of $\mathbf{E}(\nu/N_{ss})$ is then calculated, obtaining

$$\frac{\partial \left[\mathbf{E}(\nu/N_{ss})\right]_{i,k}}{\partial \nu} = j2\pi \frac{i-k}{N_{ss}} \left[\mathbf{E}(\nu/N_{ss})\right]_{i,k}.$$

Therefore, the optimal second-order small-error estimator is given by

$$\widehat{\nu} = \nu_o + \frac{\mathbf{d}_r^H \mathbf{Q}_o^{-1}}{\mathbf{d}_r^H \mathbf{Q}_o^{-1} \mathbf{d}_r} \left(\widehat{\mathbf{r}} - \mathbf{r}_o\right)$$

where $\mathbf{r}_o$ and $\mathbf{Q}_o$ were defined in (4.3) and (4.7), and the denominator is responsible for the unitary slope of $E\left\{\widehat{\nu}\right\}$.

Alternatively, a classical synchronization loop can be implemented in which the received signal is corrected using the estimated parameter $\widehat{\nu}$ (see Fig. 4.1). Thus, the discriminator can be designed assuming that the input parameter is $\nu_o = 0$ once the small-error regime is attained. Consequently, the optimal second-order discriminator is given by

$$\widehat{\nu} = \frac{\mathbf{d}_r^H \mathbf{Q}_o^{-1}}{\mathbf{d}_r^H \mathbf{Q}_o^{-1} \mathbf{d}_r} \widehat{\mathbf{r}}$$

where $\mathbf{d}_r$ and $\mathbf{Q}_o$ are computed at $\nu_o = 0$. Notice too that the last expression is simplified using that $\mathbf{d}_r^H \mathbf{Q}_o^{-1} \mathbf{r}_o = 0$. This condition is fulfilled thanks to the symmetry of matrix $\mathbf{A}\left(\nu\right)$ for the problem at hand.

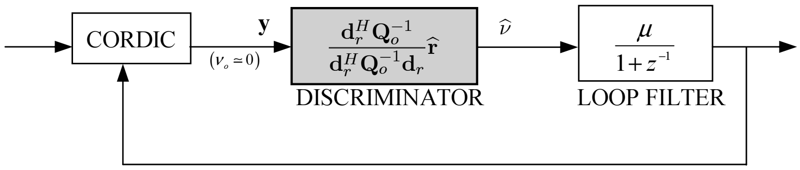

Figure 4.1: Block diagram of a (first-order) closed-loop frequency synchronizer. The optimal second-order discriminator, which was derived in this section under the small-error condition, is indicated in the figure. The CORDIC block is due to rotate the phase of the received signal according to the estimated frequency offset.

Whatever the selected scheme and the actual value of ν_o, the variance at the discriminator output is given by

$$VAR = E \left\| \widehat{\nu} - \nu_o \right\|^2 = \frac{1}{\mathbf{d}_r^H \mathbf{Q}_o^{-1} \mathbf{d}_r},$$

which constitutes the lower bound for the variance of any quadratic unbiased frequency error detector. If the nuisance parameters were normally distributed, then the above expression would correspond to the (Gaussian) UCRB bound presented in Section 2.6.1.

Notice that the discriminator variance could be reduced by including the usual loop filter (Fig. 4.1). In that case, the steady-state variance of the related closed-loop estimator is computed using the results in Section 6.1.4 following the reasoning in [Men97, Sec. 3.5.5].

The estimator performance is depicted in the following plots and compared with the ML-based estimators. The Gaussian assumption (GML) is shown to be *practically* optimal whatever the working point. Nonetheless, a minor degradation of about 0.9 dB is observed in Fig. 4.2 for positve E_s/N_0 in spite of increasing the observation time (Fig. 4.3). On the other hand, the low-SNR UML solution is rapidly limited by the self-noise as the SNR is augmented, manifesting a significant variance floor. This result is a consequence of the modulation intersymbol interference (ISI) and the finite observation time. In case of linear modulations, this high-SNR floor disappears (e.g., MPSK, QAM, etc.). Finally, the CML solution suffers from noise-enhancement at low SNR due also to the ISI.

The interest of the optimal small-error solution is more significant when dealing with a partial-response CPM modulatioon such as the LREC format [Men97, Sec. 4.2]. It can be seen that all the ML-based methods are dominated by the self-noise at high SNR (Figs. 4.4 and 4.5). The CML and GML solutions are not able to cancel out the self-noise when the number of nuisance parameters (K) is greater than the number of samples (M). In that case, the CML estimator cannot remove the self-noise term because there is no noise subspace where to project the data on. Moreover, as it will be studied in Appendix 7.E, the CML and GML solutions

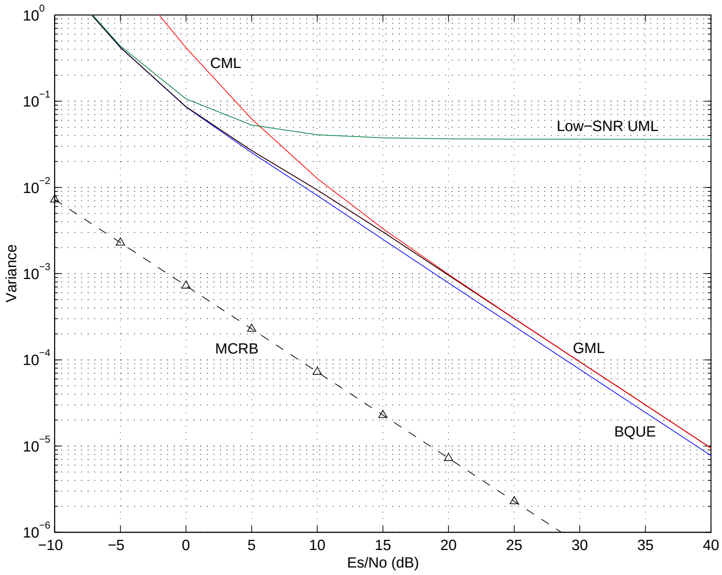

Figure 4.2: Frequency estimation variance under the small-error assumption for the optimal and GML estimators in case of MSK symbols, N_{ss}=2 and, M=4. The UCRB is not plotted for clarity since it is only slightly lower than the GML performance from Es/No=-5dB to Es/No=25dB.

are not equivalent at high SNR because the columns of $\mathbf{A}(\nu)$ are linearly dependent. Another simulation is run in which the received signal is oversampled ($N_{ss} = 4$) to guarantee that $M > K$ (Fig. 4.6). In that case, the GML and CML estimators supply self-noise free estimates at high SNR, although a significant loss is exhibited for practical SNRs.

On the other hand, the optimal second-error estimator is self-noise free under the small-error assumption, as shown for the 2REC and 3REC modulations in Fig. 4.4 and 4.5. Self-noise is removed by exploiting the pseudo-symbols fourth-order moments matrix $\mathbf{K}$. A detailed analysis on the asymptotic behaviour of second-order estimators at high SNR is given in Section 7.3.

Two classical small-error lower bounds for the variance of unbiased estimators are used to evaluate the performance of second-order techniques in the presence of nuisance parameters (Section 2.6.1). The (Gaussian) UCRB corresponds to the performance of the GML estimator in case of Gaussian nuisance parameters (Section 2.4.3). Although it has been extensively used in the literature as a valid bound in second-order estimation, simulations show how the UCRB is outperformed by the optimal second-order estimator when the nuisance parameters are discrete symbols. On the other hand, the MCRB predicts the ultimate performance of data-aided estimators that could be approached at high SNR by means of higher-order methods [Vil01b].

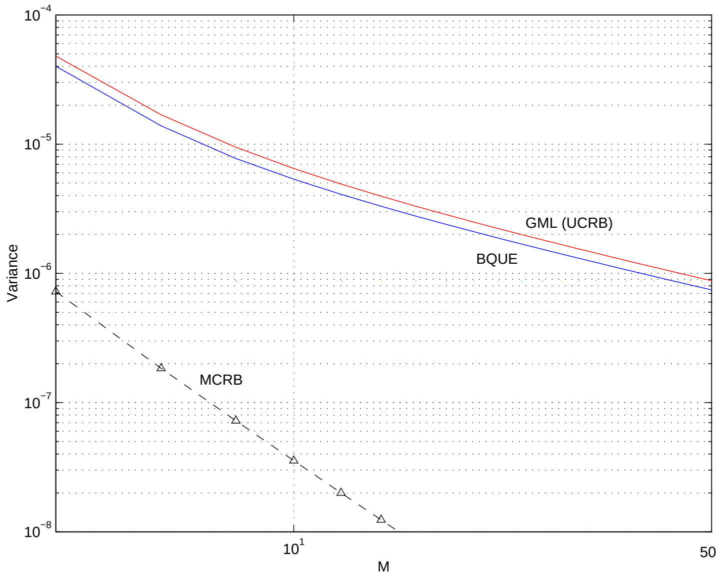

Figure 4.3: Frequency estimation variance under the small-error assumption as a function of M for the MSK modulation, Es/No=40dB and $N_{ss} = 2$.

It is worth noting that the performance predicted in the above curves is only realistic for high SNRs and/or a narrowband loop filter (Section 2.5.2). Otherwise, the studied closed-loop estimators are not able to achieve the small-error regime. This abnormal behaviour is not only associated to closed-loop schemes but it also appears in open-loop estimation in the form of *outliers* or *large-errors*.

Closed-loop schemes are sometimes able to acquire the parameter without external assistance. The necessary condition is that the estimator bias curve $E\left\{\widehat{\nu} - \nu_o\right\}$ –the so-called S-curve– uniquely intercepts the abcisa with positive slope at the origin. In Fig. 4.7, the acquisition stage of a first-order tracker with forgetting factor $\mu = 1/20$ is simulated for the 2REC modulation. The E_s/N_0 is set to 60dB in order to study the relevance of the self-noise term. Both the GML and the optimal second-order tracker are shown to acquire the parameter correctly with almost the same speed. On the other hand, the GML self-noise variance is apparent in the steady-state. The associated S-curves are also depicted in Figs. 4.8-4.10. It can be seen that all of them cross the origin and have unitary slope there.

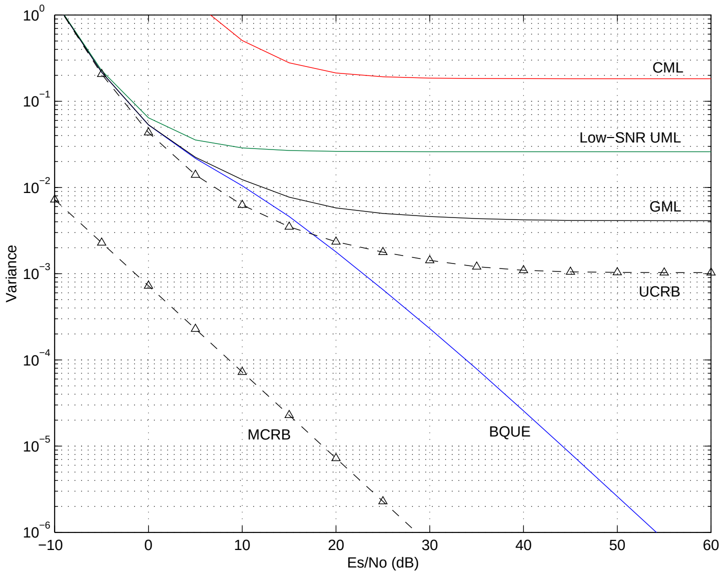

Figure 4.4: Estimators variance as a function of the Es/No for the 2REC modulation and M=8, N_{ss}=2. The number of pseudo-symbols is equal to K=12.

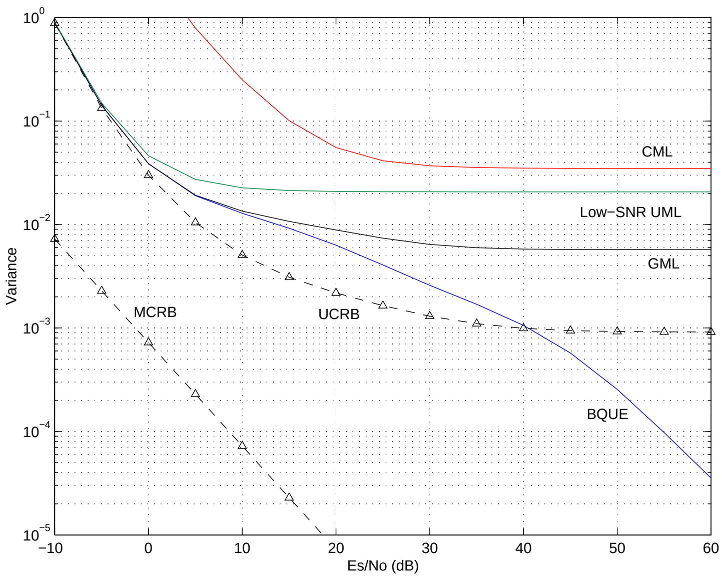

Figure 4.5: Estimators variance as a funciton of the Es/No for the 3REC modulation and M=8, N_{ss}=2. The number of pseudo-symbols is equal to K=28.

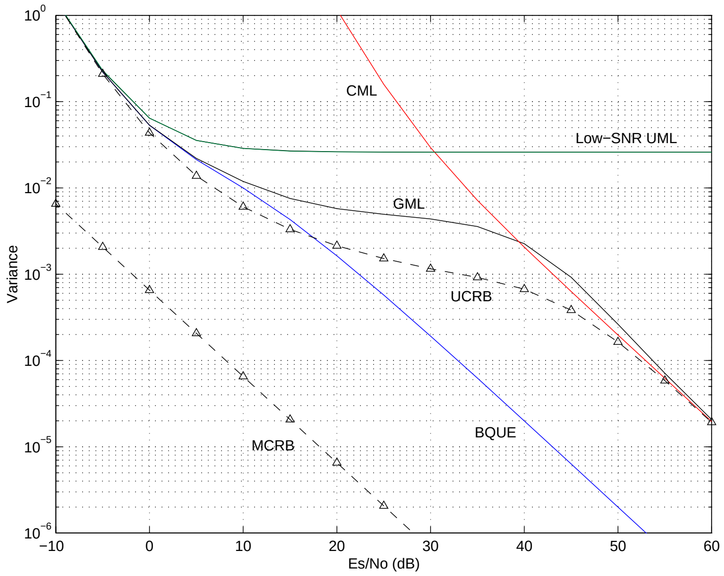

Figure 4.6: Estimators variance as a funciton of the Es/No for the 2REC modulation and M=4, N_{ss}=4. The number of pseudo-symbols is equal to K=12.

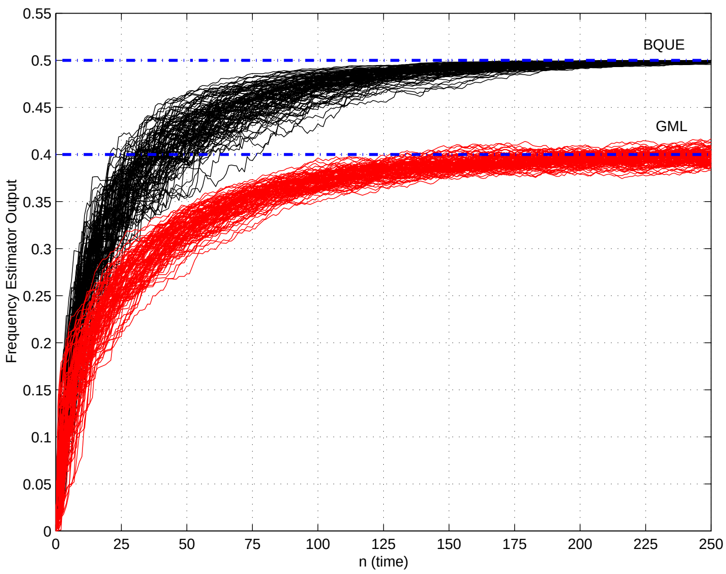

Figure 4.7: Frequency tracker output as a function of time for the 2REC modulation in a high SNR scenario (Es/No=60dB). The true frequency offset is equal to $\nu_o = 0.4$ (GML) and $\nu_o = 0.5$ (BQUE). Both trackers are initialized at $\nu = 0$ with M=8, N_{ss}=2. A first-order closed-loop is implemented with $\mu = 0.02$ the selected step-size or forgetting factor.

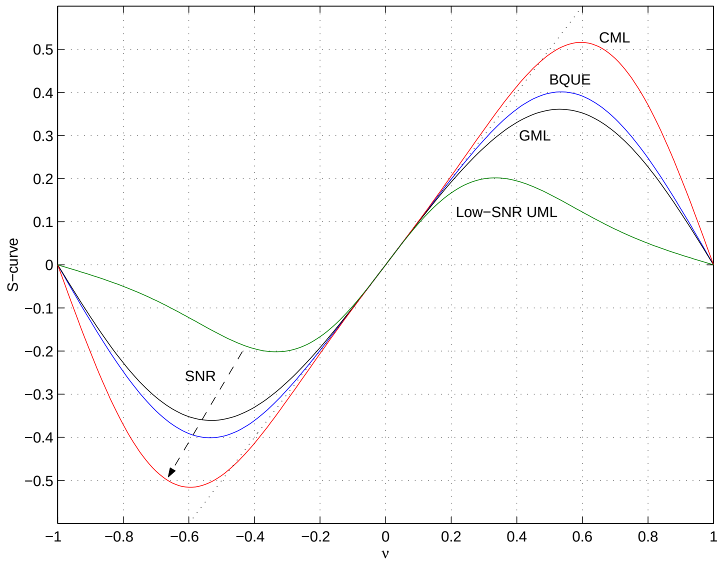

Figure 4.8: S-curve of the optimum and ML-based discriminators for the MSK modulation with M=8, N_{ss}=2, Es/No=10dB. The dashed arrow points out the tendency of the GML and BQUE S-curves as the Es/No is augmented from Es/No=0 (low-SNR UML S-curve) to Es/No=∞ (CML S-curve).

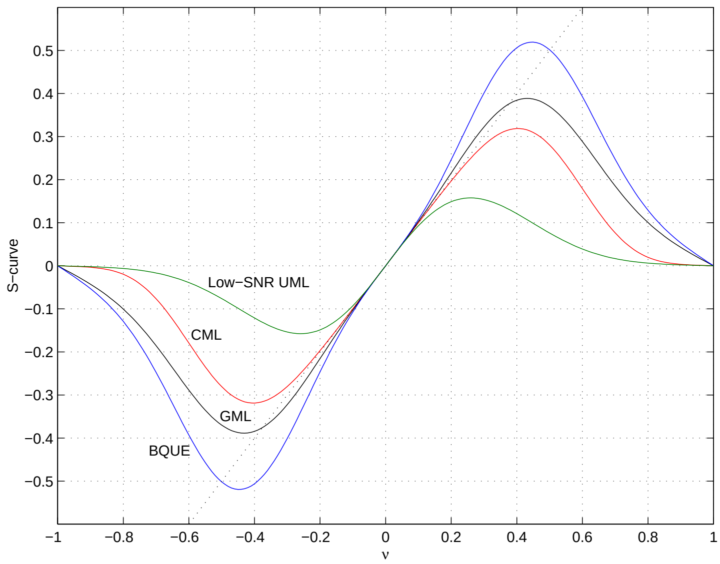

Figure 4.9: S-curve of the optimum and ML-based discriminators for the 2REC modulation with M=8, N_{ss}=2, Es/No=10dB.

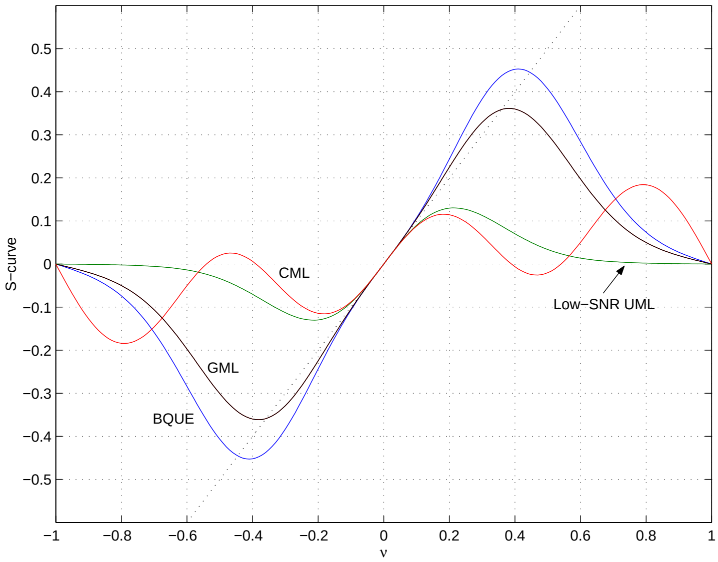

Figure 4.10: S-curve of the optimum and ML-based discriminators for the 3REC modulation with M=8, N_{ss}=2, Es/No=10dB.

4.6 Conclusions

The limitations of second-order feedforward methods in nonlinear estimation problems motivated the design of closed-loop estimators for the so-called small-error regime. Generally, second-order estimators are able to yield unbiased and self-noise free estimates once the small-error is attained after the acquisition. This important result is verified, whatever the considered estimation problem, if all the parameters are *(locally) identifiable.* Focusing only on those identifiable parameters, the prior distribution becomes irrelevant under the small-error assumption. Therefore, it can be stated that Bayesian estimators never outperform deterministic estimators in the small-error regime.

In this chapter, the *Best Quadratic Unbiased Estimator* (BQUE) is formulated considering the true distribution of the nuisance parameters. The BQUE expression is obtained analytically by expanding the constrained minimum variance solution in Chapter 3 in a Taylor's series in the neighbourhood of the true parameter where the small-error condition is satisfied. The resulting estimator is "the best" in the sense that it does not exist any other *unbiased second-order* estimator yielding a lower variance. Consequently, the BQUE performance constitutes the tigthest lower bound on the cvariance of any second-order unbiased blind estimator. Besides, it can be interpreted as the particularization of the CRB theory to second-order estimation.

The optimal second-order estimator is proved to depend on the fourth-order cumulants of the nuisance parameters. In some estimation problems, this fourth-order information becomes important to cope with the self-noise disturbance at high SNR. On the other hand, this information is omitted when the Gaussian assumption is adopted. In this chapter, the frequency estimation problem is studied concluding that the Gaussian assumption is practically optimal when we deal with a linear constellation. However, other simulations have shown that the non-Gaussian information about the nuisance parameters is needed to remove the self-noise at high SNR if the number of nuisance parameters exceeds the number of observations and a partial-response CPM transmission is considered. Some other illustrative examples will be studied in Chapter 6 in which the Gaussian assumption is questioned.

Finally, in the context of multiuser communications, the estimator peformance is seriously affected by the so-called multiple access inteference (MAI). The original BQUE solution is forced to eliminate the MAI contribution and, for this reason, it suffers from a significant noise enhancement in noisy scenarios. Thus, it is preferable to include the MAI term in the estimator optimization in order to make an optimal trade-off among the three disturbing random terms: thermal noise, self-noise and MAI. The obtained *MAI-resistant* BQUE estimator is further evaluated in Section 6.5 for the problem direction-of-arrival estimation in cellular communication systems.

Appendix 4.A Small-error matrices

Let us define $\mathbf{S}(\boldsymbol{\theta})$ and $\widetilde{\mathbf{Q}}(\boldsymbol{\theta})$ as the arguments inside the brackets of (3.24) and (3.23):

$$\mathbf{S}(\boldsymbol{\theta}) \triangleq (\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r})(\mathbf{g}(\boldsymbol{\theta}) - \mathbf{g})^H$$
$$\widetilde{\mathbf{Q}}(\boldsymbol{\theta}) \triangleq (\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r})(\mathbf{r}(\boldsymbol{\theta}) - \mathbf{r})^H .$$

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

$$\begin{aligned}
&\mathbf{S}(\boldsymbol{\theta}_o) = \mathbf{0}\\
&\left.\frac{\partial^2 \mathbf{S}(\boldsymbol{\theta})}{\partial\theta_p \partial\theta_q}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} = \left.\frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial\theta_p}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \left(\left.\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial\theta_q}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}\right)^H + \left.\frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial\theta_q}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} \left(\left.\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial\theta_p}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o}\right)^H\\
&\qquad = [\mathbf{D}_r]_p [\mathbf{D}_g]_q^H + [\mathbf{D}_r]_q [\mathbf{D}_g]_p^H ,
\end{aligned}$$

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

Then, equation (4.6) is obtained after plugging into (4.1) the following term:

$$\begin{aligned}
\sum_{p,q=1}^{P} \left.\frac{\partial^2 \mathbf{S}(\boldsymbol{\theta})}{\partial\theta_p \partial\theta_q}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} [\mathbf{C}_{\boldsymbol{\theta}}]_{p,q} &= \sum_{p,q=1}^{P} [\mathbf{D}_r]_p [\mathbf{D}_g]_q^H [\mathbf{C}_{\boldsymbol{\theta}}]_{p,q} + \sum_{p,q=1}^{P} [\mathbf{D}_r]_q [\mathbf{D}_g]_p^H [\mathbf{C}_{\boldsymbol{\theta}}]_{p,q}\\
&= \mathbf{D}_r \mathbf{C}_{\boldsymbol{\theta}} \mathbf{D}_g^H + \mathbf{D}_r \mathbf{C}_{\boldsymbol{\theta}}^T \mathbf{D}_g^H = 2\mathbf{D}_r \operatorname{Re}\{\mathbf{C}_{\boldsymbol{\theta}}\} \mathbf{D}_g^H . \qquad (4.19)
\end{aligned}$$

Proceeding in the same way with the matrix $\widetilde{\mathbf{Q}}(\boldsymbol{\theta})$, it is found that

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

Then, equation (4.5) is deduced after plugging into (4.1) the following expression:

$$\begin{aligned}
\sum_{p,q=1}^{P} \left.\frac{\partial^2 \widetilde{\mathbf{Q}}(\boldsymbol{\theta})}{\partial\theta_p \partial\theta_q}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} [\mathbf{C}_{\boldsymbol{\theta}}]_{p,q} &= \sum_{p,q=1}^{P} [\mathbf{D}_r]_p [\mathbf{D}_r]_q^H [\mathbf{C}_{\boldsymbol{\theta}}]_{p,q} + \sum_{p,q=1}^{P} [\mathbf{D}_r]_q [\mathbf{D}_r]_p^H [\mathbf{C}_{\boldsymbol{\theta}}]_{p,q}\\
&= \mathbf{D}_r \mathbf{C}_{\boldsymbol{\theta}} \mathbf{D}_r^H + \mathbf{D}_r \mathbf{C}_{\boldsymbol{\theta}}^T \mathbf{D}_r^H = 2\mathbf{D}_r \operatorname{Re}\{\mathbf{C}_{\boldsymbol{\theta}}\} \mathbf{D}_r^H . \qquad (4.20)
\end{aligned}$$

Finally, the real operator in (4.19) and (4.20) can be omitted taking into account that the vector of parameters is actually real-valued throughout this dissertation.

Appendix 4.B Proof of bias cancellation

If the Taylor expansion of $\overline{\boldsymbol{\alpha}}(\boldsymbol{\theta})$ and the target response $\mathbf{g}(\boldsymbol{\theta})$ are calculated around $\boldsymbol{\theta} = \boldsymbol{\theta}_o$, it is found that

$$\begin{aligned}\overline{\boldsymbol{\alpha}}(\boldsymbol{\theta}) &\simeq \mathbf{g}(\boldsymbol{\theta}_o) + \left.\frac{\partial \overline{\boldsymbol{\alpha}}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} (\boldsymbol{\theta} - \boldsymbol{\theta}_o) = \mathbf{g}(\boldsymbol{\theta}_o) + \mathbf{M}^H \mathbf{D}_r (\boldsymbol{\theta} - \boldsymbol{\theta}_o) \\ \mathbf{g}(\boldsymbol{\theta}) &\simeq \mathbf{g}(\boldsymbol{\theta}_o) + \left.\frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_o} (\boldsymbol{\theta} - \boldsymbol{\theta}_o) = \mathbf{g}(\boldsymbol{\theta}_o) + \mathbf{D}_g (\boldsymbol{\theta} - \boldsymbol{\theta}_o) \end{aligned} \tag{4.21}$$

with $\mathbf{D}_r$ and $\mathbf{D}_g$ defined in (4.8) and (4.9), respectively. Therefore, if (4.21) is plugged into (3.14), it follows that

$$BIAS^2 = E_{\boldsymbol{\theta}} \left\| \overline{\alpha}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\theta}) \right\|^2 = \mathrm{Tr}\left\{ \left(\mathbf{M}^H \mathbf{D}_r - \mathbf{D}_g \right) \mathbf{C}_{\boldsymbol{\theta}} \left(\mathbf{M}^H \mathbf{D}_r - \mathbf{D}_g \right)^H \right\}$$

where $\mathbf{C}_{\boldsymbol{\theta}}$ is the prior covariance matrix introduced in (4.2). Therefore, it follows that $\mathbf{M}^H \mathbf{D}_r = \mathbf{D}_g$ is a necessary and sufficient condition to ensure that $BIAS^2 = 0$ in the small-error regime if $\mathbf{C}_{\boldsymbol{\theta}}$ is a full-rank matrix.