BIAS CORRECTIONS IN LINEAR MMSE ESTIMATION WITH LARGE FILTERS

Jordi Serra and Francisco Rubio

Centre Tecnològic de Telecomunicacions de Catalunya (CTTC)
Parc Mediterrani de la Tecnologia, Av. Canal Olímpic s/n, 08860 Castelldefels (Barcelona, Spain)
URL: http://www.cttc.es/, e-Mail: {jordi.serra, francisco.rubio}@cttc.es

ABSTRACT

We investigate optimal bias corrections in the problem of linear minimum mean square error (LMMSE) estimation of a scalar parameter linearly described by a set of Gaussian multidimensional observations. The problem of finding the optimal scaling of a class of LMMSE filter implementations based on the sample covariance matrix (SCM) is addressed. By applying recent results from random matrix theory, the scaling factor minimizing the mean square error (MSE) and depending on both the unknown covariance matrix and its sample estimator is firstly asymptotically analyzed in terms of key scenario parameters, and finally estimated using the SCM. As a main result, a universal scaling factor minimizing the estimator MSE is obtained which dramatically outperforms the conventional LMMSE filter implementation. A Bayesian setting assuming random unknown parameters with known mean and variance is considered in this paper, but exactly the same methodology applies to the classical estimation setup considering deterministic parameters.

1. INTRODUCTION

A large number of signal processing problems can be addressed by performing a filtering operation on a set of multidimensional observations in order to extract a certain parameter of interest. In many of these applications, the underlying structure of the observations is linear in the unknown parameter to be estimated. By forcing the parameter estimator to be unbiased, a linear transformation can be found under the linear model assumption having the smallest mean square-error, or equivalently minimum variance, among all linear transformations. The former is known as the best linear unbiased estimator (BLUE), which in statistical signal processing is usually referred to and commonly implemented via the minimum variance distorsionless response (MVDR) filter. Under the further assumption of Gaussian observations, the previous linear estimator is also the minimum variance unbiased estimator (MVUE) [1].

Allowing for some bias in the design of optimal estimators by considering the MSE as design objective leads in general to unrealizable estimators, in the sense that they depend on the unknown parameter to be estimated. Nevertheless, examples of realizable biased estimators can be found in the literature with a MSE lower than that of the MVUE, and even below of the unbiased Cramér-Rao Bound (CRB). Seminal works following this approach are the ones by James and Stein [2, 3], who in the context of estimating the mean of a Gaussian random vector showed that the MSE can be further reduced by using a biased estimator consisting of a nonlinear shrinkage of the maximum likelihood (ML) method. In [4], the problem of estimating the unknown variance of a set of independent and identically distributed (i.i.d.) Gaussian random variables with mean zero is addressed. In particular, it is shown that introducing a linear scaling factor in the MVUE leads to a reduced MSE, even outperforming the unbiased CRB. In [5] and [6], a more general framework than in [4] is considered in order to further reduce the MSE of a general MVUE by considering a biased version of the estimator obtained via a linear scaling factor, which is optimized in terms of the minimum MSE achievable. Furthermore, it is shown that even in the case of an optimal biased estimator depending upon the unknown parameters, a minimax approach can be applied to circumvent the problem and to effectively reduce the MSE of the MVUE.

In case some distributional information is available regarding the unknown parameter, a Bayesian approach can be followed in order to find a way to further minimize the MSE of a linear estimator by effectively controlling the amount of bias introduced. This is regarded in the literature as the linear or affine minimum mean square-error (MMSE) estimator and, under some weak regularity assumptions, is uniquely defined as the conditional expectation of the unknown random parameter given the observed sample [1]. Moreover, if the sample and the random parameter are jointly Gaussian, then the linear MMSE estimator is also the MMSE.

In this paper, we will consider the previous setup in order to investigate a class of improved bias corrections in linear MMSE estimation applied to a set of samples linearly describing an unknown parameter embedded in noise. More specifically, the problem of optimally scaling the MVDR filter implementation based on the sample covariance matrix (SCM) with the aim of reducing the MSE is addressed. Our approach is based on recent results from the theory of the spectral analysis of large random matrices, or random matrix theory (RMT). First, the MSE performance measure is asymptotically approximated as a function of the linear scaling factor, the unknown theoretical covariance matrix defining the scenario, and the problem dimensions, i.e., the number of samples and the observation dimension. Then, the optimal scaling factor effectively rendering the bias correction is estimated by using the SCM. We notice that the same approach can be directly applied to the classical minimum variance estimation setup where the unknown parameter is assumed to be deterministic.

2. OPTIMAL SCALING OF LINEAR MMSE FILTERS

Consider a collection of multivariate observations $\left\{\mathbf{y}(n) \in \mathbb{C}^M, n = 1,\ldots,N\right\}$ obtained, for instance, by sampling across an antenna array with M sensors, namely, $\{y_m(n), n = 1,\ldots,N, m = 1,\ldots,M\}$, such that the observations $\mathbf{y}(n) = \left[\begin{array}{ccc} y_1(n) & \cdots & y_M(n) \end{array}\right]^T$ can be described according to the following linear data model that properly defines the structure of a vast number of estimation problems in statistical signal processing, namely,

$$\mathbf{y}(n) = x(n)\mathbf{s} + \mathbf{n}(n),$$

where $x(n)$ is an unknown parameter, denoting the signal of interest (SOI) waveform, that is observed in unknown colored noise $\mathbf{n}(n) \in \mathbb{C}^M$ after being operated upon by the known signature vector $\mathbf{s} \in \mathbb{C}^M$. Conventionally, signal and noise are assumed to be uncorrelated wide-sense stationary Gaussian random processes, with mean zero and covariance given by $\mathbb{E}\left[x(m)x(n)^H\right] = \delta_{m,n}\sigma_x^2$ and $\mathbb{E}\left[\mathbf{n}(m)\mathbf{n}(n)^H\right] = \delta_{m,n}\mathbf{R_n}$, respectively, where $\delta_{m,n}$ is the Kronecker delta function, σ_x^2 is the SOI power and $\mathbf{R_n}$ is the covariance matrix of the interference-plus-noise contribution. In particular, the observation $\mathbf{y}(n) \in \mathbb{C}^M$ may be modeling the matched filter output sufficient statistic for the received unknown symbols in, for instance, a multiuser detector, where the columns of $\mathbf{s}$ is the effective user signature associated with a certain desired user; an array processor, where $\mathbf{s}$ contains the angular frequency information (steering vectors) related to the intended source, represented by $x(n)$.

In the following, we concentrate on the problem of estimating the SOI waveform and assume that the SOI power is known (if the SOI power is unknown, an estimate of the value of σ_x^2 is customarily obtained from the data observations and used in conjuction with the SOI waveform). For the sake of ease of notation, and without loss of generality, we drop the constant σ_x^2 in the following exposition (i.e., $\sigma_x^2 = 1$). Consider the problem of estimating the signal waveform via a linear transformation of the received observations, i.e., $\hat{x}(n) = \mathbf{w}^H\mathbf{y}(n)$. The optimum MVDR/Capon filter can be obtained by solving the following linearly-constrained quadratic optimization problem, namely,

$$\mathbf{w}_{\mathsf{MVDR}} = \arg\min_{\mathbf{w}\in\mathbb{C}^M} \mathbf{w}^H\mathbf{R_n}\mathbf{w} \text{ subject to } \mathbf{w}^H\mathbf{s} = 1. \tag{1}$$

The solution to (1) can be easily obtained applying the method of Lagrange multipliers as

$$\mathbf{w}_{\mathsf{MVDR}} = \frac{\mathbf{R_n}^{-1}\mathbf{s}}{\mathbf{s}^H\mathbf{R_n}^{-1}\mathbf{s}} = \frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}}, \tag{2}$$

where $\mathbf{R} \in \mathbb{C}^{M\times M}$ is the covariance matrix of the observed process, i.e.,

$$\mathbf{R} = \mathbf{s}\mathbf{s}^H + \mathbf{R_n},$$

and the second equality in (2) can be checked using the matrix inversion lemma (in particular, the Sherman-Morrison-Woodbury identity for rank augmenting matrices).

Formally, this intuitive criterion may be formulated from a statistical estimation perspective as the problem of constructing a linear estimator minimizing the mean square-error (MSE), i.e., for $\hat{x} \equiv \hat{x}(n)$ and $x \equiv x(n)$,

$$\mathsf{MSE}(\hat{x}) = \mathbb{E}\left[|x-\hat{x}|^2\right] = \mathsf{var}(\hat{x}) + (\mathsf{bias}(\hat{x}))^2, \tag{3}$$

where the expectation is taken wrt. the random interference distribution, var stands for variance and $\mathsf{bias}(\hat{x}) = \mathbb{E}[\hat{x}] - x$. The problem of obtaining a linear transformation minimizing (3) under the unbiasedness constraint ($\mathsf{bias}(\hat{x}) = 0$) is equivalent to the optimization problem above, and the result is usually referred to as the BLUE. Interestingly enough, when the observations are Gaussian or have a linear model structure, the MVUE of the desired parameter turns out to be linear. Thus, as mentioned above, in these cases both BLUE and MVUE are equivalent.

The previous linear estimator of the signal waveform based on the filter in (2) assumes that the unknown quantities $x(n)$, $n = 1, 2, \ldots$, are fixed or deterministic. In case some information on the distribution of the unknown parameter is available for inference purposes, a probabilistic characterization of the parameter can be exploited by the Bayesian estimation framework in order to further reduce the MSE risk in (3). Indeed, under the previous statistical assumptions on the signal model, the filter minimizing (3) with the expectation being taken wrt. the joint distribution of both the signal waveform and the noise is the so-called linear minimum mean square-error (LMMSE) filter, namely,

$$\mathbf{w}_{\mathsf{LMMSE}} = \mathbf{R}^{-1}\mathbf{s}. \tag{4}$$

Note that the estimator $\hat{x}(n) = \mathbf{w}_{\mathsf{LMMSE}}^H\mathbf{y}(n)$ is no longer unbiased. In fact, the latter is an example of estimation method aiming at reducing the overall loss function by allowing for a degree of "biasedness" (see Introduction). In particular, the optimal bias-variance trade-off is fixed by the minimization of the Bayes MSE risk in (3), where the bias correction is determined as the solution of the following optimization problem:

$$\min_{\alpha}\left\{\mathsf{MSE}(\alpha\mathbf{w}_{\mathsf{MVDR}}) = \mathbb{E}\left[\left|x(n) - \alpha\mathbf{w}_{\mathsf{MVDR}}^H\mathbf{y}(n)\right|^2\right]\right\}, \tag{5}$$

where the effective linear transformation is now given by $\mathbf{w} \equiv \alpha\mathbf{w}_{\mathsf{MVDR}}$. Of course, we have $\alpha_{\mathsf{MVDR}} \equiv 1$. Moreover, the optimal α under the LMMSE criterion is clearly $\alpha_{\mathsf{LMMSE}} \equiv \mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}$, leading to the optimal filter in (4).

In general, for an arbitrary filter, observe that

$$\begin{aligned}\mathsf{MSE}(\mathbf{w}) &= 1 - 2\,\mathrm{Re}\left\{\mathbf{w}^H\mathbf{s}\right\} + \mathbf{w}^H\mathbf{R}\mathbf{w} \\ &= 1 - \mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} + \mathsf{MSE}_{\mathsf{excess}}(\mathbf{w}_{\mathsf{MMSE}}, \mathbf{w}),\end{aligned} \tag{6}$$

where

$$\mathsf{MSE}_{\mathsf{excess}}(\mathbf{w}_{\mathsf{MMSE}}, \mathbf{w}) = (\mathbf{w}_{\mathsf{MMSE}} - \mathbf{w})^H\mathbf{R}(\mathbf{w}_{\mathsf{MMSE}} - \mathbf{w}).$$

is the excess mean-square error. Note that minimizing $\mathsf{MSE}(\mathbf{w})$ is equivalent to minimizing the distance $\mathsf{MSE}_{\mathsf{excess}}(\mathbf{w}_{\mathsf{MMSE}}, \mathbf{w})$, and, from $\mathsf{MSE}_{\mathsf{excess}}(\mathbf{w}_{\mathsf{MMSE}}, \mathbf{w}_{\mathsf{MMSE}}) = 0$, the MMSE is given by

$$\mathsf{MSE}(\mathbf{w}_{\mathsf{MMSE}}) = 1 - \mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}. \tag{7}$$

An alternative performance measure particularly spread across the communications literature is the so-called signal-to-interference-plus-noise ratio (SINR), defined as

$$\mathsf{SINR}(\mathbf{w}) = \frac{|\mathbf{w}^H\mathbf{s}|^2}{\mathbf{w}^H\mathbf{R_n}\mathbf{w}} = \left(\frac{\mathbf{w}^H\mathbf{R}\mathbf{w}}{|\mathbf{w}^H\mathbf{s}|^2} - 1\right)^{-1}. \tag{8}$$

In principle, maximizing the SINR performance measure does not guarantee a good estimate of the signal waveform, which is the actual objective in this work (see discussion in [7] and also [8, Chapter 5]). However, any scaled version of the LMMSE filter in (4) maximizes the output SINR. Indeed, assuming the signal model is error-free and that exact knowledge of $\mathbf{R}$ is available, the LMMSE estimator is well-known to achieve the optimum performance trade-off in the sense of both minimizing the MSE as well as maximizing the output SINR, simultaneously, with a maximum SINR (MSINR) at the output of the MMSE filter being equal to

$$\mathsf{SINR}\left(\mathbf{w}_{\mathsf{MMSE}}\right)=\left(\frac{1}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}}-1\right)^{-1}=\mathbf{s}^H\mathbf{R}_{\mathbf{n}}^{-1}\mathbf{s}.$$

3. BIAS CORRECTIONS WITH SAMPLE FILTERS

In situations where the signal model is uncertain or the covariance matrix of the observations is unknown, only an unbalanced optimization of the previous two performance measures can be expected to be achieved by the LMMSE estimator. An analysis of this effect for the first source of mismatch is provided in [7] and also [8, Chapter 5]. Here, we concentrate on the case of estimation strategies based on the sample covariance matrix (SCM) estimator of the unknown observation covariance matrix, which is defined as

$$\hat{\mathbf{R}}=\frac{1}{N}\sum_{n=1}^{N}\mathbf{y}(n)\,\mathbf{y}^H(n). \tag{9}$$

More specifically, we focus on the generalized asymptotic regime used in [9] for the analysis of the performance of sample reduced-rank LMMSE estimators. Our approach here is based on an asymptotic interpretation of the bias factor in (5).

Consider the above linear estimation problem, where the unknown covariance matrix of the observations is replaced with a function of the SCM. Here, in order to allow for an invertible covariance matrix estimator even in the case that $M>N$, we generally consider a diagonal loading (DL) estimator of $\mathbf{R}$, namely based on a diagonally loaded SCM, i.e., $\left(\hat{\mathbf{R}}+\gamma\mathbf{I}_M\right)^{-1}$, for a given loading factor γ, and so the MVDR filter is implemented as

$$\hat{\mathbf{w}}_{\mathsf{DL-MVDR}}\equiv\frac{\left(\hat{\mathbf{R}}+\gamma\mathbf{I}_M\right)^{-1}\mathbf{s}}{\mathbf{s}^H\left(\hat{\mathbf{R}}+\gamma\mathbf{I}_M\right)^{-1}\mathbf{s}}.$$

In the case $\gamma=0$, we recover

$$\hat{\mathbf{w}}_{\mathsf{MVDR}}\equiv\frac{\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}}. \tag{10}$$

In the sequel, we will focus on the latter for the sake of notational simplicity and will omit the details on the former due to space limitations. Nevertheless, notice that all our derivations can be easily extended to the DL-base filter estimator, and final results for this case will be provided in the paper regarding the asymptotic performance analysis and the generalized consistent estimation.

Then, the sample bias correction factor is obtained from

$$\min_{\alpha}\left\{\mathsf{MSE}\left(\alpha\hat{\mathbf{w}}_{\mathsf{MVDR}}\right)=\mathbb{E}\left[\left|x(n)-\alpha\hat{\mathbf{w}}_{\mathsf{MVDR}}^H\mathbf{y}(n)\right|^2\right]\right\},$$

where, with some abuse of notation, here we have used $\mathbb{E}[\cdot]=\mathbb{E}\left[\cdot|\hat{\mathbf{R}}\right]$. The solution to the previous optimization problem can be straightforwardly shown to be

$$\hat{\alpha}_{\mathsf{LMMSE}}\equiv\frac{\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}}. \tag{11}$$

Consequently, if the covariance matrix is unknown, the MSE-optimal linear filter in the Bayesian estimation framework is $\hat{\mathbf{w}}_{\mathsf{LMMSE}}=\hat{\alpha}_{\mathsf{LMMSE}}\hat{\mathbf{w}}_{\mathsf{MVDR}}$. In particular, observe that

$$\mathsf{MSE}\left(\hat{\mathbf{w}}_{\mathsf{LMMSE}}\right)=1-\frac{\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}}. \tag{12}$$

Clearly, the SINR measure is invariant to scaling. However, by comparing (12) with the MSE achieved by the conventional implementation of the LMMSE filter, i.e.,

$$\mathsf{MSE}\left(\hat{\mathbf{R}}^{-1}\mathbf{s}\right)=1-2\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}+\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}, \tag{13}$$

a strictly smaller MMSE can be shown to be achieved by $\hat{\mathbf{w}}_{\mathsf{LMMSE}}$, since

$$\mathsf{MSE}\left(\hat{\mathbf{w}}_{\mathsf{LMMSE}}\right)<\mathsf{MSE}\left(\hat{\mathbf{R}}^{-1}\mathbf{s}\right).$$

To see this, we just need to show that

$$\frac{\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}}>2\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}-\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s},$$

but this readily follows by completing the square as

$$\begin{aligned}&\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2-2\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}+\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2\\&\qquad=\left(\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2-\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2\right)^2,\end{aligned}$$

and noting that the quadratic term in the RHS is strictly positive.

The quantity in (11) depends on both the SCM as well as the unknown theoretical covariance matrix. In practice, the implementation of $\hat{\alpha}_{\mathsf{LMMSE}}$ is conventionally based on replacing $\mathbf{R}$ with $\hat{\mathbf{R}}$. In this case, we clearly have $\hat{\alpha}_{\mathsf{LMMSE}}\left(\hat{\mathbf{R}}\right)=\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}$, which leads to the sample implementation of $\mathbf{w}_{\mathsf{MMSE}}$ obtained by directly replacing $\mathbf{R}$ with $\hat{\mathbf{R}}$. The theoretical foundations of the conventional implementation of $\hat{\alpha}_{\mathsf{LMMSE}}$ can be set as follows. The SCM is the minimum variance unbiased estimator of the theoretical covariance matrix, as well as the maximum likelihood estimator for Gaussian observations. For a sufficiently large number of samples N, implementations of the optimal scaling factor in (11) based on directly replacing the unknown theoretical covariance matrix with its sample estimate can be considered to provide a fairly accurate approximation. However, such an assumption does hardly hold under realistic scenario conditions given in a practical setting, where the number of samples per observation dimension is finite.

In this work, contrary to conventional practice, we consider an asymptotic approximation of the solution in (11), such that the ratio between the sample size and the observation dimension is fixed or constant. In particular, we deal here with the following general problem. Consider a bounded scalar function of the unknown theoretical covariance matrix $\mathbf{R}$, say $f(\mathbf{R})$, which represents a parameter that needs to be estimated. Under the practically more relevant assumption $M,N \to +\infty$ at a constant rate $c = M/N \in (0,+\infty)$, the conventional estimator given by $f\left(\hat{\mathbf{R}}\right)$ is not consistent, i.e., $f\left(\hat{\mathbf{R}}\right) \to g(\mathbf{R}) \neq f(\mathbf{R})$, as $N = N(M) \to \infty$ (in the sequel, we will only consider strong consistency and, therefore, almost sure stochastic convergence of the estimators). Motivated by this fact, we propose an estimator consisting of a certain function of the SCM, say $h\left(\hat{\mathbf{R}}\right)$, and which is consistent in the previous, practically more meaningful doubly-asymptotic regime, i.e., $h\left(\hat{\mathbf{R}}\right) \to f(\mathbf{R})$, with probability one as $N = N(M) \to \infty$.

The two previous ideas on the asymptotic convergence (or consistency) analysis of conventional SCM-based estimators, and the construction of improved consistent estimators based on the SCM and a fixed ratio c, are developed next. Before proceeding, we notice that the distribution of the random quantity in (11) is known in the literature (see [10]).

4. ASYMPTOTIC PERFORMANCE ANALYSIS

In this section, we apply recent results from RMT to derive asymptotic deterministic equivalents of the quantities $\mathsf{MSE}(\hat{\mathbf{w}}_{\mathsf{LMMSE}})$ in the limiting regime defined by $M,N \to +\infty$ with $c = M/N \in (0,+\infty)$. To that effect, it is enough to study the convergence of the quantities $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$ describing the the previous performance measure. Regarding the asymptotic convergence of these two quantities, we have the following ($a \asymp b$ means both quantities are asymptotic equivalents, i.e., $|a-b| \to 0$, and the convergence is with probability one):

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp (1-c)^{-1}\,\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}, \tag{14}$$

and

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp (1-c)^{-3}\,\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}. \tag{15}$$

For a proof of (14) and (15) we refer the reader to [11, Proposition 1] (see also [8, Chapter 4]) and [9, Theorem 1] (see also [12]), from where also the result for the more general case of a DL-based covariance matrix estimator can be obtained.

Finally, based on the previous results we notice that the asymptotic (deterministic) approximation of the MMSE achieved by the *proposed* bias correction is

$$\mathsf{MSE}(\hat{\mathbf{w}}_{\mathsf{LMMSE}}) = 1 - \vartheta_p(c)\,\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}, \tag{16}$$

where $\vartheta_p(c) \equiv (1-c)$. The previous measure is to be compared with the asymptotic limit of the MSE achieved by the *conventional* LMMSE filter implementation in (13), i.e.,

$$\begin{aligned}\mathsf{MSE}\left(\hat{\mathbf{R}}^{-1}\mathbf{s}\right) &= 1 - 2(1-c)^{-1}\,\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} + (1-c)^{-3}\,\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}\\ &= 1 - \vartheta_c(c)\,\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}, \end{aligned} \tag{17}$$

where $\vartheta_c(c) \equiv \left(2(1-c)^{-1} - (1-c)^{-3}\right)$. For a quantitative comparison between proposed and conventional approaches, please see the section Numerical Results.

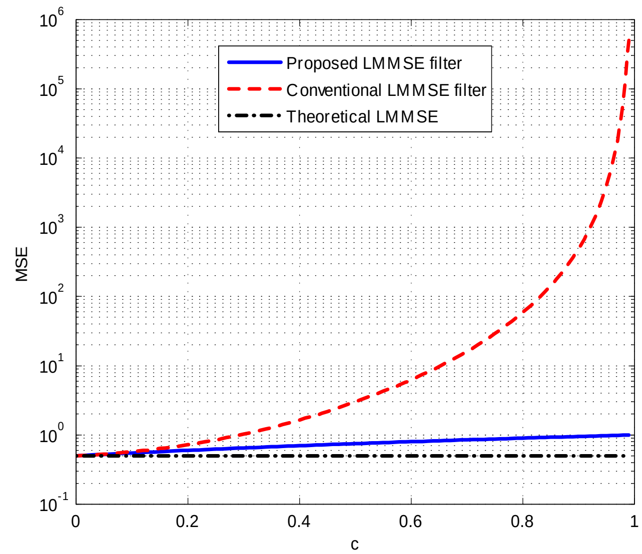

Figure 1: Asymptotic MSE achieved by the conventional ($\hat{\mathbf{R}}^{-1}\mathbf{s}$) and the proposed ($\hat{\alpha}_{\mathsf{LMMSE}}\hat{\mathbf{R}}^{-1}\mathbf{s}$) LMMSE filters, compared to theoretical MMSE.

5. CONSISTENT ESTIMATION OF OPTIMAL SCALING FACTOR

In order to implement the proposed filter, the bias correction in (11) given in terms of the unknown covariance matrix has to be estimated by using the available SCM. For that purpose, as introduced above, we find an expression only depending on the signature vector s and the matrix $\hat{\mathbf{R}}$ (both available for estimation purposes) that converges in our general asymptotic regime to the desired quantity. In particular, from the asymptotic analysis in the previous section, it suffices to find an estimator converging to

$$\frac{\left(\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right)^2}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}} \asymp (1-c)\,\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}.$$

From above, it straightforwardly find that

$$(1-c)^2\,\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp (1-c)\,\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s},$$

so that the proposed LMMSE filter implementation is $(1-c)^2\,\hat{\mathbf{R}}^{-1}\mathbf{s}$. Interestingly enough, observe that the latter corresponds to the classical implementation of the LMMSE filter up to the nontrivial scaling $(1-c)^2$. More importantly, we notice that the optimal scaling minimizing the MSE achieved by any practical implementation of the SCM-based LMMSE filter is *universal*, and so it does not depend on the theoretical covariance matrix $\mathbf{R}$, but only on the ratio c.

6. NUMERICAL RESULTS

In this section, the performance of the proposed filter is evaluated by means of numerical simulations. In Figure 1, the asymptotic approximation of the MSE achieved by both the

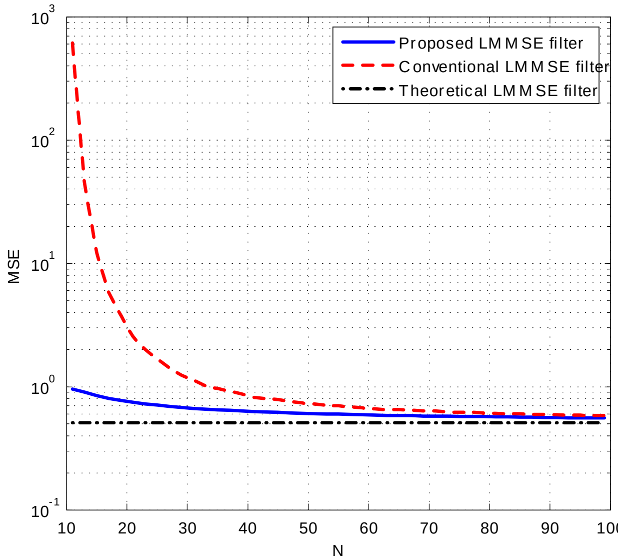

Figure 2: Finite sample-size MSE performance of the conventional ($\hat{\mathbf{R}}^{-1}\mathbf{s}$) and the proposed ($\hat{\alpha}_{\mathsf{LMMSE}}\hat{\mathbf{R}}^{-1}\mathbf{s}$) LMMSE filters, compared to theoretical MMSE ($M = 10$).

proposed and the conventional LMMSE filters, defined by $\hat{\alpha}_{\mathsf{LMMSE}}\hat{\mathbf{R}}^{-1}\mathbf{s}$ and $\hat{\mathbf{R}}^{-1}\mathbf{s}$, respectively, and given by (16) and (17) as a function of the ratio c is depicted along with the theoretical MMSE in (7). It can be observed that when $c = 0$ both filters converge to the MSE of the theoretical LMMSE filter, i.e., $\mathbf{R}^{-1}\mathbf{s}$. This corresponds to the case of classical asymptotics, so that the sample covariance matrix converges to the true covariance matrix. On the other hand, when the number of available samples N is decreased with respect to the dimension of the observations M, i.e., the value of the ratio c approaches 1, the performance of the proposed LMMSE filter dramatically outperforms that of the conventional one, as the former is optimized for any value of c whereas the latter is only optimal for $c = 0$.

In Figure 2, the performance of the proposed method in the case of a finite sample-size and in the context of an array signal processing application is considered. Specifically, the scenario consists of the SOI and four interferers in additive white Gaussian noise with variance normalized to 1. The powers of the SOI and all interferers are fixed to 1, and their direction-of-arrival is 0^o, 20^o, 30^o, 50^o and 60^o, respectively. The steering vectors are normalized such that $\|\mathbf{s}\|_2 = 1$. For simulation purposes, M was fixed to 10 and the MSE performance measure defined in (6) is shown for both the conventional and the proposed LMMSE filter implementations. The theoretical MMSE is also shown. Similar conclusions as for the previous simulation are observed, therefore validating the improved performance of the proposed bias correction in a realistic scenario compared to the classical approach.

7. CONCLUSIONS

The problem of finding the optimal scaling of a class of LMMSE filter implementations, based on a general SCM-based estimator of the actual covariance matrix, including diagonal loading, has been addressed. The proposed optimal scaling factor is a correction of the bias of the Bayesian MMSE estimator that optimizes the bias-variance tradesoff in an attempt to further reduce the overall MSE when the estimator is constructed based on the SCM.The scaling factor minimizing the MSE and depending of both the unknown covariance matrix and its sample estimator is firstly asymptotically analyzed in terms of key scenario parameters, and finally approximated via the SCM constructed using the set of available samples. To that effect, recent results from random matrix theory on the asymptotic analysis of large sample covariance matrices have been applied. As the main contribution of the paper, for the case of a non-loaded SCM estimator, a universal optimal scaling factor has been obtained which dramatically outperforms the conventional LMMSE filter implementation.

REFERENCES

[1] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

[2] C. Stein, “Inadmissibility of the usual estimator for the mean of a multivariate normal distribution,” in *Proc. of 3th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1956, pp. 197–206.

[3] W. James and C. Stein, “Estimation of quadratic loss,” in *Proc. of 4th Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1961, pp. 361–379.

[4] P. Stoica and R. Moses, “On biased estimators and the unbiased cramér-rao lower bound,” *Signal Processing Elsevier*, vol. 21, no. 4, pp. 349–350, October 1990.

[5] S. Kay and Y. C. Eldar, “Rethinking biased estimation,” *IEEE Signal Processing Magazine*, vol. 25, no. 3, pp. 133–136, May 2008.

[6] Y. Eldar, “Uniformly improving the cramér-rao bound and maximum-likelihood estimation,” *IEEE Transactions on Signal Processing*, vol. 54, no. 8, pp. 2943–2956, August 2006.

[7] Y. Eldar, A. Nehorai, and P. L. Rosa, “A competitive mean-squared error approach to beamforming,” *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5143–5154, November 2007.

[8] J. Li and P. Stoica, Eds., *Robust Adaptive Beamforming*. Wiley-Interscience, John Wiley and Sons, 2005.

[9] F. Rubio and X. Mestre, “Consistent reduced-rank lmmse estimation with a limited number of samples per observation dimension,” *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 2889–2902, Aug. 2009.

[10] I. S. Reed, J. Mallett, and L. Brennan, “Rapid convergence rate in adaptive arrays,” *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-10, no. 8, pp. 853Ű–863, November 1974.

[11] X. Mestre and M. Lagunas, “Finite sample size effect on MV beamformers: optimum diagonal loading factor for large arrays,” *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 69–82, January 2006.

[12] F. Rubio, “Generalized consistent estimation in arbitrarily high dimensional signal processing,” Ph.D. dissertation, Universitat Politècnica de Catalunya, Barcelona, April 2008.