Chapter 3

Optimal Second-Order Estimation

In this chapter, optimal second-order estimators are formulated considering that the estimator is provided with some side information about the unknown parameters. This side information can be exploited to improve the estimator accuracy or imposed by the designer in order to constrain the estimator mean response. Although the formulation in both cases will be the same, the classification above becomes crucial from a theoretical viewpoint. In that way, the adopted framework allows unifying the Bayesian and classical estimation theories.

In the first case, the Bayesian approach is adopted to model the unknown parameters as random variables of known probability density function $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. Thus, $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ provides the available statistical information on the parameters *prior to* the observation of the data. Bayesian estimators resort to this prior information when the observation is severely corrupted by the noise in low SNR scenarios. On the other hand, the prior contribution is scarse if the observation is rather informative. The above side information is supposed to be obtained in a previous estimation stage providing both the estimate and its accuracy. In that case, Gaussian priors are usually employed having in mind that the output of a consistent estimator becomes asymptotically Gaussian distributed on account of the Central Limit Theorem. When the parameter is constrained to a given interval, the folded Gaussian distribution is more appropiate [Rib97]. In particular, the folded Gaussian p.d.f. converges to the uniform distribution when all the available knowledge is the parameter range.

Bayesian estimation has received a lot of attention in the past decades but it has always raised a lot of controversy because the parameters are actually deterministic unknowns in a typical estimation problem (Section 2.1). The Bayesian approach is realistic if the parameters can be modelled as ergodic realizations of the a priori distribution $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. In this kind of applications, adaptive filters or *trackers* must be designed in order to track the parameter temporal evolution (Section 2.5.2). If the observation is linear in the parameters and the prior is Gaussian, the optimal linear tracker is the well-known Kalman filter [And79] [Kay93b]. Unfortunately, most estimation problems in communications are nonlinear and the suboptimal Extended Kalman filter (EKF) must be used instead. In Chapter 5, the EKF formulation is generalized to design *blind* second-order trackers based on the Bayesian interpretation of the results in this chapter.

If the classical estimation theory is adopted, the side information can be used to constrain the estimator mean response. In that case, $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ is a *weighting* function introduced by the designer to define new custom-built optimization criteria. Anyway, the formulation in both cases will be identical and $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ will be referred to as the prior distribution in spite of dealing with deterministic parameters. Likewise, $E_{\boldsymbol{\theta}}\{\cdot\}$ will denote indistintly the Bayesian expectation with respect to the random vector $\boldsymbol{\theta}$ or solely the following averaging

$$E_{\boldsymbol{\theta}}\{\mathcal{F}(\boldsymbol{\theta})\} \triangleq \int \mathcal{F}(\boldsymbol{\theta})\, f_{\boldsymbol{\theta}}(\boldsymbol{\theta})\, d\boldsymbol{\theta}, \tag{3.1}$$

if the parameters are deterministic amounts.

Based on the a priori distribution $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ and the known *linear* signal model introduced in Section 2.4, the objective is to find the optimal second-order estimator of $\boldsymbol{\alpha} \in \mathbb{R}^Q$ where

$$\boldsymbol{\alpha} = \mathbf{g}(\boldsymbol{\theta})$$

is an arbitrary transformation of $\boldsymbol{\theta} \in \mathbb{R}^P$. With this aim, the general expression of any second-order estimator of $\boldsymbol{\alpha}$ is presented next:

$$\widehat{\boldsymbol{\alpha}} = \mathbf{b} + \mathbf{M}^H \widehat{\mathbf{r}} \tag{3.2}$$

where

$$\widehat{\mathbf{r}} \triangleq \mathrm{vec}\left(\widehat{\mathbf{R}}\right) = \mathrm{vec}\left(\mathbf{y}\mathbf{y}^H\right) \tag{3.3}$$

is the column-wise vectorization of the sample covariance matrix $\widehat{\mathbf{R}}$ introduced in Section 2.4.1 and, $\mathbf{b}$ and $\mathbf{M}$ are the estimator coefficients corresponding to the independent and the quadratic term, respectively. Notice that the linear term is not considered because the nuisance parameters $\mathbf{x}$ are usually zero-mean random variables in the context of NDA estimation (2.4).

If the transmitted constellation is polarized or some training symbols are transmitted, the linear term $\mathbf{L}^H\mathbf{y}$ should be included following a semi-blind approach, improving so the estimator performance at low SNR [Mes02, Ch.3][Gor97][Car97]. Notice too that the $\mathrm{vec}(\cdot)$ operator can be applied successively to formulate higher-order estimators in order to improve the estimator performance in high SNR scenarios [Vil01b].

Finally, it is worth noting that a circular constellation is assumed. In that case, the *improper* covariance matrix $E\{\mathbf{y}\mathbf{y}^T\}$ is equal to zero and, therefore, no information can be drawn from the term $\mathrm{vec}(\mathbf{y}\mathbf{y}^T)$ where $(\cdot)^T$ stands for the transpose [Sch03][Pic96]. In Appendix 3.A, the results in this section are extended to encompass important noncircular constellations holding

$E\left\{\mathbf{y}\mathbf{y}^T\right\} \neq \mathbf{0}$ (e.g., PAM, BPSK or CPM). Moreover, the design of *quadratic* carrier phase synchronizers for the noncircular CPM modulation will be addressed in detail in Section 6.2.

Henceforth, the objective is to determine the estimator optimal coefficients $\mathbf{M}$ and $\mathbf{b}$ under a given performance criterion. Two criteria will be analyzed next; the first one is the usual minimum mean squared error (MMSE) criterion that minimizes the aggregated contribution of variance and bias. The MMSE criterion usually leads to biased estimators, mainly in low SNR scenarios in which the noise-induced variance is dominant. Unfortunately, in some applications bias is not tolerated (e.g., navigation applications) and some constraints must be introduced to compensate this bias. In that case, the proposed alternative is to minimize the estimator MSE subject to the minimum bias constraint. This chapter presents a convenient framework from which different estimation strategies can be devised as a trade-off between bias and variance. With this purpose, the following definitions are introduced in the next section.

3.1 Definitions and Notation

In this section, the mean square error (MSE) and variance figures are computed for the linear signal model introduced in Section 2.4 and for second-order estimation. Thus, the MSE associated to the generic second-order estimator in (3.2) is given by

$$MSE\left(\boldsymbol{\theta}\right) = E\left\|\widehat{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right) - \mathbf{g}\left(\boldsymbol{\theta}\right)\right\|^2 = E\left\|\mathbf{b} + \mathbf{M}^H\widehat{\mathbf{r}}\left(\boldsymbol{\theta}\right) - \mathbf{g}\left(\boldsymbol{\theta}\right)\right\|^2 \tag{3.4}$$

where the expectation is computed over the noise $\mathbf{w}$ and the nuisance parameters $\mathbf{x}$. The estimator MSE can be divided into the bias and variance contributions so that

$$MSE\left(\boldsymbol{\theta}\right) = BIAS^2\left(\boldsymbol{\theta}\right) + VAR\left(\boldsymbol{\theta}\right)$$

where the squared bias and variance are defined as follows:

$$BIAS^2\left(\boldsymbol{\theta}\right) = \left\|\overline{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right) - \mathbf{g}\left(\boldsymbol{\theta}\right)\right\|^2 = \left\|\mathbf{b} + \mathbf{M}^H\mathbf{r}\left(\boldsymbol{\theta}\right) - \mathbf{g}\left(\boldsymbol{\theta}\right)\right\|^2 \tag{3.5}$$

$$VAR\left(\boldsymbol{\theta}\right) = E\left\|\widehat{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right) - \overline{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right)\right\|^2 = E\left\|\mathbf{M}^H\left(\widehat{\mathbf{r}}\left(\boldsymbol{\theta}\right) - \mathbf{r}\left(\boldsymbol{\theta}\right)\right)\right\|^2 = \mathrm{Tr}\left\{\mathbf{M}^H\mathbf{Q}\left(\boldsymbol{\theta}\right)\mathbf{M}\right\} \tag{3.6}$$

with $\overline{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right)$ the estimator mean value

$$\overline{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right) \triangleq E\left\{\widehat{\boldsymbol{\alpha}}\left(\boldsymbol{\theta}\right)\right\} = \mathbf{b} + \mathbf{M}^H\mathbf{r}\left(\boldsymbol{\theta}\right) \tag{3.7}$$

and

$$\mathbf{r}\left(\boldsymbol{\theta}\right) \triangleq E\left\{\widehat{\mathbf{r}}\left(\boldsymbol{\theta}\right)\right\} = \mathrm{vec}\left(\mathbf{A}\left(\boldsymbol{\theta}\right)\mathbf{A}^H\left(\boldsymbol{\theta}\right) + \mathbf{R}_w\right) \tag{3.8}$$

$$\mathbf{Q}\left(\boldsymbol{\theta}\right) \triangleq E\left\{\left(\widehat{\mathbf{r}}\left(\boldsymbol{\theta}\right) - \mathbf{r}\left(\boldsymbol{\theta}\right)\right)\left(\widehat{\mathbf{r}}\left(\boldsymbol{\theta}\right) - \mathbf{r}\left(\boldsymbol{\theta}\right)\right)^H\right\} \tag{3.9}$$

the mean and the covariance matrix of the *vectorized* sample covariance matrix $\widehat{\mathbf{r}}$, respectively. Notice that $\mathbf{r}\left(\boldsymbol{\theta}\right)$ corresponds to the (vectorized) covariance matrix of $\mathbf{y}$ whereas $\mathbf{Q}\left(\boldsymbol{\theta}\right)$ gathers the central fourth-order moments of $\mathbf{y}$. The vectorization is fundamental to derive a closed form for the matrix $\mathbf{Q}(\boldsymbol{\theta})$. In Appendix 3.B, it is found that

$$\mathbf{Q}(\boldsymbol{\theta}) = \mathbf{R}^*(\boldsymbol{\theta}) \otimes \mathbf{R}(\boldsymbol{\theta}) + \mathcal{A}(\boldsymbol{\theta})\mathbf{K}\mathcal{A}^H(\boldsymbol{\theta}) \tag{3.10}$$

where $\mathcal{A}(\boldsymbol{\theta}) \triangleq \mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})$, $\mathbf{R}(\boldsymbol{\theta})$ was introduced in Section 2.4.3 and,

$$\mathbf{K} \triangleq E_{\mathbf{x}}\{\mathrm{vec}\left(\mathbf{x}\mathbf{x}^H\right)\mathrm{vec}^H\left(\mathbf{x}\mathbf{x}^H\right)\} - \mathrm{vec}\left(\mathbf{I}_K\right)\mathrm{vec}^H\left(\mathbf{I}_K\right) - \mathbf{I}_{K^2} \tag{3.11}$$

is the matrix containing the fourth-order cumulants (kurtosis) of the nuisance parameters $\mathbf{x}$. It is worth realizing that $\mathbf{Q}(\boldsymbol{\theta})$ and $\mathbf{K}$ are calculated analytically for the linear signal model introduced in Section 2.4, avoiding so the problematic estimation of fourth-order statistics.

In the case of zero-mean, *circular* complex nuisance parameters, the matrix $\mathbf{K}$ is the following diagonal matrix:

$$\mathbf{K} = (\rho - 2)\,\mathrm{diag}\left(\mathrm{vec}\left(\mathbf{I}_K\right)\right) \tag{3.12}$$

where the scalar

$$\rho \triangleq \frac{E\left\{\left|\left[\mathbf{x}\right]_k\right|^4\right\}}{E^2\{\left|\left[\mathbf{x}\right]_k\right|^2\}}$$

is the fourth- to second-order moment ratio (Appendix 3.C). If the nuisance parameters are not circular (e.g., for the CPM modulation), the expectation in (3.11) has to be computed numerically –and *offline*– from the known p.d.f. of $\mathbf{x}$. Moreover, if the nuisance parameters are discrete, as it happens in digital communications, the computation of $\mathbf{K}$ needs only a small number of realizations of $f_{\mathbf{x}}(\mathbf{x})$.

It is well-known that matrix $\mathbf{K}$ is zero for normally distributed nuisance parameters for which $\rho = 2$. Otherwise, matrix $\mathbf{K}$ provides the complete non-Gaussian information about the nuisance parameters that second-order estimators are able to exploit. In fact, the GML estimator is sometimes outperformed at high SNR if the second term of (3.10) is considered. This remark was actually the motivation of this thesis and will be analyzed intensively in the following chapters.

Unfortunately, the vector $\mathbf{b}$ and matrix $\mathbf{M}$ minimizing the bias, variance or MSE figures are generally a function of the *unknown* vector of parameters $\boldsymbol{\theta}$ and, therefore, the resulting estimator is not realizable. Accordingly, the estimator coefficients have to be optimized from a convenient average of these figures of merit over all the possible values of $\boldsymbol{\theta}$. In that sense, the prior $f_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ introduced previously is applied to obtain the following averaged MSE, bias and variance:

$$MSE \triangleq E_{\boldsymbol{\theta}}\left\{MSE(\boldsymbol{\theta})\right\} = BIAS^2 + VAR = E_{\boldsymbol{\theta}}E\left\|\mathbf{b} + \mathbf{M}^H\widehat{\mathbf{r}}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\theta})\right\|^2 \tag{3.13}$$

$$BIAS^2 \triangleq E_{\boldsymbol{\theta}}\left\{BIAS^2(\boldsymbol{\theta})\right\} = E_{\boldsymbol{\theta}}\left\|\mathbf{b} + \mathbf{M}^H\mathbf{r}(\boldsymbol{\theta}) - \mathbf{g}(\boldsymbol{\theta})\right\|^2 \tag{3.14}$$

$$VAR \triangleq E_{\boldsymbol{\theta}}\left\{VAR(\boldsymbol{\theta})\right\} = \mathrm{Tr}\left\{\mathbf{M}^H\mathbf{Q}\mathbf{M}\right\} \tag{3.15}$$

with

$$\mathbf{Q} \triangleq E_{\boldsymbol{\theta}} \left\{ \mathbf{Q}\left(\boldsymbol{\theta}\right) \right\}.$$

The estimator variance in (3.15) is independent of $\mathbf{b}$. Therefore, $\mathbf{b}$ can be selected to minimize the bias contribution without degrading the estimator variance. It is found that the optimum $\mathbf{b}$ is given by

$$\mathbf{b}_{opt} \triangleq \arg\min_{\mathbf{b}} BIAS^2 = \arg\min_{\mathbf{b}} MSE = \mathbf{g} - \mathbf{M}^H \mathbf{r} \tag{3.16}$$

where

$$\mathbf{g} \triangleq E_{\boldsymbol{\theta}} \left\{ \mathbf{g}\left(\boldsymbol{\theta}\right) \right\} \tag{3.17}$$

$$\mathbf{r} \triangleq E_{\boldsymbol{\theta}} E \left\{ \widehat{\mathbf{r}}\left(\boldsymbol{\theta}\right) \right\} \tag{3.18}$$

If now $\mathbf{b}_{opt}$ is substituted into (3.13) and (3.14), we obtain that

$$BIAS^2 = E_{\boldsymbol{\theta}} \left\| \mathbf{M}^H \left(\mathbf{r}\left(\boldsymbol{\theta}\right) - \mathbf{r} \right) - \left(\mathbf{g}\left(\boldsymbol{\theta}\right) - \mathbf{g} \right) \right\|^2 = \sigma_g^2 + \mathrm{Tr}\left\{ \mathbf{M}^H \widetilde{\mathbf{Q}} \mathbf{M} - \mathbf{M}^H \mathbf{S} - \mathbf{S}^H \mathbf{M} \right\} \tag{3.19}$$

$$VAR = \mathrm{Tr}\left\{ \mathbf{M}^H \mathbf{Q} \mathbf{M} \right\} \tag{3.20}$$

$$MSE = BIAS^2 + VAR = \sigma_g^2 + \mathrm{Tr}\left\{ \mathbf{M}^H \left(\mathbf{Q} + \widetilde{\mathbf{Q}} \right) \mathbf{M} - \mathbf{M}^H \mathbf{S} - \mathbf{S}^H \mathbf{M} \right\} \tag{3.21}$$

with the following definitions

$$\sigma_g^2 \triangleq E_{\boldsymbol{\theta}} \left\| \mathbf{g}\left(\boldsymbol{\theta}\right) - \mathbf{g} \right\|^2 \tag{3.22}$$

$$\widetilde{\mathbf{Q}} \triangleq E_{\boldsymbol{\theta}} \left\{ \left(\mathbf{r}\left(\boldsymbol{\theta}\right) - \mathbf{r} \right) \left(\mathbf{r}\left(\boldsymbol{\theta}\right) - \mathbf{r} \right)^H \right\} \tag{3.23}$$

$$\mathbf{S} \triangleq E_{\boldsymbol{\theta}} \left\{ \left(\mathbf{r}\left(\boldsymbol{\theta}\right) - \mathbf{r} \right) \left(\mathbf{g}\left(\boldsymbol{\theta}\right) - \mathbf{g} \right)^H \right\}. \tag{3.24}$$

The expectation with respect to the prior $f_{\boldsymbol{\theta}}\left(\boldsymbol{\theta}\right)$ poses serious problems when calculating the analytical expressions of $\mathbf{g}$, $\mathbf{r}$, $\mathbf{Q}$, $\widetilde{\mathbf{Q}}$ and $\mathbf{S}$. In Appendix 3.D, this problem is solved when the parameter dependence is phasorial. In the following sections, the MMSE estimator and the minimum bias-variance estimator are formulated, and further analyzed, assuming that these vectors and matrices have been computed somehow.

3.2 Second-Order MMSE Estimator

The second-order MMSE estimator is obtained minimizing the overall MSE in equation (3.21). It follows that the optimum matrix $\mathbf{M}$ is given by

$$\mathbf{M}_{mse} = \arg\min_{\mathbf{M}} MSE = \left(\mathbf{Q} + \widetilde{\mathbf{Q}} \right)^{-1} \mathbf{S} \tag{3.25}$$

where the inversion is guaranteed assuming that the noise covariance matrix $\mathbf{R}_w$ is positive definite. Notice that the above expression corresponds to the linear Bayesian MMSE estimator of $\boldsymbol{\alpha}$ based on the sample covariance vector $\widehat{\mathbf{r}}$ where $\mathbf{Q}+\widetilde{\mathbf{Q}}$ is the autocorrelation matrix of $\widehat{\mathbf{r}}$ and, $\mathbf{S}$ the cross-correlation between $\widehat{\mathbf{r}}$ and $\boldsymbol{\alpha}$ [Kay93b].

If (3.25) is now plugged into (3.21), the minimum MSE is found to be

$$MSE_{\min} = \sigma_g^2 - \mathrm{Tr}\left\{\mathbf{S}^H\left(\mathbf{Q}+\widetilde{\mathbf{Q}}\right)^{-1}\mathbf{S}\right\} \tag{3.26}$$

where σ_g^2 is the initial (prior) uncertainty about the parameter $\boldsymbol{\alpha}$ and the second term is the MSE improvement after processing the data vector $\mathbf{y}$. It is easy to show that this second term vanishes as the noise variance is increased.

3.3 Second-Order Minimum Variance Estimator

The aim of this section would be obtaining the minimum variance unbiased (MVU) estimator (Section 2.2). However, in most cases it is not possible to cancel out the bias term unless the covariance vector $\mathbf{r}(\boldsymbol{\theta})$ is an affine transformation of $\boldsymbol{\alpha}\in\mathbb{R}^Q$, that is, $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{W}\mathbf{g}(\boldsymbol{\theta}) + \mathbf{v}$ for any value of $\mathbf{W}$ and $\mathbf{v}$. If so, it is straightforward to verify that the estimator bias (3.14) is removed by setting $\mathbf{M}^H\mathbf{W} = \mathbf{I}_Q$. Unfortunately, this situation is unusual and quadratic estimators are normally degraded by some residual bias. Taking into account this limitation, in this section the minimum variance estimator is deduced subject to those constraints minimizing the estimator bias. Thus, let us first obtain the equation that $\mathbf{M}$ must verify to yield minimum bias:

$$\frac{dBIAS^2}{d\mathbf{M}^*} = \widetilde{\mathbf{Q}}\mathbf{M} - \mathbf{S} = \mathbf{0} \tag{3.27}$$

Generally, the constraints obtained in (3.27) form an underdetermined system of equations because $R \triangleq rank\left(\widetilde{\mathbf{Q}}\right) < M^2$ and $\mathbf{S}\in\mathbb{C}^{M^2\times Q}$ in (3.24) lies, by definition, in the column span of $\widetilde{\mathbf{Q}}\in\mathbb{C}^{M^2\times M^2}$. Hence, (3.27) is actually imposing RQ design constraints on the matrix $\mathbf{M}\in\mathbb{C}^{M^2\times Q}$ that, after the diagonalization of $\widetilde{\mathbf{Q}} = \mathbf{V}\Sigma\mathbf{V}^H$, can be formulated as follows:

$$\mathbf{V}^H\mathbf{M} = \mathcal{S} \tag{3.28}$$

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

Therefore, since equation (3.27) is only forcing R constraints, the remaining degrees of freedom in $\mathbf{M}$ can be used to optimize the estimator variance. Specifically, the aim is to minimize the estimator variance subject to the constraints on $\overline{\boldsymbol{\alpha}}(\boldsymbol{\theta})$ given in (3.27) or (3.28), that is,

$$\begin{aligned}&\mathbf{M}_{var} = \arg\min_{\mathbf{M}} VAR = \arg\min_{\mathbf{M}} \mathbf{M}^H\mathbf{Q}\mathbf{M}\\ &\text{subject to } \widetilde{\mathbf{Q}}\mathbf{M} = \mathbf{S} \text{ or } \mathbf{V}^H\mathbf{M} = \mathcal{S},\end{aligned} \tag{3.29}$$

which yields the following solution:

$$\mathbf{M}_{var} = \mathbf{P}^H \mathbf{S} = \mathcal{P}^H \mathcal{S} \tag{3.30}$$

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

$$\mathbf{P} \triangleq \left(\widetilde{\mathbf{Q}}\mathbf{Q}^{-1}\widetilde{\mathbf{Q}}\right)^{\#} \widetilde{\mathbf{Q}}\mathbf{Q}^{-1} \tag{3.31}$$

$$\mathcal{P} \triangleq \left(\mathbf{V}^H \mathbf{Q}^{-1} \mathbf{V}\right)^{-1} \mathbf{V}^H \mathbf{Q}^{-1}. \tag{3.32}$$

Thus, after substitutions in (3.2), the minimum variance estimator is given by

$$\widehat{\boldsymbol{\alpha}} = \mathbf{g} + \mathbf{S}^H \mathbf{P}\left(\widehat{\mathbf{r}} - \mathbf{r}\right) = \mathbf{g} + \mathcal{S}^H \mathcal{P}\left(\widehat{\mathbf{r}} - \mathbf{r}\right) \tag{3.33}$$

where $\mathbf{P}$ is projecting the sample covariance vector $(\widehat{\mathbf{r}} - \mathbf{r})$ onto the minimum-bias subspace generated by matrix $\widetilde{\mathbf{Q}}$ in (3.27) (see Fig. 3.1). Plugging now (3.30) into (3.20), the minimum variance is equal to

$$VAR_{\min} = \mathrm{Tr}\left\{\mathbf{S}^H \left(\widetilde{\mathbf{Q}}\mathbf{Q}^{-1}\widetilde{\mathbf{Q}}\right)^{\#} \mathbf{S}\right\} = \mathrm{Tr}\left\{\mathcal{S}^H \left(\mathbf{V}^H \mathbf{Q}^{-1}\mathbf{V}\right)^{-1} \mathcal{S}\right\} \tag{3.34}$$

where the argument inside the trace operator is the covariance matrix of the estimation error:

$$\begin{aligned} E_{\boldsymbol{\theta}} E\left\{\left(\widehat{\boldsymbol{\alpha}}(\boldsymbol{\theta}) - \overline{\boldsymbol{\alpha}}(\boldsymbol{\theta})\right)\left(\widehat{\boldsymbol{\alpha}}(\boldsymbol{\theta}) - \overline{\boldsymbol{\alpha}}(\boldsymbol{\theta})\right)^H\right\} &= \mathbf{S}^H \left(\widetilde{\mathbf{Q}}\mathbf{Q}^{-1}\widetilde{\mathbf{Q}}\right)^{\#} \mathbf{S} \\ &= \mathcal{S}^H \left(\mathbf{V}^H \mathbf{Q}^{-1}\mathbf{V}\right)^{-1} \mathcal{S} \end{aligned} \tag{3.35}$$

Finally, plugging (3.30) into (3.19), the residual bias can be expressed in any of these alternative forms:

$$\begin{aligned} BIAS^2_{\min} &= \sigma_g^2 - \mathrm{Tr}\left\{\mathbf{M}^H \mathbf{S}\right\} = \sigma_g^2 - \mathrm{Tr}\left\{\mathbf{M}^H \widetilde{\mathbf{Q}}\mathbf{M}\right\} \\ &= \sigma_g^2 - \mathrm{Tr}\left\{\mathbf{S}^H \mathbf{P}\mathbf{S}\right\} = \sigma_g^2 - \mathrm{Tr}\left\{\mathcal{S}^H \Sigma \mathcal{S}\right\} \\ &= \sigma_g^2 - \mathrm{Tr}\left\{\mathbf{S}^H \widetilde{\mathbf{Q}}^{\#} \mathbf{S}\right\} \end{aligned} \tag{3.36}$$

The last equation is obtained from $\mathbf{M}^H \widetilde{\mathbf{Q}}\mathbf{M}$ using that[1]

$$\mathbf{P}\widetilde{\mathbf{Q}} = \widetilde{\mathbf{Q}}^{\#}\widetilde{\mathbf{Q}}$$

[1]The following identity is obtained from the diagonalization of $\widetilde{\mathbf{Q}} = \mathbf{V}\Sigma\mathbf{V}^H$. It is found that [Mag98, Ch.2]:

$$\begin{aligned} \widetilde{\mathbf{Q}} &= \mathbf{V}\Sigma^{-1}\mathbf{V}^H \\ \widetilde{\mathbf{Q}}\mathbf{Q}^{-1}\widetilde{\mathbf{Q}} &= \mathbf{V}\Sigma\mathbf{V}^H \mathbf{Q}^{-1}\mathbf{V}\Sigma\mathbf{V}^H \\ \left(\widetilde{\mathbf{Q}}\mathbf{Q}^{-1}\widetilde{\mathbf{Q}}\right)^{\#} &= \mathbf{V}\Sigma^{-1}\left(\mathbf{V}^H \mathbf{Q}^{-1}\mathbf{V}\right)^{-1}\Sigma^{-1}\mathbf{V}^H \end{aligned}$$

and, thus,

$$\begin{aligned} \mathbf{P}\widetilde{\mathbf{Q}} &= \left(\widetilde{\mathbf{Q}}\mathbf{Q}^{-1}\widetilde{\mathbf{Q}}\right)^{\#}\widetilde{\mathbf{Q}}\mathbf{Q}^{-1}\widetilde{\mathbf{Q}} = \mathbf{V}\mathbf{V}^H \\ \widetilde{\mathbf{Q}}^{\#}\widetilde{\mathbf{Q}} &= \widetilde{\mathbf{Q}}\widetilde{\mathbf{Q}}^{\#} = \mathbf{V}\mathbf{V}^H. \end{aligned}$$

taking into account that $\mathbf{V}^H\mathbf{V} = \mathbf{I}_R$.

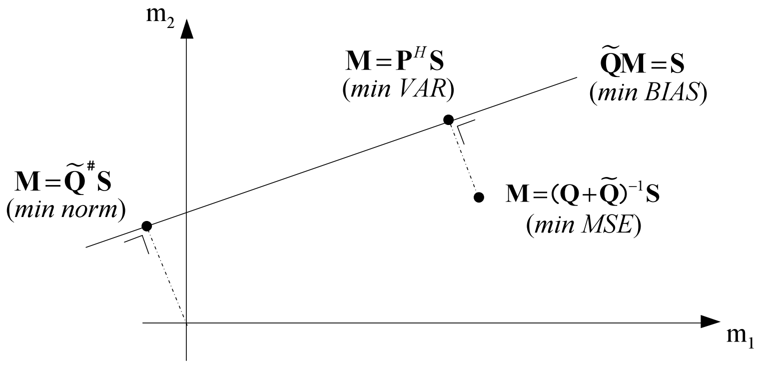

Figure 3.1: Geometric interpretation of the second-order estimators deduced in this chapter for a uniparametric, hypothetical problem in which $\mathbf{M}$ had only 2 coefficients: m_1 and m_2.

is actually the orthogonal projector onto the R-dimensional subspace generated by $\widetilde{\mathbf{Q}}$. The resulting expression is then simplified using the following property of the pseudo-inverse: $\widetilde{\mathbf{Q}}^{\#}\widetilde{\mathbf{Q}}\widetilde{\mathbf{Q}}^{\#} = \widetilde{\mathbf{Q}}^{\#}$ [Mag98, Eq. 5.2].

Notice that any matrix $\mathbf{M}$ solving (3.27) or (3.28) yields the same bias, for example, $\mathbf{M} = \widetilde{\mathbf{Q}}^{\#}\mathbf{S}$. Indeed, among all of them, $\mathbf{M}_{var}$ (3.30) is the one yielding minimum variance (Fig. 3.1).

3.4 A Case Study: Frequency Estimation

In this section, the second-order MMSE and minimum variance estimators are applied to estimate the carrier frequency offset in the context of digital synchronization. This problem has been chosen because closed form expressions exist based on the results in Appendix 3.D.

The signal model for frequency synchronization fits the general linear model in Section 2.4, in which the transfer matrix $\mathbf{A}(\boldsymbol{\theta})$ is given by

$$[\mathbf{A}(\nu)]_k = \exp\left(j2\pi\nu\mathbf{d}_M/N_{ss}\right) \odot [\mathbf{A}]_k$$

where ν and N_{ss} are, respectively, the normalized carrier frequency offset and sampling rate, matrix $\mathbf{A}$ generates the actual modulation and, $\mathbf{d}_M \triangleq [0, ..., M-1]^T$. The precise content of matrix $\mathbf{A}$ in digital synchronization will be detailed in Section 6.1.2.

In addition, a uniform prior is assumed for the unknown carrier frequency ν as the following one:

$$f_\nu(\nu) = \begin{cases} \Delta^{-1} & |\nu| \leq \Delta/2 \\ 0 & \text{otherwise} \end{cases}$$

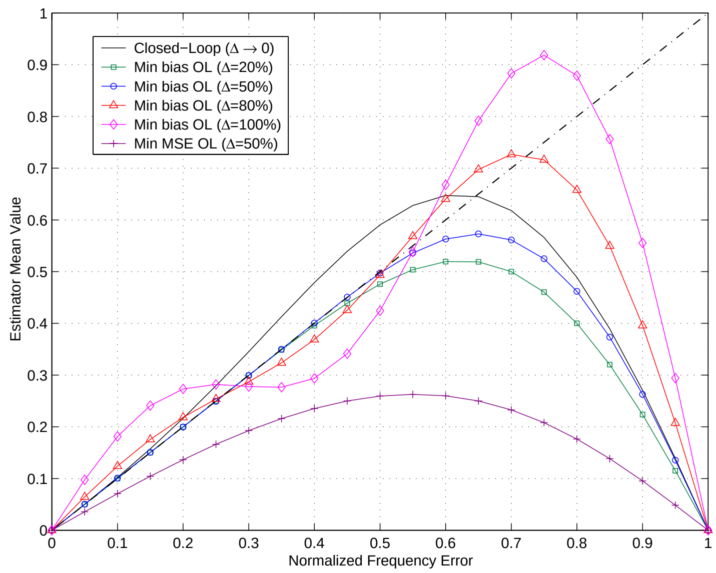

Figure 3.2: Estimator mean response for different values of Δ. The simulation parameters are M=4, N_{ss}=2, Es/No=10dB.

with $\Delta \leq N_{ss}$ determining the frequency offset range[2]. Notice that Δ constitutes the sole prior knowledge about the parameter.

In the following figures, the MMSE and minimum variance estimators are compared in terms of bias, variance and MSE.

The results in this section were partially presented in the following conferences:

- "Sample Covariance Matrix Based Parameter Estimation for Digital Synchronization". J. Villares, G. Vázquez. Proceedings of the *IEEE Global Communications Conference 2002* (Globecom 2002). November 2002. Taipei (Taiwan).
- "Sample Covariance Matrix Parameter Estimation: Carrier Frequency, A Case Study". J. Villares, G. Vázquez. Proceedings of the IEEE *International Conference on Accoustics, Speech and Signal Processing* (ICASSP). April 2003. Hong Kong (China).

3.4.1 Bias Analysis

The estimator mean response $E\{\widehat{\nu}\}$ is plotted as a function of the parameter value for different values of Δ. Fig. 3.2 shows how the minimum variance solution minimizes the estimator bias

[2]Sometimes Δ will be specified as a percentage of N_{ss}.

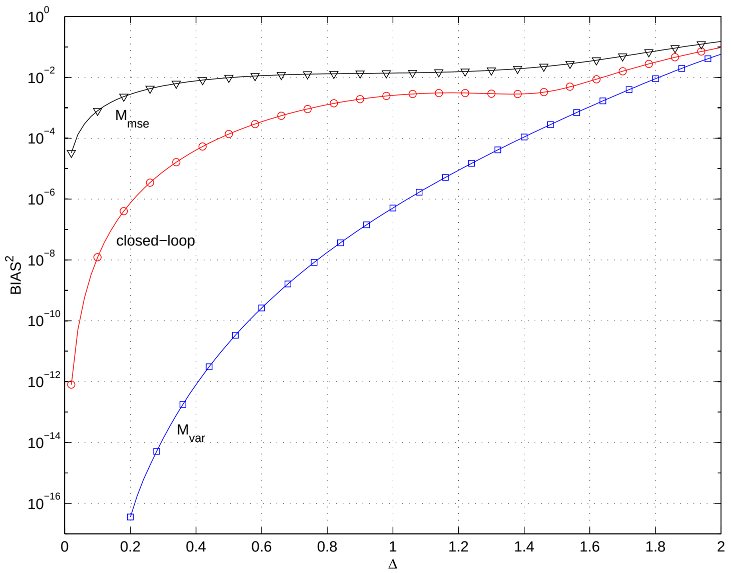

Figure 3.3: Averaged squared bias as a function of the prior range Δ for Es/No=10dB, $M=4$ and N_{ss}=2.

within the prior range. The estimator mean response oscilates around the unbiased response cancelling out the bias of $2\min(LN_{ss}, M) - 1$ points within the prior interval $(-\Delta/2, \Delta/2]$, with L the effective pulse duration (in symbols). These points are automatically selected in order to minimize the overall estimator bias (3.19). This basic result is shortly proved in Appendix 3.E and states that the residual bias is a function of the following ratio

$$\frac{\Delta}{\min(LN_{ss}, M)}$$

Therefore, if the prior range Δ is fixed, the estimator bias can be reduced by oversampling the received signal and/or, if possible, reducing the trasmission bandwidth, i.e., increasing L. Surprisingly, the bias cannot be reduced by augmenting the observation time in the studied frequency estimation problem for $M \geq LN_{ss}$ (Appendix 3.E).

Regarding Fig. 3.2, one concludes that the bias term increases dramatically if Δ/N_{ss} exceeds 0.5 (50%) for the simulated MSK modulation ($L=2$). In the same figure, the mean response of the MMSE estimator is plotted showing how it is clearly biased. This bias is found to increase if the SNR is reduced because, in that case, the MMSE estimator trades more bias for variance. Finally, the S-curve for the closed-loop estimator deduced in the next chapter is depicted. In that case, the estimator is only required to yield unbiased estimates around the origin ($\nu = 0$). As it will be studied with more detail in Chapter 4, the closed-loop solution is obtained considering the asymptotic case in which $\Delta \to 0$.

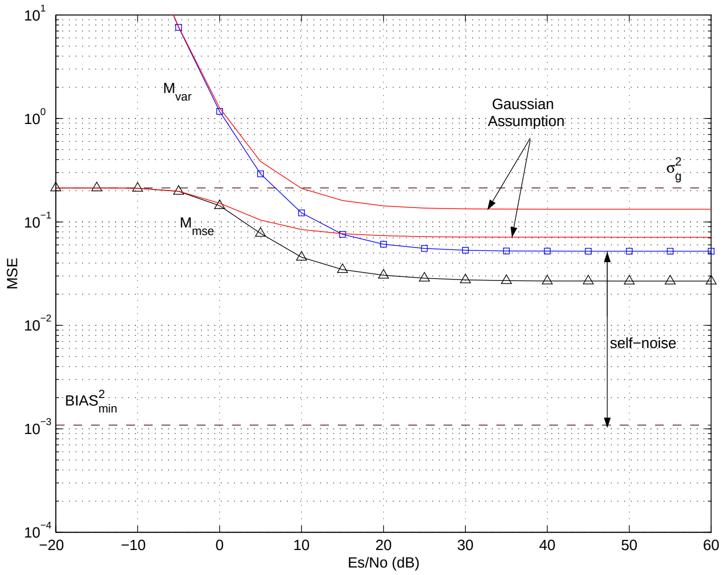

Figure 3.4: Normalized MSE for the MMSE and the minimum variance frequency estimators deduced in (3.25) and (3.30) for the MSK modulation. The corresponding estimators deduced under the Gaussian assumption are also plotted for comparison. The simulation parameters are $M = 8$, $N_{ss} = 2$ and $\Delta = 1.6$.

Another interesting simulation is presented in Fig. 3.3 in which the squared bias $BIAS^2$ (3.19) is plotted as a function of Δ for the MMSE estimator ($\mathbf{M}_{mse}$), the minimum variance estimator ($\mathbf{M}_{var}$) and the closed-loop small-error estimator ($\Delta \rightarrow 0$) deduced in Chapter 4. The SNR is set to 10 dB and, therefore, the noise induced variance is very significative. This fact justifies the relaxation of the MMSE estimator with respect to the bias term. Notice that the three estimators are able to cancel out the bias term if the prior range approaches zero ($\Delta \rightarrow 0$). This simple remark is of paramount importance in the following sections and motivates the need of *closed-loop* algorithms for second-order blind parameter estimation (Chapter 4).

3.4.2 MSE Performance

In this section, the performance of second-order frequency estimators is evaluated in terms of their mean square error (3.21). Observing the following figures, the next remarks are relevant:

- *A priori knowledge.* The performance of the MMSE estimator is upper bounded at low SNR by the *a priori* mean square error σ_g^2 (Fig. 3.4). In such a noisy scenario, the MMSE solution becomes biased with the aim of limiting the variance increase caused by the noise-induced variability. As the SNR increases, the observation provides more

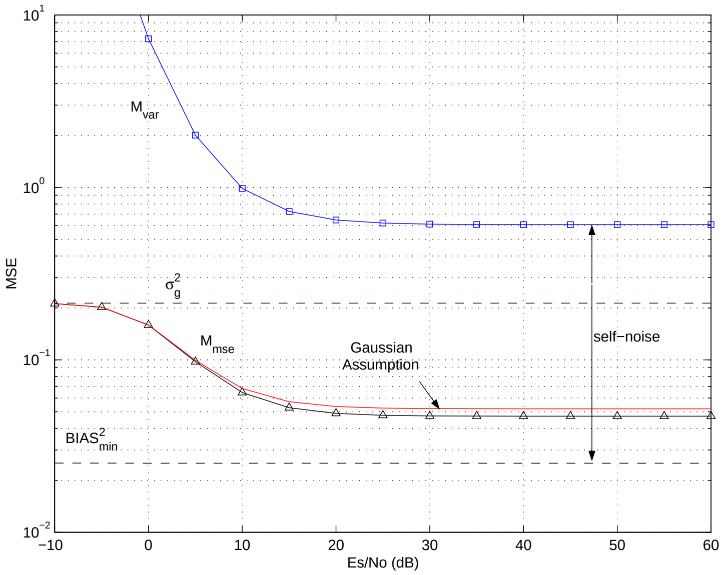

Figure 3.5: Normalized MSE for the MMSE and the minimum variance frequency estimators deduced in (3.25) and (3.30) for the 16-QAM modulation. The transmitted pulse is a square-root raised cosine with roll-off 0.75 truncated at $\pm 5T$. The corresponding estimators deduced under the Gaussian assumption are also plotted for comparison. The simulation parameters are $M = 8$, $N_{ss} = 2$, $K = 13$ and $\Delta = 1.6$.

information about the parameter of interest and this information is exploited to reduce the average MSE.

- *Self-Noise.* For finite observations (M finite), the studied quadratic estimators manifest a significant *variance floor* at high SNR due to the so-called *self-noise* (Fig. 3.4). Remember that self-noise refers to the random fluctuations caused by the unknown nuisance parameters $\mathbf{x}$ in blind estimation schemes (See Section 2.4.1). Effectively, the feed-forward estimators presented in this section are unable to cancel out the self-noise for all the possible values of ν. On the other hand, the self-noise free condition is guaranteed in the case of *closed-loop* ($\Delta \to 0$) second-order frequency estimators as shown in Fig. 3.7. Consequently, the amount of information that can be drawn from the current sample $\mathbf{y}$ is very limited in the studied case due to the presence of self-noise. In fact, the level of the high-SNR floor is a function of the observation time M (Fig. 3.8) as well as the prior range Δ (Fig. 3.7).

- *Modulation.* If the figures 3.4-3.6 are compared, one concludes that the performance of the MMSE estimator is practically insensitive to the actual distribution of the transmitted symbols. However, the incurred minimum bias, $BIAS^2_{\min}$, depends on the transmitted

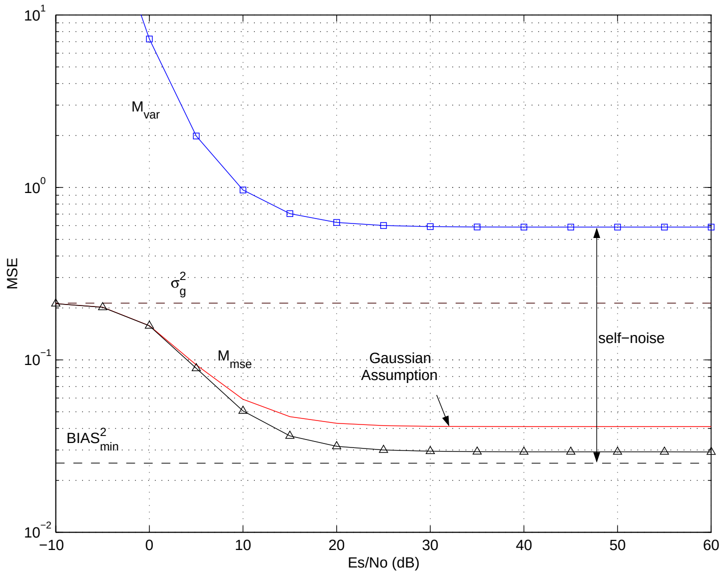

Figure 3.6: Normalized MSE for the MMSE and the minimum variance frequency estimators deduced in (3.25) and (3.30) for the MPSK modulation. The transmitted pulse is a square-root raised cosine with roll-off 0.75 truncated at $\pm 5T$. The corresponding estimators deduced under the Gaussian assumption are also plotted for comparison. The simulation parameters are $M = 8$, $N_{ss} = 2$, $K = 13$ and $\Delta = 1.6$.

pulse. For the considered Nyquist pulse of roll-off 0.75, the minimum variance solution becomes significantly degraded with respect to the MSK performance for any SNR (Fig. 3.4). Specifically, the bias and self-noise contribution is more significative for the simulated MPSK and 16-QAM modulations.

- *Bias vs. variance trade-off.* The MMSE solution outperforms the minimum variance solution because it is not forced to minimize the bias. On the contrary, it tolerates some bias if the variance term can be attenuated in return, minimizing so the overall MSE. This trade-off is more significant in the low SNR regime but it is also observed at high SNR on account of the self-noise variance. If the self-noise variance is reduced by increasing M, the minimum variance solution converges to the MMSE solution at high SNR (Fig. 3.8).

- *Consistency.* For large samples ($M \to \infty$, with N_{ss} constant), the estimator variance is completely removed whatever the actual SNR and the residual MSE is the estimator bias computed in (3.36). Therefore, consistent second-order estimation is not possible unless the bias term $BIAS^2_{\min}$ vanishes as explained in Section 3.3. This asymptotic result applies

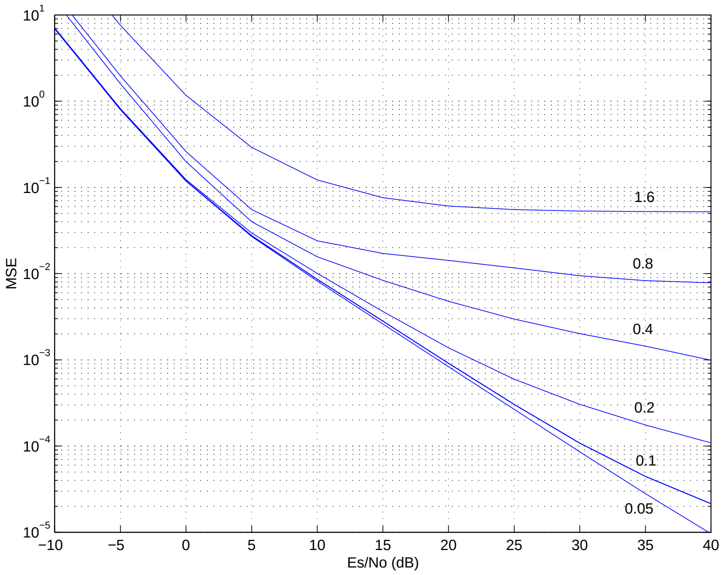

Figure 3.7: MSE corresponding to the minimum variance solution for different values of the parameter range Δ=0.05, 0.1, 0.2, 0.4, 0.8 and 1.6. The received signal is MSK-modulated and M=8 samples are processed with N_{ss}=2.

to both the MMSE and minimum variance solutions (Fig. 3.8). Formally,

$$\lim_{M\to\infty} MSE = \lim_{M\to\infty} BIAS_{\min}^2 = \sigma_g^2 - \lim_{M\to\infty} \mathrm{Tr}\left\{\mathbf{S}^H \widetilde{\mathbf{Q}}^{\#} \mathbf{S}\right\}$$

where the last term becomes constant for $M \geq N_{ss}L$. Notice that the MSE curves in Fig. 3.8 would eventually converge to the bias floor shown in Fig. 3.4 if the M-axis were expanded, i.e., $\lim_{M\to\infty} BIAS_{\min}^2 \approx 10^{-3}$.

- *Gaussian assumption.* The Gaussian assumption is checked in Fig. 3.4 showing that it yields a significant loss for medium-to-high SNRs. On the other hand, it converges to the optimal solution as the SNR approaches to zero. Regarding Fig. 3.8, the Gaussian assumption also supplies asymptotically ($M \to \infty$) self-noise free estimators but it suffers a constant penalty for any finite SNR. This loss is less significative in the case of considering a linear modulation as shown in Figs. 3.5-3.6.

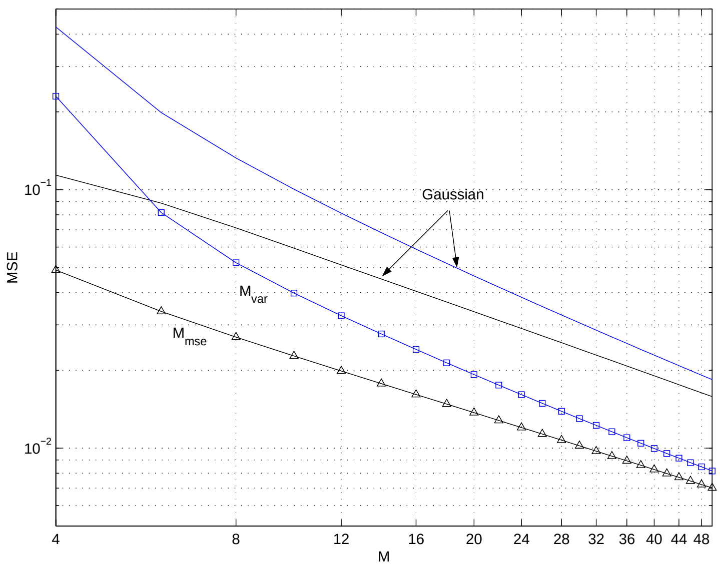

Figure 3.8: Normalized MSE for the MMSE and the minimum variance frequency estimators deduced in (3.25) and (3.30). The modulation is MSK, Es/No=40dB, $N_{ss} = 2$ and $\Delta = 1.6$.

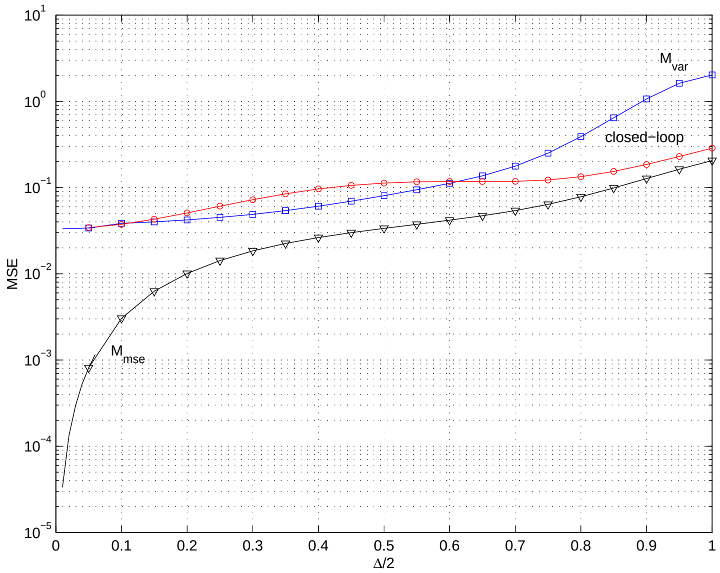

Figure 3.9: MSE as a function of Δ for the MSK modulation. The simulation parameters are Es/No=10dB, $M = 4$ and N_{ss}=2.

3.5 Conclusions

This chapter was devoted to design feedforward second-order estimators, adopting the well-known Bayesian approach. The coefficients of the quadratic estimator were selected to minimize the estimator MSE or the estimator variance on the average, where this average involves the a priori distribution of the unknown parameters. In the optimization of the estimator coefficients, the actual distribution of the nuisance parameters was considered avoiding the usual Gaussian assumption.

The applicability of the studied second-order estimators in nonlinear estimation problems becomes generally limited due to the impossibility of cancelling the bias term. Indeed, *consistent* second-order estimators are mostly unfeasible due to the persistent bias term. Moreover, if the observation time is finite, a variance floor appears at high SNR due to presence of the random nuisance parameters. This floor depends on the actual distribution of the nuisance parameters and can be reduced exploiting their actual distribution, especially in case of CPM signals.

Nonetheless, most of these conclusions depend on the actual parameterization and the assumed prior distribution. In this chapter, the problem of blind frequency synchronization was chosen to illustrate these conclusions by means of analytical and numerical results. In this case study, the minimization of the estimator bias –within the parameter range– is proved to be limited by the effective duration of the transmitted pulse. On the other hand, open-loop second-order frequency estimators exhibited the referred variance floor at high SNR, whereas self-noise free closed-loop frequency estimators exist in the literature even for limited observation times.

Beyond the practical interest of open-loop second-order estimators, the formulation in this chapter constitutes the basis for the deduction of optimum quadratic closed-loop estimators in Chapter 4.

Appendix 3.A Second-order estimation in noncircular transmissions

In the main text, optimal second-order estimators have been deduced for complex, circular constellations. The nuisance parameters circularity can be assumed for any bandpass modulation if the carrier phase is uniformly distributed and this random term is incorporated into the vector of nuisance parameters $\mathbf{x}$. In that case, the expectation of $\mathbf{y}\mathbf{y}^T$ becomes zero and does not provide information about the parameter of interest. However, optimal second-order estimators should also exploit the improper correlation matrix $\mathbf{y}\mathbf{y}^T$ in case of baseband transmissions or noncircular bandpass modulations, provided that the carrier phase is known or estimated. Precisely, the carrier phase estimation is addressed in Section 6.2 using quadratic schemes in case of MSK-type modulations. Other important noncircular modulations are the CPM format, any staggered modulation (e.g., offset QPSK), any real-valued constellation such as BPSK or ASK, trellis coded modulations (TCM) as well as other coded transmissions [Pro95].

The analysis of noncircular or improper complex random variables has been carried out in [Sch03][Pic96] and references therein. *Widely-linear* estimators are proposed in [Sch03][Pic95] in which the vector

$$\mathbf{z} \triangleq \begin{bmatrix} \mathbf{y} \\ \mathbf{y}^* \end{bmatrix}$$

is linearly processed. This extended signal model has been applied in the field of communications by some authors, e.g., [Gel00][Tul00] [Ger01].

Therefore, all the results in this thesis can be extended by considering the following sample covariance matrix

$$\widehat{\mathbf{R}} = \mathbf{z}\mathbf{z}^T = \begin{bmatrix} \mathbf{y}\mathbf{y}^T & \mathbf{y}\mathbf{y}^H \\ \mathbf{y}^*\mathbf{y}^T & \mathbf{y}^*\mathbf{y}^H \end{bmatrix}$$

to obtain the optimal *widely-quadratic* estimator. When stacking the sample covariance matrix, it is worth realizing that $\mathbf{y}^*\mathbf{y}^T$ could be omitted from $\widehat{\mathbf{r}} = \mathrm{vec}(\widehat{\mathbf{R}})$ because the term $\mathbf{y}\mathbf{y}^H$ provides the same information.

To compute the coefficients of the optimal second-order estimator, it is necessary to obtain the covariance of $\widehat{\mathbf{r}} = \mathrm{vec}(\widehat{\mathbf{R}})$ following the guidelines in Appendix 3.B.

Appendix 3.B Deduction of matrix $\mathbf{Q}(\boldsymbol{\theta})$

The expression of $\mathbf{Q}(\boldsymbol{\theta})$ in (3.9) can be written as follows:

$$\mathbf{Q}(\boldsymbol{\theta}) = E\left\{\widehat{\mathbf{r}}\widehat{\mathbf{r}}^H\right\} - \mathbf{r}\mathbf{r}^H \tag{3.37}$$

where

$$\widehat{\mathbf{r}} = \operatorname{vec}\left\{(\mathbf{A}\mathbf{x}+\mathbf{w})(\mathbf{A}\mathbf{x}+\mathbf{w})^H\right\} = \operatorname{vec}\left\{\mathbf{A}\mathbf{x}\mathbf{x}^H\mathbf{A}^H + \mathbf{A}\mathbf{x}\mathbf{w}^H + \mathbf{w}\mathbf{x}^H\mathbf{A}^H + \mathbf{w}\mathbf{w}^H\right\}$$

and the dependence on $\boldsymbol{\theta}$ is omitted for the sake of brevity.

Taking into account the noise is circular and zero mean, i.e.,

$$\begin{aligned} E\{\mathbf{w}\} &= \mathbf{0} \\ E\left\{\mathbf{w}\mathbf{w}^T\right\} &= \mathbf{0} \\ E\left\{w_i w_j^* w_k\right\} &= E\left\{w_i^* w_j^* w_k\right\} = 0, \end{aligned}$$

only six terms, out of the sixteen in $\widehat{\mathbf{r}}\widehat{\mathbf{r}}^H$, survive to the expectation in (3.37). These terms can be classified as follows:

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

Then, using the following three properties [Mag98, Chapter 2]:

$$\operatorname{vec}\left(\mathbf{A}\mathbf{B}\mathbf{C}^H\right) = \left(\mathbf{C}^* \otimes \mathbf{A}\right)\operatorname{vec}(\mathbf{B}) \tag{3.38}$$

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{A}\mathbf{C}\otimes\mathbf{B}\mathbf{D} \tag{3.39}$$

$$\operatorname{vec}\left(\mathbf{a}\mathbf{b}^H\right)\operatorname{vec}^H\left(\mathbf{a}\mathbf{b}^H\right) = \left(\mathbf{b}^*\otimes\mathbf{a}\right)\left(\mathbf{b}^*\otimes\mathbf{a}\right)^H = \left(\mathbf{b}\mathbf{b}^H\right)^*\otimes\mathbf{a}\mathbf{a}^H \tag{3.40}$$

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

$$\begin{aligned} E\left\{\widehat{\mathbf{r}}\widehat{\mathbf{r}}^H\right\} = {} & \mathcal{A}E\left\{\operatorname{vec}\left(\mathbf{x}\mathbf{x}^H\right)\operatorname{vec}^H\left(\mathbf{x}\mathbf{x}^H\right)\right\}\mathcal{A}^H + \\ & + \mathbf{R}_w^*\otimes\mathbf{A}\mathbf{A}^H + \left(\mathbf{A}\mathbf{A}^H\right)^*\otimes\mathbf{R}_w + \operatorname{vec}\left(\mathbf{A}\mathbf{A}^H\right)\operatorname{vec}^H(\mathbf{R}_w) + \operatorname{vec}(\mathbf{R}_w)\operatorname{vec}^H\left(\mathbf{A}\mathbf{A}^H\right) \\ & + \mathbf{R}_w^*\otimes\mathbf{R}_w + \operatorname{vec}(\mathbf{R}_w)\operatorname{vec}^H(\mathbf{R}_w) \end{aligned} \tag{3.41}$$

where $\mathcal{A} \triangleq \mathbf{A}^*\otimes\mathbf{A}$ and the following property of Gaussian vectors is used (Appendix 3.C):

$$E\left\{\operatorname{vec}\left(\mathbf{w}\mathbf{w}^H\right)\operatorname{vec}^H\left(\mathbf{w}\mathbf{w}^H\right)\right\} = \mathbf{R}_w^*\otimes\mathbf{R}_w + \operatorname{vec}(\mathbf{R}_w)\operatorname{vec}^H(\mathbf{R}_w). \tag{3.42}$$

Therefore, grouping terms in (3.41) and having in mind that $\mathbf{R} = \mathbf{A}\mathbf{A}^H + \mathbf{R}_w$ (2.23), the following expression is obtained:

$$\begin{aligned} E\left\{\widehat{\mathbf{r}}\widehat{\mathbf{r}}^H\right\} &= \mathcal{A}E\left\{\operatorname{vec}\left(\mathbf{x}\mathbf{x}^H\right)\operatorname{vec}^H\left(\mathbf{x}\mathbf{x}^H\right)\right\}\mathcal{A}^H \\ &- \left(\mathbf{A}\mathbf{A}^H\right)^*\otimes\mathbf{A}\mathbf{A}^H - \operatorname{vec}\left(\mathbf{A}\mathbf{A}^H\right)\operatorname{vec}^H\left(\mathbf{A}\mathbf{A}^H\right) \\ &+ \mathbf{R}^*\otimes\mathbf{R} + \operatorname{vec}\left(\mathbf{R}\right)\operatorname{vec}^H\left(\mathbf{R}\right) \end{aligned}$$

Finally, using once more (3.38) and (3.39) in order to write the negative terms above as a function of $\mathcal{A}$ and, plugging this result into (3.37), the expression proposed in (3.10) is obtained:

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

Appendix 3.C Fourth-order moments

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

$$\mathbf{Q}_v = E\left\{\mathrm{vec}\left(\mathbf{v}\mathbf{v}^H\right)\mathrm{vec}^H\left(\mathbf{v}\mathbf{v}^H\right)\right\} \tag{3.43}$$

whose elements are given by

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

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

If all these elements are arranged in $\mathbf{Q}_v$, three components are identified:

$$\mathbf{Q}_v = \mathrm{vec}\left(\mathbf{R}_v\right)\mathrm{vec}^H\left(\mathbf{R}_v\right) + \mathbf{R}_v^* \otimes \mathbf{R}_v + \mathrm{diag}\left(\mathrm{vec}\left(\Gamma\right)\right)$$

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

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

$$\mathbf{Q}_v = \mathrm{vec}\left(\mathbf{R}_v\right)\mathrm{vec}^H\left(\mathbf{R}_v\right) + \mathbf{R}_v^* \otimes \mathbf{R}_v + \mu^2\left(\rho - 2\right)\mathrm{diag}\left(\mathrm{vec}\left(\mathbf{I}_L\right)\right) \tag{3.44}$$

In particular, the fourth-order moments of $\mathbf{x}$ in (3.12) are given by (3.44) having in mind that the symbols autocorrelation is $E\left\{\mathbf{x}\mathbf{x}^H\right\} = \mathbf{I}_K$ and, thus, we have that $\mathbf{R}_v = \mathbf{I}_K$ and $\mu = 1$. On the other hand, if $\mathbf{v}$ is a complex Gaussian vector, as the noise vector $\mathbf{w}$ in the adopted signal model, the third term in (3.44) can be removed taking into account that $\rho = 2$ in the Gaussian case, hence proving equation (3.42):

$$E\left\{\mathrm{vec}\left(\mathbf{w}\mathbf{w}^H\right)\mathrm{vec}^H\left(\mathbf{w}\mathbf{w}^H\right)\right\} = \mathbf{R}_w^* \otimes \mathbf{R}_w + \mathrm{vec}\left(\mathbf{R}_w\right)\mathrm{vec}^H\left(\mathbf{R}_w\right).$$

Appendix 3.D Bayesian average in frequency estimation

Let us assume that the scalar parameter λ is estimated from the following observation:

$$\mathbf{y} = \exp\left(j2\pi\lambda\mathbf{d}_M\right) \odot \mathbf{A}\mathbf{x} + \mathbf{w}$$

where $\mathbf{d}_M \triangleq [0, ..., M-1]^T$ and $\mathbf{A}$ stands for $\mathbf{A}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\mathbf{0}}$. Therefore, the observation covariance matrix is given by

$$\mathbf{R}(\lambda) = E\left\{\mathbf{y}\mathbf{y}^H\right\} = \mathbf{E}(\lambda) \odot \mathbf{A}\mathbf{A}^H + \mathbf{R}_w$$

where $\mathbf{E}(\lambda)$ is defined as

$$[\mathbf{E}(\lambda)]_{i,k} = e^{j2\pi\lambda(i-k)}. \tag{3.45}$$

Let us consider that the prior is uniform in the interval $\lambda \in (-\Delta/2, \Delta/2]$ with $\Delta \leq 1$. In that case, it is possible to obtain closed-form expressions for those matrices appering in $\mathbf{b}_{opt}$ (3.16), $\mathbf{M}_{mse}$ (3.25) and, $\mathbf{M}_{var}$ (3.30). The resulting expressions are listed next:

$$\begin{aligned}
g &= E_\lambda\left\{g(\lambda)\right\} = E_\lambda\left\{\lambda\right\} = 0 \\
\sigma_g^2 &= E_\lambda\left\{\lambda^2\right\} = \Delta^2/12 \\
\mathbf{R} &= \mathbf{E} \odot \mathbf{A}\mathbf{A}^H + \mathbf{R}_w \\
\mathbf{Q} &= \mathbf{R}^* \otimes \mathbf{R} + \left(\mathbf{E}_q - \mathbf{E}^* \otimes \mathbf{E}\right) \odot \mathcal{A}\mathcal{A}^H + \mathbf{E}_q \odot \mathcal{A}\mathbf{K}\mathcal{A}^H \\
\widetilde{\mathbf{Q}} &= \left[\mathbf{E}_q - \operatorname{vec}(\mathbf{E}) \operatorname{vec}^H(\mathbf{E})\right] \odot \operatorname{vec}\left(\mathbf{A}\mathbf{A}^H\right) \operatorname{vec}^H\left(\mathbf{A}\mathbf{A}^H\right) \\
\mathbf{s} &= \operatorname{vec}\left(\mathbf{E}_s \odot \mathbf{A}\mathbf{A}^H\right)
\end{aligned}$$

with

$$\begin{aligned}
\mathbf{E} &\triangleq E_\lambda\left\{\mathbf{E}(\lambda)\right\} \\
\mathbf{E}_q &\triangleq E_\lambda\left\{\mathbf{E}^*(\lambda) \otimes \mathbf{E}(\lambda)\right\} \\
\mathbf{E}_s &\triangleq E_\lambda\left\{\mathbf{E}(\lambda)\lambda\right\}
\end{aligned}$$

whose elements are given next [Vil03a]:

$$\begin{aligned}
[\mathbf{E}]_{i,k} &= \operatorname{sinc}\left((i-k)\Delta\right) \\
[\mathbf{E}_q]_{i+Mj,k+Ml} &= \operatorname{sinc}\left((i-j+l-k)\Delta\right) \\
[\mathbf{E}_s]_{i,k} &= \begin{cases} 0 & i = k \\ \frac{j}{2\pi(i-k)}\left[\operatorname{sinc}\left((i-k)\Delta\right) - \cos\left(\pi(i-k)\Delta\right)\right] & i \neq k \end{cases}
\end{aligned}$$

and the sinc $(\cdot)$ operator is defined as $\operatorname{sinc}(x) \triangleq \sin(\pi x)/(\pi x)$ with $\operatorname{sinc}(0) = 1$.

Appendix 3.E Bias study in frequency estimation

In this appendix, the minimum bias solution is studied in detail for the frequency estimation problem. The coefficients $\mathbf{m} \triangleq \text{vec}(\mathcal{M})$ minimizing the estimator bias in (3.19) have to satisfy the minimum-bias constraints in (3.27). After some trivial manipulations, this equation can be written as

$$E_\nu \{\mathbf{B}(\nu) \overline{\alpha}^*(\nu)\} = E_\nu \{\mathbf{B}(\nu) \nu\} \tag{3.46}$$

where $\mathbf{B}(\nu) \triangleq \mathbf{A}(\nu) \mathbf{A}^H(\nu) = \mathbf{E}(\nu/N_{ss}) \odot \mathbf{A}\mathbf{A}^H$ (Appendix 3.D) and

$$\overline{\alpha}^*(\nu) = (\mathbf{r}(\nu) - \mathbf{r})^H \mathbf{m} = \text{Tr}\left\{\mathbf{B}^H(\nu) \mathcal{M}\right\} - C$$

stands for the estimator mean value (3.7) as a function of the parameter ν and

$$C \triangleq E_\nu \left\{\text{Tr}\left\{\mathbf{B}^H(\nu) \mathcal{M}\right\}\right\}$$

is a constant term, which is independent of the parameter ν. Notice also that $\overline{\alpha}(\nu)$ is actually real-valued despite the complex conjugation in (3.46), that is kept for the sake of generality.

Regarding the obtained minimum bias equation (3.46), it is straightforward to realize that any unbiased estimator verifies (3.46). Unfortunately, the converse is not usually possible and (3.46) supplies the least squares fitting of $\overline{\alpha}(\nu)$ to the ideal linear response $\overline{\alpha}(\nu) = \nu$ within the prior domain (i.e., $|\nu| < \Delta/2$).

Furthermore, if some elements of $\mathbf{B}(\nu)$ are connected by an affine transformation, i.e., $[\mathbf{B}(\nu)]_{i_2,j_2} = C_a [\mathbf{B}(\nu)]_{i_1,j_1} + C_b$ for any value of C_a and C_b, the system of equations in (3.46) becomes underdetermined, as it was equation (3.27). Indeed, this is exactly what happens in the frequency estimation case since the diagonal entries of $\mathbf{B}(\nu)$ share the same phasor (3.45). Thus, it is possible to reduce (3.46) to $2M-1$ equations corresponding to the diagonals of $\mathbf{B}(\nu)$. Nonetheless, the uppest and lowest diagonals are equal to zero if $M > N_{ss}L$ with L the *effective* transmitted pulse duration (in symbols). Therefore, the minimization of the estimator bias requires to fulfill the following $2K+1$ equations:

$$E_\nu \left\{\overline{\alpha}^*(\nu) e^{j2\pi\nu k/N_{ss}}\right\} = E_\nu \left\{\nu e^{j2\pi\nu k/N_{ss}}\right\} \qquad k \in [-K, K]$$

or, equivalently,

$$\int_{-R/2}^{R/2} V(f) e^{j2\pi fk} df = N_{ss} \int_{-R/2}^{R/2} f e^{j2\pi fk} df \qquad k \in [-K, K] \tag{3.47}$$

where $K \triangleq \min(M, LN_{ss}) - 1$, $f \triangleq \nu/N_{ss}$, $R \triangleq \Delta/N_{ss}$ is the carrier uncertainty relative to the Nyquist bandwidth and,

$$V(f) \triangleq \text{Tr}\left\{\mathbf{B}^H(N_{ss}f) \mathcal{M}\right\} = \sum_{k=-K}^{K} \sum_i [\mathcal{M}]_{i,i+k} [\mathbf{B}^*(N_{ss}f)]_{i,i+k}$$

$$= \sum_{k=-K}^{K} \left(\sum_i [\mathcal{M}]_{i,i+k} \left[\mathbf{A}\mathbf{A}^H\right]^*_{i,i+k}\right) e^{-j2\pi fk}$$

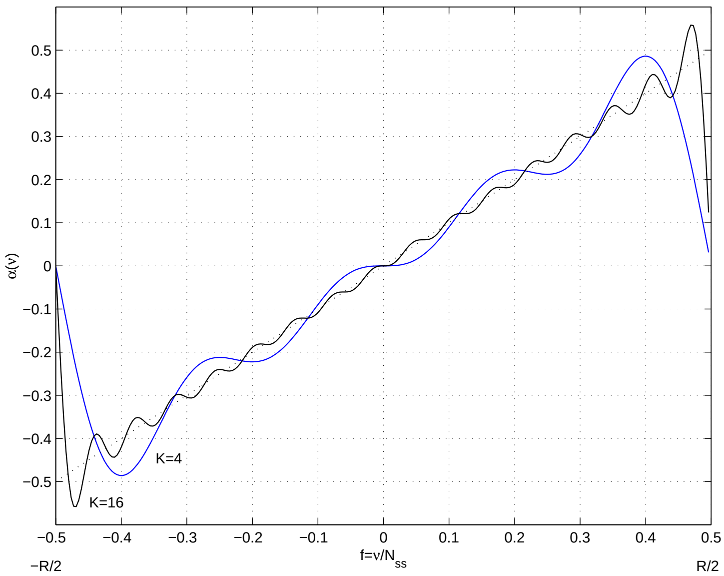

Figure 3.10: Mean value of the frequency estimator corresponding to the minimum bias solution for K=4 and 16.

is the Fourier transform of the sequence $v[k]$ defined as

$$v[k] \triangleq \mathfrak{F}^{-1}\left\{V\left(f\right)\right\} = \begin{cases} \sum_i \left[\mathcal{M}\right]_{i,i+k} \left[\mathbf{A}\mathbf{A}^H\right]^*_{i,i+k} & |k| \leq K \\ 0 & otherwise. \end{cases} \tag{3.48}$$

Notice that in (3.47) we have taken into account that $\overline{\alpha}^*\left(\nu\right) = V\left(v/N_{ss}\right) - C$ where C must be null to guarantee the *odd* symmetry of the harmonic expansion of f in the right-hand side of (3.47).

Thus, equation (3.47) states that the $2K+1$ central terms of the discrete Fourier series of $N_{ss}f$ and $V(f)$, filtered in the interval $\pm R/2$, must be identical in order to minimize the estimator bias. Formally, this means that the sequence $v[k]$ must be equal to the inverse discrete Fourier transform of $N_{ss}f$ as stated in the next equation:

$$v[k] = \frac{jN_{ss}}{2\pi k}\left[\delta\left[k\right] - \left(-1\right)^k\right] \qquad |k| < K$$

Ideally, if K were arbitrarily long, (3.47) would imply the identity of $\overline{\alpha}\left(\nu\right)$ and ν within the prior interval $|\nu| < \Delta/2$ or, in other words,

$$V\left(f\right)|_{K\to\infty} = \lim_{K\to\infty}\sum_{k=-K}^{K} v[k]e^{-j2\pi fk} = N_{ss}f \qquad |f| < R/2 \tag{3.49}$$

whatever the value of R. However, since K is finite and is limited by the transmitted pulse duration L, the value at which the above Fourier series can be truncated without noticiable distortion is a function of the ratio $R = \Delta/N_{ss}$; the smaller R, the less terms are required for the same distortion of $\overline{\alpha}(\nu)$. In the limit ($R \to 0$), the Taylor expansion of (3.49) around $f = 0$ ensures that $K = 1$ is sufficient to hold exactly (3.49) with $v[1] = -v[-1] = jN_{ss}/(2\pi)$. Otherwise, if (3.49) is truncated taking too few elements, $\overline{\alpha}(\nu)$ will suffer from ripple and the Gibbs effect, i.e., the overshooting at the discontinuity points $|\nu| = \pm\Delta/2$, as shown in Fig. 3.10 for the most critical situation in which $R = 1$.

Finally, notice that the effective duration L is inversely proportional to the effective signal bandwidth. Because the minimum transmission bandwidth in bandpass communications is $1/T$ Hz (i.e., 0% roll-off), it follows that the main lobe of the signal autocorrelation lasts $2T$ seconds and, thus, in practice the Fourier series in (3.49) becomes truncated approximately at $K = N_{ss}$ or, in the best case, at a few multiples of N_{ss}.