Optimal Linear Shrinkage corrections of sample LMMSE and MVDR estimators

Author: Jorge Serra Puertas
Advisor: Montse Nájar

A master thesis
presented to the Universitat Politècnica de Catalunya
in fulfillment of the
master thesis requirement for the degree of
European Master of Research
on Information and
Communication Technologies (MERIT)

Barcelona, Spain, 2012

Abstract

A problem that appears in a vast number of signal processing applications is the estimation of an unknown parameter, observed through a linear model. The inference of this parameter is commonly based on a linear transformation of the available data, i.e. on a linear filtering. E.g. the aim of beamforming in array signal processing is to steer the beampattern of the antenna array towards a given direction to obtain the signal associated to a desired source. This is accomplished by means of a linear spatial filtering. In array processing another application is Direction of Arrival (DOA) estimation. Which usually implies exploring a range of directions and estimating the power or signal amplitude associated to each of them. Clearly to attain this aim a linear spatial filtering is needed as well. The design of the linear filters is based on the optimization of a measure of performance that in signal processing and in general in statistical inference is widely accepted to be the Mean Square Error (MSE). Thus, the optimal estimator is obtained by means of the optimization of the MSE and constrained to the available statistical information about the parameter of interest. This leads to obtain two notable estimators that will serve as a reference throughout this master thesis. On the one hand, when there is information about the first two moments of the parameter of interest one obtains the Linear Minimum Mean Square Error (LMMSE). On the other hand, when such statistcal information is not available one obtains the Capon or Minimum Variance Distortionless Response (MVDR) method, which obviously obtains worse performance than the LMMSE due to this lack of statistical information.

Altough the LMMSE and MVDR are the optimal methods, they are not realizable in general since they depend on the inverse of the correlation of the observations, which is not known. The common approach to circumvent this problem is to substitute it for the inverse of the sample correlation matrix (SCM), leading to the sample LMMSE and sample MVDR. This approach is optimal whenever the number of available realizations of the observed signal tends to infinity for a fixed observation dimension or at least when the number of samples is much greater than the observation dimension. Nonetheless, in a practical setting this large sample size regime scenario hardly holds and the sample methods undergo large performance degradation as the sample covariance is not a well conditioned estimator. The small sample size regime may be due to short stationarity constraints or due to a system with high observation dimension. These situations have appeared traditionally in applications such as radar or adaptive beamforming. Moreover, they will be also more and more met in the future generations of wireless communications such as beyond LTE systems where an increasing number of users demand to transmit and receive more and more information with hard timing constraints. Indeed, this demand will exceed the capacity of the current wireless systems and among the possible solutions

the research community is thinking of increasing the number of antenna array elements. Another scenario where one may find a small sample size regime is in green technologies, where one would desire to have estimators that require less and less number of available observations while maintaining a good performance.

Therefore, the aim of this master thesis is to propose corrections of the sample LMMSE and MVDR methods that allow to circumvent their performance degradation in small sample size regime and take profit of their optimality in the large sample size regime. To this end we are equipped with two powerful tools shrinkage estimation and random matrix theory. With shrinkage estimation we are defining the linear corrections of the sample LMMSE and MVDR methods and guarantee the robustness to the small sample size regime. As direct optimization of these shrinkage methods leads to unrealizable estimators then we propose to obtain a consistent estimate of these optimal shrinkage estimators within the general asymptotics regime where both the observation dimension and the sample size may grow, but at a fixed rate. I.e. we use random matrix theory to deal with either small, intermedium or large sample size regimes and to obtain a consistent estimate. Moreover, another advantage is that this approach based on shrinkage estimation and random matrix theory does not rely on any assumptions about the distribution of the observations. Furthermore, the numerical simulations highlight that the proposed methods outperform the sample LMMSE and MVDR methods in any of the sample size regimes considered herein and that the improvement in the small sample size situation is huge. Finally, for a particular case of shrinkage estimation of the sample LMMSE, we propose an alternative to random matrix theory based on multivariate analysis. Namely, based on the knowledge of the moments of an inverse Wishart distribution we obtain a shrinkage filter that optimizes the average MSE for Gaussian observations. The proposed approach dramatically outperforms the sample LMMSE in the small sample size regime. Indeed, it outperforms the sample LMMSE in any of the sample size regimes considered in this master thesis.

Dedication

Dedicated in loving memory to Josep Serra Vall and Pilar T. Rodríguez López.

Acknowledgements

I would like to thank all the people who made this Master Thesis possible.

Table of Contents

Notation

In general, uppercase boldface letters, $\mathbf{A}$, denote matrices, lowercase boldface letters, $\mathbf{a}$, denote column vectors and italics, a, denote scalars and generic random variables.

$\mathbf{A}^T, \mathbf{A}^*, \mathbf{A}^H$	Transpose, complex conjugate and Hermitian of a matrix $\mathbf{A}$, respectively.
$\mathbf{A}^{-1}$	Inverse of $\mathbf{A}$.
$\mathbf{A}^{1/2}$	Positive definite Hermitian square-root of $\mathbf{A}$, i.e. $\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{A}$.
$\mathrm{Tr}\left[\mathbf{A}\right]$	Trace of a matrix $\mathbf{A}$.
$\left\Vert\mathbf{A}\right\Vert_F$	Frobenius norm of a matrix $\mathbf{A}$, $\left\Vert\mathbf{A}\right\Vert_F = \left(\mathrm{Tr}\left[\mathbf{A}^H\mathbf{A}\right]\right)^{1/2}$.
$\left\Vert\mathbf{a}\right\Vert$, $\left\Vert\mathbf{a}\right\Vert_2$	Euclidean norm of a vector $\mathbf{a}$, $\left\Vert\mathbf{a}\right\Vert \triangleq \left\Vert\mathbf{a}\right\Vert_2 = \left(\mathbf{a}^H\mathbf{a}\right)^{1/2}$.
$\left[\mathbf{a}\right]_i, \mathbf{a}_i$	i-th entry of a vector $\mathbf{a}$.
$\left[\mathbf{A}\right]_{i,j}$	i,j-th entry of a matrix $\mathbf{A}$, corresponding to the i-th row and the j-th column.
$\left[\mathbf{A}\right]_{i,:}$	i-th row of a matrix $\mathbf{A}$.
$\left[\mathbf{A}\right]_{:,j}$	j-th column of a matrix $\mathbf{A}$.
$\mathbb{R}, \mathbb{C}, \mathbb{C}^+$	Denote, respectively, the set of real numbers, complex numbers and $\left\{z \in \mathbb{C} : \mathrm{Im}\left[z\right] > 0\right\}$.
$\mathbb{R}^M, \mathbb{C}^M$	The set of M-dimensional vectors with entries in $\mathbb{R}$ and $\mathbb{C}$, respectively.
$\mathbb{R}^{M\times N}, \mathbb{C}^{M\times N}$	The set of $M \times N$ matrices with real and complex valued entries, respectively.

$\mathbf{I}_M$	The $M \times M$ identity matrix.
$\mathbb{E}[\mathbf{A}]$	Expectation of a random matrix $\mathbf{A}$.
j	Imaginary unit, $j = \sqrt{-1}$.
$\#\{\cdot\}$	Cardinality of a set.
$\mathbf{a} \propto \mathbf{b}$	$\mathbf{a}$ is proportional to $\mathbf{b}$, i.e. $\mathbf{a} = \beta \mathbf{b}$ being β a given scalar.
$\rightarrow$	Convergence.
iid	A set of random variables are independent and identically distributed.
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.
$\mathcal{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Multivariate complex gaussian distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$.
$\|a\|$	Modulus of a complex number a.
$\mathcal{I}_{\Psi}(\omega)$	Indicator function. Suppose that Ω is a set with typical element ω and let Ψ be a subset of Ω. Then the indicator function of Ψ, denoted by $\mathcal{I}_{\Psi}(\omega)$, is defined as 1 if $\omega \in \Psi$ and 0 otherwise.
$\mathcal{CW}_M(N, \boldsymbol{\Sigma})$	Complex Wishart distribution with N degrees of freedom and scale parameter $\boldsymbol{\Sigma} \in \mathbb{C}^{M \times M}$.
$\mathcal{CW}_M^{-1}(N, \boldsymbol{\Sigma})$	Inverse complex Wishart distribution with N degrees of freedom and scale parameter $\boldsymbol{\Sigma} \in \mathbb{C}^{M \times M}$.
$a \asymp b$	a and b are asymptotically equivalent, i.e. $\|a - b\| \rightarrow 0$ almost surely.

Acronyms

AWGN	Additive White Gaussian Noise
BLUE	Best Linear Unbiased Estimator
DL	Diagonal Loading
DOA	Direction of Arrival
ESD	Empirical Spectrum Distribution
GSA	General Statistical Analysis
LASSO	Least Absolute Shrinkage and Selection Operator
LMMSE	Linear Minimum Mean Square Error
LS	Least Squares
MIMO	Multiple Input Multiple Output
ML	Maximum Likelihood
MMSE	Minimum Mean Square Error
MSE	Mean Square Error
MUSIC	MUltiple SIgnal Classification
MVDR	Minimum Variance Distortionless Response
RMT	Random Matrix Theory
SCM	Sample Correlation Matrix
SIR	Signal to Interference Ratio
SMI	Sample Matrix Inversion
SNR	Signal to Noise Ratio
ULA	Uniform Linear Array

Chapter 1

Introduction

1.1 Research Motivation and Contribution

The problem of linear estimation of an unknown parameter observed through a linear model is ubiquitous in signal processing. E.g. beamforming and Direction of Arrival (DOA) estimation in array signal processing [1] or spectral analysis [2]. Indeed this problem appears in many other fields of science and it can be traced back to the least-squares approach proposed by Gauss. In order to assess the performance of the designed estimator, the mean square error (MSE) is usually accepted as the common measure of performance [3]. As a consequence a myriad of estimators have been designed in the literature with the common aim of obtaining a good MSE performance. In this regard, assume that the first two moments of the parameter to estimate are available. Then, among the linear estimators, the one achieving the lowest MSE is the so-called Linear Minimum Mean Squared Error (LMMSE) estimator, [3]. Indeed, it is the minimum mean square error (MMSE) estimator when the joint distribution between the parameter to estimate and the observations is gaussian. When there is not a priori information about the first two moments of the parameter to estimate the conventional approach is as follows. As in this case the parameter to estimate is modeled as deterministic, direct optimization of the MSE leads to unrealizable methods, as they depend on the parameter to estimate. Then, the conventional approach is to impose an unbiasedness constraint on the MSE to avoid the dependence on the parameter of interest. This is equivalent to optimize the variance of the method and leads to the well known Best Linear Unbiassed Estimator (BLUE) [3], which in the signal processing literature coincides with the Capon method, also known as Minimum Variance Distortionless Response (MVDR) estimator [1]. Nonetheless, the price to pay for the lack of knowledge about the a priori information of the parameter to estimate, is that the performance of the MVDR is worse than the one of the LMMSE in terms of MSE. In order to improve the BLUE, attemps have been made to design methods that are biased but closer to the MSE than the BLUE. E.g. the Tikhonov regularizer [4] [5], the shrunken estimator [6], the covariance shaping least-squares estimator [7] and minimax MSE estimators [8–13].

According to the discussion in the last paragraph, LMMSE and MVDR are the optimal methods in linear estimation, depending on the available information about the parameter to be estimated. Unfortunately, they are not realizable in general, as they depend on the correlation of the noise plus interference terms, through the correlation of the observations, which is not known in most of practical applications. In order to circumvent this problem, the standard approach is based on a two stage procedure. First, the correlation of the observations $\mathbf{R}$ is estimated by means of the sample correlation matrix (SCM) $\hat{\mathbf{R}}$. Second, the true unknown correlation of the observations is substituted for the SCM in the expressions of the LMMSE or the MVDR methods. In the literature dealing with the MVDR implementation this technique is also known as sample matrix inversion (SMI) technique [1]. The underlying rationale is based on the optimal properties of the SCM. Namely, for Gaussian observations it is the maximum likelihood (ML) estimator of the true correlation matrix, see [14, Theorem 4.1] and as a consequence the MVUE of $\mathbf{R}$ for a sufficiently large number of samples N compared to the observation dimension M.

Nevertheless, in practical scenario conditions the assumption that the sample size is large compared to the observation dimension, i.e. $N >> M$, does hardly hold. In fact N may be comparable to M or even lower, leading to the so-called small sample size regime. E.g., in adaptive beamforming and DOA estimation [15] [16], where the dimension of an antenna array may be comparable to the number of available snapshots or even larger; this is due to short stationarity properties of the observed signal or due to a large array of antennas. This paradigm is also found within the context of beyond 4G wireless communications networks. E.g. in [17] a large array of antennas is considered in the context of multiuser MIMO scenarios. Also in [18], in the context of cognitive wireless networks, the authors consider that the number of available samples may be comparable to the number of receiving antennas when estimating the energy of multiple sources. One of the reasons is that the processing of dynamic information in secondary networks must be as fast as possible to avoid disruption in the primary networks.

Unfortunately, when the sample size is comparable to the observation dimension, the traditional implementation of the optimal estimators based on the sample estimate of $\mathbf{R}$ leads to a severe performance degradation, see [1], [15], [19], [20] and references therein. In fact, they may display worse performance than the Bartlett or phased array estimator which is a naive strategy based on directly replacing the theoretical covariance matrix by a scaled identity matrix, i.e. $\mathbf{w} \propto \mathbf{s}$. The reason for this performance degradation may be explained as follows. The sample LMMSE and MVDR rely on directly substituting $\mathbf{R}^{-1}$ for the inverse of the SCM in the expressions of the LMMSE and MVDR, respectively. Nonetheless, the SCM is not a well conditioned estimator, i.e. in the small sample size regime inverting the SCM severely amplifies the estimation error. Thus, in small sample size situations the sample methods are no longer optimal and require a calibration that counteract their severe performance degradation.

In order to circumvent this problem several approaches have been suggested in the literature to improve the estimation of $\mathbf{R}$. Those methods are mainly based on a regularization technique called diagonal loading (DL), consisting on adding a positive real number α to the diagonal entries of the SCM, i.e. $\check{\mathbf{R}} = \hat{\mathbf{R}} + \alpha\mathbf{I}$, see [15], [21], [22] and [23]. However, the correct choice of the loading factor α is controversial and usually rather ad hoc methods depending on the practical setting of the application are used. An exception is the work of Mestre [19], that based on random matrix theory (RMT) founds a DL factor for the MVDR beamformer that optimizes the SNR. In fact, DL based techniques can be regarded as a particular case of linear shrinkage estimators that aim to reduce the estimation error based on shrinking the SCM towards a constant estimator $\mathbf{R}_0$. Namely, the shrinkage estimator, $\check{\mathbf{R}}$, of the correlation of the observations, $\mathbf{R}$, reads $\check{\mathbf{R}} = \alpha_1\hat{\mathbf{R}} + \alpha_2\mathbf{R}_0$, see [24] or [25], and $\mathbf{R}_0$ may be obtained from a priori knowledge stemming from the problem at hand. The idea of those estimators is to blend the SCM, whose estimation error mostly comes from an estimation variance, with an estimator diplaying certain amount of bias but zero variance, that is a constant estimator related to a priori information about the parameter to estimate. This yields to a gain in estimation variance that more than compensates the increase in bias and thus the overall error is diminished. However, those methods aim to improve directly the estimation of the covariance matrix of the observations, which is not the final target herein. Moreover, there is another important drawback related to the sample estimators. Namely, considering the practical more relevant assumption where $M, N \to \infty$ at a constant rate $M/N \to c \in (0, 1)$, they lead in general to estimators that are not consistent as it can be derived from results in [15, Chapter4], [26], [20] or [19] and also as it is shown herein.

Therefore, the aim throughout all this master thesis is to design new estimators that overcome the limitations of the conventional sample LMMSE and MVDR methods. These are the performance degradation in the small sample size regime, and in the case of the former method, the lack of consistency within the doubly asymptotic regime where $M, N \to \infty$ at a constant rate $M/N \to c \in (0, 1)$. As it will be seen in numerical simulations of the upcoming chapters, the proposed estimators not only achieve this aim but also dominate the conventional sample LMMSE and MVDR methods. Where by definition, an estimator $\hat{x}$, of a parameter x, is said to dominate another estimator $\check{x}$ on a set S if its MSE is smaller at least for some values of x in S and it is never larger for any value of x in S [27], [28]. The design of the proposed estimators will be based in three tools shrinkage estimation, recent results from random matrix theory and some results within the field of multivariate statistical analysis.

Shrinkage estimation is considered herein to define the structure of the estimators to be designed, as they are known to be robust to the small sample size regime and they achieve in general a lower estimation error than the sample estimators, see [28]. The inception of those methods can be traced back to the works of Stein [29], [30] and later on of Brown [31], see [28] for a thorough discussion on this topic. The rationale behind shrinkage estimation can be summarized as follows. In general in estimation theory, a large amount of methods rely on a function of the sample moments, e.g. on a function of the sample mean or the sample covariance. The foundations of that approach are based on the Glivenko-Cantelli theorem that states that for a set of iid random variables, the empirical distribution tends to the true distribution for a large number of observations [32]. Thus, as the sample estimators are moments of the empirical distribution, when the sample size is large they tend to the moments of the true unknown distribution and as a consequence they are optimal. Therefore, when the number of samples is low, a large degradation of the sample based estimators may be expected. Namely the error comes from the estimation variance as these methods are unbiased in general. Therefore, the idea of shrinkage estimation is to diminish the estimation variance by introducing a bias such that the overall estimation error is lower than that of the sample estimators. This can be accomplished by a linear scaling of the sample methods. Or more in general by blending by means of a weighted average the sample estimators, i.e. $\mathbf{w}_t \propto \hat{\mathbf{R}}^{-1}\mathbf{s}$, that display large variance but no bias, with constant estimators, obtained from a priori knowledge stemming from the problem at hand, which display zero variance but large bias, e.g. the Bartlett filter $\mathbf{w}_b = \mathbf{s}$. I.e. the general class of shrinkage estimators $\hat{x} = \mathbf{w}^H\mathbf{y}$ such that $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2\mathbf{s}$ is considered in this master thesis. This approach, leads to a gain in variance that more than compensates for the increase in bias an as a consequence the overall MSE decreases.

Nonetheless, direct optimization of the MSE, when plugging the shrinkage estimators into it, leads to optimal weigths α_1, α_2 dependent on the true but unknown correlation $\mathbf{R}$, i.e. to unrealizable methods. In [33] in an analogous problem, within the context of portfolio optimization in finance, they propose to substitute $\mathbf{R}$ for its sample estimate in the optimal weigths α_1, α_2, but this approach is suboptimal as an estimation risk is implicit. Indeed, this approach leads to obtain the sample MVDR and LMMSE methods again and the positive effects of shrinkage estimation vanish.

This gives rise to the second and most important tool considered in this thesis to design the estimators, Random Matrix Theory (RMT), namely large dimensional random matrix theory. This field originated from the works of Wigner [34] and subsequently by Marčenko and Pastur [35]. In this context it is also worth mentioning the general statistical analysis introduced by Girko in [36] and [37]. More recently, RMT has been successfully applied in signal processing and information theory emerging as a promising new paradigm, see [38], [39], [40] and references therein. Herein, some recent results from RMT, derived in [19], [15, Chapter 4], [38] and [20, Theorem 1], are considered. They pave the way to obtain consistent estimates of the optimal weigths α_1, α_2 of the shrinkage estimators, under the general and more practical asymptotics where M , $N \to \infty$ at a constant rate $M/N \to c \in (0,1)$. I.e. the RMT based approach introduced herein leads to obtain shrinkage estimators that are realizable, consistent, robust to the small sample size regime and that outperform the traditional implementations of the LMMSE and MVDR. Moreover, another advantage of this RMT approach is that it does not rely on any assumption about the distribution of the observations.

Finally, for the particular case where the observations are Gaussian distributed and the LMMSE shrinkage estimator is of the form $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$, an alternative to RMT, based on multivariate statistical analysis, is presented. Namely, as $\mathbf{w}$ depends on $\hat{\mathbf{R}}$, the MSE is a conditional expectation, relying on the knowledge of $\hat{\mathbf{R}}$, and as a consequence a random quantity. I.e. for each possible value of $\hat{\mathbf{R}}$ a given MSE is obtained. Thus, in this case the proposed approach is based on obtaining the shrinkage factor α which minimizes the average MSE. Provided that the observed data be Gaussian, the solution is obtained by using the knowledge of the summary statistics of a complex inverse Wishart distribution. Unlike the RMT approach, which obtains an asymptotically optimal solution, this design not only obtains a consistent estimator, but also an optimal solution for the finite regime.

1.2 Signal Model

Next, we present the general model of the observed data that will be considered throughout all the master thesis to design the proposed estimators of the unknown parameter $x(n) \in \mathbb{C}$. Namely, let $x(n)$ be observed through the stochastic process $\mathbf{y}(n) \in \mathbb{C}^M$ by means of the next affine transformation,

$$\mathbf{y}(n) = x(n)\mathbf{s} + \mathbf{n}(n), \quad 1 \leq n \leq N \tag{1.1}$$

Where $\mathbf{s} \in \mathbb{C}^M$ is a known deterministic vector, $\mathbf{n}(n) \in \mathbb{C}^M$ is a stochastic process and N is the number of available measures. E.g. in the context of array signal processing $\mathbf{y}(n)$ is the output of an antenna array, $\mathbf{s}$ is the steering vector and $\mathbf{n}(n)$ contains the noise plus interference signals [1]. The next model assumptions are supposed to hold for any of the designed estimators throughout all this master thesis,

(a) *$x(n)$ and $\mathbf{n}(n)$ are uncorrelated. Moreover, $\mathbb{E}[\mathbf{n}(n)] = \mathbf{0}$ and $\mathbb{E}\left[\mathbf{n}(n)\mathbf{n}(n)^H\right] = \mathbf{R_n}$.*
(b) *As a consequence of (a) $\mathbf{R} \triangleq \mathbb{E}\left[\mathbf{y}(n)\mathbf{y}(n)^H\right] = \gamma \mathbf{s}\mathbf{s}^H + \mathbf{R_n}$; $\gamma \triangleq \mathbb{E}\left[|x(n)|^2\right]$* and $\|\mathbf{s}\|^2 = 1$.
(c) *The set of observations $\{\mathbf{y}(n)\}_{n=1}^N$ are iid.*
(d) *The number of samples is higher than the observation dimension, i.e. $N > M$ or in other words the ratio between the observation dimension and the sample size fulfills $M/N \in (0,1)$.*

Moreover, for the estimators in chapters 3 and 5 the next assumption is also needed,

(e) $\gamma \triangleq \mathbb{E}\left[|x(n)|^2\right]$ is known.

Finally, a part from assumptions *(a)-(e)* the next assumption is also assumed to hold for the estimator designed in chapter 5,

(f) *The set of observations $\{\mathbf{y}(n)\}_{n=1}^N$ is distributed according to a complex gaussian distribution. Namely, $\mathbf{y}(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R})$.*

1.3 Optimal Linear Estimators

In this master thesis, the family of estimators of $x(n)$ based on a linear transformation or linear filtering of $\mathbf{y}(n)$ is considered. Namely, denoting by $\hat{x}(n)$ the estimation of $x(n)$ and $\mathbf{w}$ the linear filter, these estimators read,

$$\hat{x}(n) = \mathbf{w}^H \mathbf{y}(n) \tag{1.2}$$

Moreover, throughout this thesis, in order to establish a criterion of optimality among the estimators, the MSE is considered as a measure of performance. In this regard, consider the optimization of the MSE for the family of linear estimators in (1.2), i.e.,

$$\mathbf{w}_{opt} = \arg\min_{\mathbf{w}} \mathrm{MSE}(\mathbf{w}) \triangleq \arg\min_{\mathbf{w}} \mathbb{E}\left[\left|x(n) - \mathbf{w}^H\mathbf{y}(n)\right|^2\right] \tag{1.3}$$

furthermore, let assume the data model in (1.1) with assumptions *(a)-(e)* for the observed signal $\mathbf{y}(n)$. Then one obtains the well known LMMSE method [3],

$$\hat{x}_l(n) = \mathbf{w}_l^H \mathbf{y}(n);\ \mathbf{w}_l{=}\gamma \mathbf{R}^{-1}\mathbf{s} \tag{1.4}$$

Being $\gamma \triangleq \mathbb{E}\left[|x(n)|^2\right]$ the second raw moment of the signal to be estimated and $\mathbf{R} \triangleq \mathbb{E}\left[\mathbf{y}(n)\mathbf{y}(n)^H\right]$ the correlation of the observed signal $\mathbf{y}(n)$. LMMSE possesses important optimality features. Namely, first it is the method that achieves the lowest MSE among the set of linear estimators. Second, it is the minimum MSE estimator when the joint distribution between $x(n)$ and $\mathbf{y}(n)$ is gaussian. Nevertheless, expression (1.4) highlights that LMMSE assumes implicitly some a priori knowledge about the second moment of $x(n)$. Therefore, in the applications where such an information is not available the LMMSE estimator is not realizable.

In order to circumvent the lack of knowledge about γ the most popular approach is the Capon method, also known as MVDR in the array signal processing literature, see e.g. [1] or [41]. The rationale behind this methods is as follows. After some manipulations, the expression of the MSE in (1.3) reads

$$\mathrm{MSE}\left(\mathbf{w}\right) = \mathbf{w}^H \mathbf{R}_\mathbf{n} \mathbf{w} {+} \gamma \left|1 - \mathbf{w}^H \mathbf{s}\right|^2 \tag{1.5}$$

Therefore, imposing the constraint $\mathbf{w}^H\mathbf{s} = 1$ avoids the dependence of the cost function on the unknown quantity γ. Indeed, when $x(n)$ is modeled as a deterministic parameter $\mathbf{w}^H\mathbf{s} = 1$ is actually an unbiasedness constraint in the MSE optimization. Thus, the MVDR estimator arises from the next optimization problem

$$\begin{aligned} \mathbf{w}_c = \arg\min_{\mathbf{w}}\ & \mathbf{w}^H \mathbf{R}_\mathbf{n} \mathbf{w} \\ s.t.\ & \mathbf{w}^H \mathbf{s} = 1 \end{aligned} \tag{1.6}$$

Now, observe that under assumptions *(a)-(e)* exposed in (1.1) the optimization in (1.6) is equivalent to the one where $\mathbf{R}_\mathbf{n}$ is replaced by $\mathbf{R}$. Thus, applying the method of Lagrange multipliers to (1.6), it is easy to obtain the well known expression for the MVDR estimator, see [1],

$$\hat{x}_c(n) = \mathbf{w}_c^H \mathbf{y}(n);\ \mathbf{w}_c{=}\frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H \mathbf{R}^{-1}\mathbf{s}} \tag{1.7}$$

At this point, it is worth mentioning that when $x(n)$ is modeled as a deterministic parameter, the MVDR method is the Best Linear Unbiased Estimator (BLUE) and if the noise is Gaussian distributed it is the Minimum Variance Unbiased Estimator (MVUE), see e.g. [3].

Therefore, MVDR and LMMSE methods are the optimal linear estimators in terms of MSE depending on whether the unbiasedness constraint $\mathbf{w}^H\mathbf{s}$ =1 is applied or not, respectively. Nonetheless, in practice they are not realizable since they depend on the correlation of the observations $\mathbf{R}$ which on its turn depends on the unknown noise covariance $\mathbf{R_n}$. In order to circumvent this problem, the traditional approach is based on a two step strategy. First, given the set of N available observations $\{\mathbf{y}(n)\}_{n=1}^{N}$, the unknown $\mathbf{R}$ is estimated by means of the SCM, $\hat{\mathbf{R}}$.

$$\hat{\mathbf{R}} \triangleq \frac{1}{N}\sum_{n=0}^{N-1}\mathbf{y}(n)\mathbf{y}^H(n) \tag{1.8}$$

Second, the SCM is substituted in the theoretical expressions of the LMMSE and Capon methods, (1.4) and (1.7) respectively. This yields the traditional sample implementations of the LMMSE and MVDR estimators,

$$\begin{aligned}\hat{x}_{l,t}(n) &= \hat{\mathbf{w}}_l^H\mathbf{y}(n);\ \hat{\mathbf{w}}_l{=}\gamma\hat{\mathbf{R}}^{-1}\mathbf{s}\\ \hat{x}_{c,t}(n) &= \hat{\mathbf{w}}_c^H\mathbf{y}(n);\ \hat{\mathbf{w}}_c{=}\tfrac{\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}}\end{aligned} \tag{1.9}$$

This strategy relies on the optimal properties of the SCM. Namely, for Gaussian observations, $\hat{\mathbf{R}}$ is the ML estimator of $\mathbf{R}$ and also its MVUE for a sufficiently large number of samples N compared to the observation dimension M, [14, Theorem 4.1]. Indeed, considering the asymptotic regime where M is fixed and $N \to \infty$ the sample estimators are consistent, i.e. $\hat{\mathbf{w}}_l{=}\gamma\hat{\mathbf{R}}^{-1}\mathbf{s} \to \mathbf{w}_l{=}\gamma\mathbf{R}^{-1}\mathbf{s}$ and $\hat{\mathbf{w}}_c{=}\frac{\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}} \to \mathbf{w}_c{=}\frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}}$. Unfortunately, in practice N may be comparable to M. In these situations, the SCM is no longer a good estimate. This problem is exacerbated by the inverse involved in the LMMSE and Capon methods and leads to a large performance degradation of the sample based implementations. Indeed, in the lower sample size regime where $M \approx N$ the traditional implementations may display worse performance than a naive strategy such as the Bartlett filter $\mathbf{w} = \mathbf{s}$ that does not take profit of the available measured information, as it is based on substituting the unknown $\mathbf{R}$ for the identity matrix in the optimal filters. Moreover, considering the more practical general asymptotics where $M, N \to \infty$ at a constant rate $M/N \to c \in (0, 1)$, the sample LMMSE is not consistent. This is shown below in chapter 3, alternatively it can be derived from the results in [19, Appendix I], (cf. [15, Chapter4], [20] and [26]).

1.4 Problem Statement

The aim throughout all this master thesis is to design estimators that take as a reference the sample LMMSE and MVDR estimators but that overcome their drawbacks. Namely their performance degradation in the small sample size regime and the lack of consistency in the case of the sample LMMSE, as it was commented in the previous subsection. The rationale to take the sample estimators as a reference is that for large sample size regimes, $N \gg M$, they are optimal as they tend to the theoretical expressions in (1.4), (1.7). In order to deal with the performance degradation in the small sample size regime, the use of shrinkage estimators is proposed herein, as they are known to be robust to those situations [28] and they permit to preserve the optimal properties of the sample LMMSE and MVDR for large sample size. I.e. a possible choice for the structure of the proposed estimators is $\hat{x}(n) = \mathbf{w}^H \mathbf{y}(n);\ \mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$. Note that the shrinkage factor α introduces a bias in the estimation with the aim of reducing the overall estimation error. Moreover, intuitively for large sample size regime α would lead to obtain the sample LMMSE and MVDR and for small sample size regime would produce a linear correction of these sample methods. Indeed, a generalization of these shrinkage estimators arises considering the fact that for $M \approx N$, the sample LMMSE and MVDR display in general worse performance than a naive Bartlett filter, i.e. $\mathbf{w} = \mathbf{s}$, see [1]. Therefore, in this master thesis we will consider the next general class of shrinkage estimators that combine the sample estimators given by $\mathbf{w} \propto \hat{\mathbf{R}}^{-1}\mathbf{s}$, which are known to be optimal for $N \gg M$ with a Bartlett filter $\mathbf{w} = \mathbf{s}$, which gives better performance than the sample methods when $M \approx N$.

$$\hat{x}_s(n) = \mathbf{w}_s^H \mathbf{y}(n);\ \mathbf{w}_s = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s} \tag{1.10}$$

Being α_1 and α_2 the shrinkage weights to be determined. It is worth remarking that the proposed estimator incorporates the benefits of shrinkage estimation. On the one hand, when using the sample based implementations of the optimal methods, i.e. $\mathbf{w} \propto \hat{\mathbf{R}}^{-1}\mathbf{s}$, the estimation error of the filters mostly comes from the estimation variance. On the other hand, the Bartlett filter has zero variance but a large bias in the estimation of the optimal $\mathbf{w}$. Therefore, by blending the sample based LMMSE or MVDR with the Bartlett filter, by means of weighted averages, we are introducing some bias in the estimation of the optimal filters, but the gain in variance more than compensates for the increase in bias an as a consequence the overall MSE decreases.

At this point, recall that the performance criterion of the designed estimators is the MSE. Therefore, herein we propose to design the shrinkage estimators in (1.10) by minimizing the MSE or more in general a functional of the MSE, $f(\mathrm{MSE}(\mathbf{w}))$. This leads to the problem formulation that will be tackled in this thesis.

Problem statement:

Assume that a set of observations $\{\mathbf{y}(n)\}_{n=1}^{N}$ fulfilling the model in (1.1) is available. Then, obtain an estimation of the unknown parameter $x(n)$ in (1.1) that minimize a functional $f(\mathrm{MSE}(\mathbf{w}))$ of the MSE in (1.5) when the estimation is based on the shrinkage filtering in (1.10) and subject to a set of constraints $\mathcal{C}$ on $\mathbf{w}$. This problem is mathematically formulated as follows,

$$\hat{x}_s(n) = \mathbf{w}_s^H \mathbf{y}(n); \qquad \begin{array}{c} \mathbf{w}_s = \arg\min_{\mathbf{w}} f(\mathrm{MSE}(\mathbf{w})) \\ s.t.\ \mathbf{w} \in \mathcal{C}, \mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s} \end{array} \tag{1.11}$$

Where, according to (1.5), $\mathrm{MSE}(\mathbf{w}) = \mathbf{w}^H \mathbf{R_n} \mathbf{w} + \gamma \left|1 - \mathbf{w}^H \mathbf{s}\right|^2$.

Remark 1: This problem formulation embraces both the case where knowledge about the second moment of $x(n)$ is available, i.e. a shrinkage of the sample LMMSE filter, and the case where it is not, i.e. a shrinkage of the sample MVDR filter. In the former case there are not constraints $\mathcal{C}$, whereas in the latter $\mathbf{w} \in \mathcal{C} \Leftrightarrow \mathbf{w}^H\mathbf{s} = 1$.

Remark 2: With regard to $f(\mathrm{MSE}(\mathbf{w}))$, in chapters 3 and 4 $f(\mathrm{MSE}(\mathbf{w})) = \mathrm{MSE}(\mathbf{w})$. Moreover, in these chapters, as the optimization of $\mathrm{MSE}(\mathbf{w})$ yields unrealizable methods, subsequently RMT tools are used to obtain (M, N)-consistent estimates of the optimal methods. On the other hand, in chapter 5 an alternative to the RMT approach is proposed. Namely, for the particular case where $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s}$, provided that there are not constraints $\mathcal{C}$ and provided that the observations be gaussian, we consider the average MSE to design the proposed method, i.e. $f(\mathrm{MSE}(\mathbf{w})) = \mathbb{E}_{\hat{\mathbf{R}}}\left[\mathbb{E}_{x,\mathbf{n}}\left[\left|x(n) - \mathbf{w}^H\mathbf{y}(n)\right|^2 \mid \hat{\mathbf{R}}\right]\right] \triangleq \mathbb{E}_{\hat{\mathbf{R}}}[\mathrm{MSE}(\mathbf{w})]$.

1.5 Organization of the Master Thesis

The organization of the master thesis is as follows. In chapter 2, a brief introduction to the main tools used herein, i.e. RMT and shrinkage estimation, is presented. Moreover in this chapter certain RMT results are presented, they are the cornerstone to derive the proposed shrinkage LMMSE and MVDR filters. Chapter 3 proposes optimal shrinkage corrections for large sample LMMSE filters, based on RMT. The work of this chapter is in part available in the next conference and journal papers,

- J. Serra and F. Rubio, "Bias Corrections in Linear MMSE Estimation with Large Filters," *in Proceedings of the European Signal Processing Conference (EUSIPCO 2010)*, 23-27 August 2010, Aalborg (Denmark).

- J. Serra and M. Nájar,"Optimal Linear Shrinkage of large sample LMMSE and MVDR filters," (submitted to) IEEE Transactions on Signal Processing.

Chapter 4 proposes an optimal shrinkage correction for large sample MVDR filters, which is based on RMT. The material of this chapter has been included in the next journal paper, which is now under review,

- J. Serra and M. Nájar,"Optimal Linear Shrinkage of large sample LMMSE and MVDR filters," (submitted to) IEEE Transactions on Signal Processing.

Next, in chapter 5 an alternative to the RMT approach is presented. Namely, for the particular case where the observed data is Gaussian distributed, we present a shrinkage of the sample LMMSE that optimizes the average MSE by using summary statistics of the complex inverse Wishart distribution. The method proposed in this chapter appears in the next conference paper,

- J. Serra and M. Nájar,"Optimal Linear Correction in LMMSE Estimation Using Moments of the Complex Inverse Wishart Distribution," *in Proc. IEEE Statistical Signal Processing Conference (SSP 2012)*, 5-8 August 2012, Ann Arbor, MI, (USA).

Finally, chapter 6 presents the concluding remarks and future topics of research.

Chapter 2

Technical Background

2.1 Introduction

In this chapter, a review of the techniques used to design the proposed estimators in the upcoming chapters is exposed. Namely, these are shrinkage estimation and random matrix theory. Both tools cover extensive topics and not only have been applied to signal processing and wireless communications, but also to other fields of science, as it will be commented below in this chapter. Therefore, herein the focus will be put on exposing the most important features of these techniques in the context of this master thesis. Namely, a historical background will be given and certain definitions and propositions motivating the use of these tools for the design of the proposed estimators will be dealt with. Moreover, several results that are the cornerstone for design of the proposed estimators are presented. The organization of this chapter is as follows. Section 2.2 deals with the RMT tool and section 2.3 exposes the theory of shrinkage estimation.

2.2 Random Matrix Theory

2.2.1 Historical Background

The theory of random matrices is a vast field that studies the properties of matrices whose entries follow a given joint probability distribution. Namely, within this field different topics are addressed or have been addressed. These are the study of small size matrices with joint Gaussian entries, e.g. see [42], [43] and [44]; the study of small and large random matrices with invariance properties (e.g. free probability theory [45] [46], combinatronics [47] [48] and Gaussian methods); And finally the study of large random matrices with independent entries [35], [49] and [50].

The study of random matrices may be traced back within the mathematical field of multivariate statistical analysis. Namely, due to the work that J. Wishart conducted on fixed-size matrices with Gaussian entries in [42]. Nonetheless, the seed that subsequently produced a plethora of research in random matrix theory stems from problems that appeared within the field of nuclear physics in 1950s. Namely, in quantum mechanics, the quantum energy levels are not observable, but may be characterized through the eigenvalues of a matrix of observations. It turns out that the empirical distribution of the eigenvalues (ESD) has a very complicated form when the dimension of the matrix is high. Nonetheless, it was observed, by means of numerical simulations that the ESD tends to a non-random limit when the dimensions of the matrix increase without bound. Anyway these were conjectures, based on given observations, and it was not until 1958 that E. Wigner, with his pathbreaking publication [34], showed that the empirical distribution of the eigenvalues of a large random matrix, called Wigner matrix nowadays, tends to a semi-circle. With this work he founded the field of random matrix theory, which deals with the asymptotic study of eigenvalues of random matrices. Subsequently, another publication that was of paramount importance for the development of the theory of large dimensional random matrices, was presented by Marčenko and Pastur in 1967, [35]. Since then, a plethora of research have been conducted by researchers such as Bai or Silverstein, see [51] and references therein. For this master thesis purposes, it is also worth mentioning the work of Girko, as he developed a new statistical inference framework, known as G-estimation, which is based on random matrix theory and complex integration, see e.g. [36] or [37].

Random matrix theory has found applications in fields as diverse as nuclear physics [52], mathematical finance [53] or computational biology [54]. In wireless communications and information theory, random matrix theory has found several applications, see [40], [55] and references therein for a thorough description. E.g. it has been used in the analysis of the capacity of MIMO systems, see [39], [56] and also in the energy estimation of multiple sources in cognitive wireless networks [18]. In signal processing, RMT has been applied to study the performance of subspace based methods and to propose new subspace algorithms such as the G-MUSIC, see [16]. Also to analyze the performance of the sample estimates of eigenvalues and eigenvectors of covariance matrices [38], and to propose new estimators of them that cope with the performance degradation in the small sample size regime, see [50], [19].

2.2.2 Large Dimensional Random Matrix Theory

The Stieljes Transform

In this master thesis, recent results from large dimensional random matrix theory will be used to derive the proposed estimators. This discipline is devoted to study the asymptotic distribution of eigenvalues and eigenvectors of random matrices, whose dimensions grow without bound. Namely, let define the empirical distribution function of the eigenvalues λ of a given random matrix $\mathbf{X} \in \mathbb{C}^{M \times M}$ as,

$$F^{\mathbf{X}}(\lambda) = \frac{1}{M}\#\{\lambda_m \leq \lambda; m = 1, \ldots, M\} = \frac{1}{M}\sum_{m=1}^{M} \mathcal{I}_{\lambda_m \leq \lambda}(\lambda) \tag{2.1}$$

Where # denotes the cardinality of a set and $\mathcal{I}$ denotes the indicator function. Then random matrix theory studies the convergence of $F^{\mathbf{X}}(\lambda)$ towards a limiting probability distribution function, when the dimensions of $\mathbf{X}$ increase without bound. In this regard, a tool of paramount importance is the Stieljes transform, which is defined as follows,

Definition 2.1 *Let F be a real probability distribution function and $z \in \mathbb{C}$ be taken outside the support of F. Then the Stieljes transform of F at point z, denoted by $m_F(z)$, is defined as,*

$$m_F(z) \triangleq \int \frac{1}{t - z} dF(t) \tag{2.2}$$

■

Even more interesting for our purposes is the expression of the Stieljes transform for the eigenvalue distribution of hermitian matrices, which is shown in the next definition.

Definition 2.2 *Let $\mathbf{X} \in \mathbb{C}^{M \times M}$ be an hermitian matrix with eigenvalues $\lambda_1, \ldots, \lambda_M$. Moreover, let $F^{\mathbf{X}}$ be an eigenvalue distribution function of $\mathbf{X}$ as defined in (2.1). Then, the Stieljes transform of $F^{\mathbf{X}}$, denoted by $m_{\mathbf{X}}$ is given by,*

$$m_{\mathbf{X}} \triangleq \frac{1}{M}\sum_{k=1}^{M} \frac{1}{\lambda_k - z} = \frac{1}{M}\,\mathrm{Tr}\left[(\mathbf{X} - z\mathbf{I})^{-1}\right] \tag{2.3}$$

■

The Stieljes transform allows to simplify the asymptotic analysis of the eigenvalue distribution. I.e. it plays an analogous role to the Fourier transform, which simplifies the study of signals in the frequency domain instead of the temporal domain. Namely, in order to study the convergence of the eigenvalue distribution, say $F^{\mathbf{B}}$, towards a limiting eigenvalue distribution, say $F^{\mathbf{L}}$, a possible procedure is as follows. First, the Stieljes transform of $F^{\mathbf{B}}$, say $m_{\mathbf{B}}$, is found. Then, one founds that $m_{\mathbf{B}}$ converges to the Stieljes transform of $F^{\mathbf{L}}$, i.e .$m_{\mathbf{L}}$. Finally, one obtains the limiting distribution $F^{\mathbf{L}}$ from $m_{\mathbf{L}}$. This last step can be based on applying the inverse Stieljes transform and the next property of the Stieljes transform, where $\rightarrow$ denotes weakly convergence,

$$F^{\mathbf{B}} \rightarrow F^{\mathbf{L}} \Leftrightarrow m_{\mathbf{B}} \rightarrow m_{\mathbf{L}} \tag{2.4}$$

This procedure is applied for instance to obtain the Marčenko-Pastur law, which is exposed next.

Proposition 2.1 *(Marčenko Pastur Law) [57, Theorem 1.1] Let* $\mathbf{B}$ *be a random matrix fulfilling the next structure* $\mathbf{B} \triangleq \frac{1}{N}\mathbf{X}\mathbf{X}^H \in \mathbb{C}^{M\times M}$ *with* $\mathbf{X} \in \mathbb{C}^{M\times N}$*. Moreover, let* $[\mathbf{X}]_{i,j}$ *be a random matrix such that the real and imaginary parts of the entries are i.i.d according to a gaussian with zero mean, variance 1/2, i.e.* $[\mathbf{X}]_{i,j} \sim \mathcal{CN}(0,1)$*, i.e.* $\mathbf{B}$ *is distributed according to a complex Wishart with* N *degrees of freedom and scale parameter* $\mathbf{I}_M$*, namely* $N^{-1}\mathbf{B} \sim \mathcal{CW}_M(N, \mathbf{I}_M)$*. Then, as* $M\ ,N \rightarrow \infty$ *with* $M\ /N \rightarrow c < +\infty$*, the Stieljes transform of the eigenvalue distribution function* $F^{\mathbf{B}}$ *associated to* $\mathbf{B}$*, denoted by* $m_{\mathbf{B}}$*, converges weakly to the next Stieljes transform* $m_{\mathbf{L}}$*,*

$$m_{\mathbf{L}}(z) = \frac{1-c-z+\sqrt{(z-1-c)^2-4c}}{2cz}, \forall z \in \mathbb{C}^+ \tag{2.5}$$

With an associated eigenvalue pdf given by,

$$f_{\mathbf{L}}(x) = \begin{cases} \frac{1}{2\pi c x}\sqrt{(b-x)(x-a)} & \text{if } a \le x \le b \\ 0, & \text{otherwise} \end{cases} \tag{2.6}$$

Where $a = (1-\sqrt{c})^2$ *and* $b = (1+\sqrt{c})^2$*.*

■

Indeed, the Marčenko-Pastur law its a particular case of [57, Theorem 1.1] that predicts a limiting eigenvalue distribution of random matrices of the form $\mathbf{B} = \frac{1}{N}\mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2}$, provided that $\mathbf{R}$ be a hermitian square positive-definite matrix and that the entries of $\mathbf{X}$ be iid with zero mean, variance 1/2 and with bounded moments. Nonetheless, in general one may not obtain a close analytical form for the limiting pdf of the eigenvalues $f_{\mathbf{L}}(x)$ and has to resort to numerical methods such as the fixed-point to solve a transcendental equation.

In figure 2.1 the Marčenko-Pastur law is exemplified by displaying the eigenvalue pdf in (2.6) for different values of c, namely c =0.1, 0.2 and 0.5. One can see that when $M/N \to c \to 0$ the support of the limiting pdf tends to concentrate in a single mass in 1. This case corresponds to classical asymptotic analysis where the observation dimension M is fixed, whereas the sample size $N \to \infty$, and by the law of large numbers in this case $\mathbf{B} \to \mathbf{I}_M$, i.e. the eigenvalues are all equal to 1 with multiplicity M. On the other hand, when both M and N grow large and their limiting ratio c increases, the eigenvalue density no longer concentrates in a single mass, namely it tends to spread.

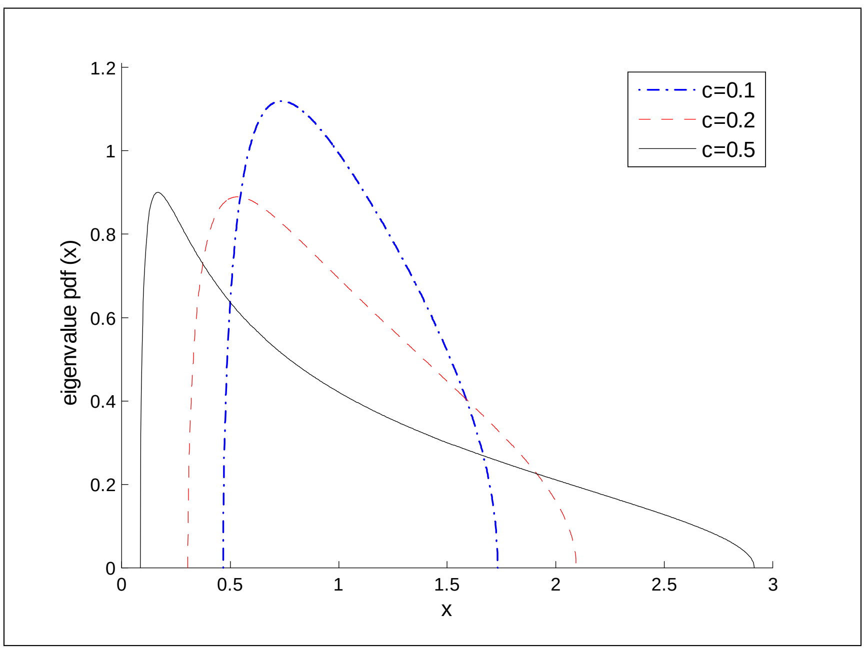

Figure 2.1: Marcenko-Pastur Law: Limiting eigenvalue pdf in (2.6) for different values of $c = M/N$.

Following with the illustration of the Marčenko-Pastur law, in figure 2.2 we represent the histogram of the eigenvalues of $N^{-1}\mathbf{B}$, defined in proposition 2.1, and the associated limiting pdf predicted by the Marčenko-Pastur law (2.6), when $M = 500$, $N = 5000$ and $c = 0.1$. One can observe that the limiting pdf envisaged by the Marčenko-Pastur law is a good match of the empirical histogram as both M and N grow large at a fixed rate c, and that no eigenvalue is outside that pdf.

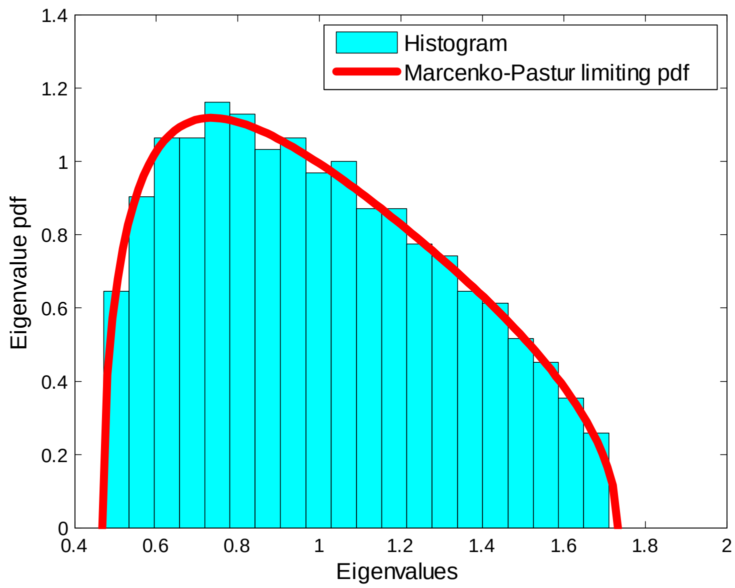

Figure 2.2: Comparison between eigenvalue pdf of $N^{-1}\mathbf{B}$ and its limiting pdf given by Marcenko Pastur law in (2.6), when $M = 500$, $N = 5000$ and $c = 0.1$.

The Stieljes transform presented in (2.3) is appropiate to study the asymptotic behavior of eigenvalues. Nonetheless, for the purposes of this master thesis it is more convenient to study both the asymptotic properties of the eigenvalues and the eigenvectors associated to a given random matrix. To this end, let define the next spectral function associated to the random matrix $\mathbf{X} \in \mathbb{C}^{M \times M}$, which is a generalization of the empirical eigenvalue distribution function in (2.1),

$$G^{\mathbf{X}}(\lambda) = \sum_{m=1}^{M} \mathbf{a}^H \mathbf{e}_m \mathbf{e}_m^H \mathbf{b} \mathcal{I}_{\lambda_m \leq \lambda}(\lambda) \tag{2.7}$$

Where $\mathbf{e}_m$ and λ_m are the *m-th* eigenvector and eigenvalue of $\mathbf{X}$, respectively. Moreover, $\mathbf{a} \in \mathbb{C}^M$ and $\mathbf{b} \in \mathbb{C}^M$ are two generic deterministic vectors. Interestingly enough, this spectral function has an associated Stieljes transform given by the next expression, that was introduced by Girko in e.g. [58],

$$\mathbf{a}^H(\mathbf{X} - z\mathbf{I}_M)^{-1}\mathbf{b} = \sum_{m=1}^{M} \frac{\mathbf{a}^H \mathbf{e}_m \mathbf{e}_m^H \mathbf{b}}{\lambda_m - z} \tag{2.8}$$

Now we can see the importance of (2.8). It resembles the quadratic forms that usually appear in statistical signal processing and that depend on the sample correlation matrix $\hat{\mathbf{R}}$, e.g. $\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}$. Therefore, (2.8) paves the way to study the asymptotic properties of quadratic forms depending on $\hat{\mathbf{R}}$.

G-estimation

G-analysis, also known as G-estimation or general statistical analysis (GSA) is an statistical inference framework that builds on random matrix theory and complex integration methods and was introduced by Girko in e.g. [36]. It provides a framework to derive estimators that are consistent in the doubly asymptotic regime where both the observation dimension M and the sample size N grow large at the same rate, i.e. $M, N \to \infty$ and $M/N \to c \in (0, \infty)$. E.g. in array signal processing M may be the number of antennas and N the number of available snapshots to design a given estimation algorithm. This new statistical inference framework differs from classical estimation that considers M fixed and $N \to \infty$ to derive a consistent estimator. Therefore, the estimators derived under the GSA framework are usually called (M, N)-*consistent*. Moreover, GSA naturally deals with the small sample size regime, i.e. situations where M and N are comparable, and where classical estimators tend to perform poorly. Also due to this reason GSA paves the way to obtain estimators that converge much faster when M grows large than classical estimators.

A great deal of results in G-estimation build upon the so-called "G_{25}-estimator" proposed by Girko in e.g. [37]. Namely, let consider a given covariance matrix $\mathbf{R} \in \mathbb{C}^{M \times M}$ and its sample estimate, i.e. the sample covariance matrix $\hat{\mathbf{R}}$. Moreover, let define the next function, which is a real Stieljes transform of a certain spectral function, analogous to its complex counterpart in (2.8),

$$\mathcal{T}_{\mathbf{R}}(x) = \mathbf{a}^H(\mathbf{I}_M + x\mathbf{R})^{-1}\mathbf{b} = \frac{1}{M} \sum_{m=1}^{M} \frac{\mathbf{a}^H \mathbf{e}_m \mathbf{e}_m^H \mathbf{b}}{1 + x\lambda_m}, \quad x \geq 0 \tag{2.9}$$

With $\mathbf{a}, \mathbf{b} \in \mathbb{C}^M$ two generic deterministic vectors and $\mathbf{e}_m, \lambda_m$ the *m-th* eigenvector and eigenvalue of $\mathbf{R} \in \mathbb{C}^{MxM}$, respectively. Then, an (M, N)-consistent estimator of (2.9) when $M, N \to \infty$ and $M/N \to c \in (0, \infty)$ reads as follows [37],

$$\hat{\mathcal{T}}_{\mathbf{R}}(x) = \mathbf{a}^H (\mathbf{I}_M + \theta(x)\hat{\mathbf{R}})^{-1}\mathbf{b} \tag{2.10}$$

Where $\theta(x)$ is the positive solution of the next equation,

$$\theta(x) \left[1 - c + \frac{c}{M} \operatorname{Tr} \left[(\mathbf{I}_M + \theta(x)\hat{\mathbf{R}})^{-1}\right]\right] = x$$

The importance of the G_{25}-estimator in our framework is that a lot of signal processing and communications expressions can be expressed in terms of (2.9). As a consequence, the G_{25}-estimator paves the way to obtain the (M, N)-consistent estimators of those quantities. E.g. the inverse correlation matrix $\mathbf{R}^{-1}$ or quadratic forms of $\mathbf{R}^k, k > 0$, see [19],

$$\left[\mathbf{R}^{-1}\right]_{i,j} = \lim_{x\to\infty} x\mathbf{u}_i^H (\mathbf{I}_M + x\mathbf{R})^{-1}\mathbf{u}_j \tag{2.11}$$

$$\mathbf{a}^H\mathbf{R}^k\mathbf{b} = \frac{(-1)^k}{k!} \left[\frac{d^k}{dx^k}\mathbf{a}^H(\mathbf{I}_M + x\mathbf{R})^{-1}\mathbf{b}\right]_{|_{x=0}} \tag{2.12}$$

Where $\mathbf{u}_j$ is an all zeros column vector with a 1 in the *j-th* position. In general, the common procedure in G-estimation to obtain (M, N)-consistent estimators of a given random quantity is as follows. First, one expresses the parameter to estimate, say ϕ, as a function of the Stieljes transform of a deterministic matrix $\mathbf{T}$ hidden in the signal model, i.e. $\phi = f(m_{\mathbf{T}})$. Second, one finds that $m_{\mathbf{T}}$ is asymptotically equivalent to the Stieljes Transform of the available matrix of observations $\mathbf{Y}$, i.e. $m_{\mathbf{T}} \asymp g(m_{\mathbf{Y}})$. Finally, one estimates ϕ as $\hat{\phi} = f((g(m_{\mathbf{Y}}))$.

2.2.3 Fundamental Results for the Master Thesis

In this section asymptotic equivalences that pave the way to obtain (M, N)-consistent estimators in the subsequent chapters of this master thesis are presented. Namely, they build upon large dimensional random matrix theory and are based on finding the convergence of certain random quantities depending on the sample estimate of the correlation matrix, under the doubly asymptotic regime where both M and N grow large at a given rate.

Lemma 2.1 *Let $a \asymp b$ mean that $|a-b| \to 0$ almost surely, let $\mathbf{s} \in \mathbb{C}^M$ be a generic deterministic vector, $\mathbf{R} \in \mathbb{C}^{M\times M}$ denote a generic correlation matrix and $\hat{\mathbf{R}} \in \mathbb{C}^{M\times M}$ its sample estimate as defined in (1.8). Moreover, let assume that $\mathbf{s}^H\mathbf{s}$ =1. Then, within the framework of general asymptotics, i.e. $M, N \to \infty$ at a constant rate $M/N \to c \in (0,1)$, the next equivalences hold,*

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp (1-c)^{-1}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} \tag{2.13}$$

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp (1-c)^{-3}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} \tag{2.14}$$

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \asymp \mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp (1-c)^{-1} \tag{2.15}$$

$$\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s} \asymp \mathbf{s}^H\mathbf{R}\mathbf{s} \tag{2.16}$$

Proof: The proofs are presented in the Appendix of this chapter.

■

2.3 Shrinkage Estimation

2.3.1 Introduction: the James-Stein method

Among the methods that estimate the moments of a given distribution, the sample estimators are one of the most widely used. The rationale behind is that the sample estimates are moments of the empirical distribution, which converges almost surely to the true distribution when the number of iid observations tends to infinity and the observation dimension is fixed, according to the Glivenko-Cantelli theorem. Indeed, it is well known that, when the data is Gaussian, the sample mean and the sample covariance are the ML estimators of the mean and the covariance of the true distribution, respectively.

Nonetheless, in a pathbreaking and astonishing publication [29], Stein proved that the sample mean is not an admissible estimator of the mean of a multivariate gaussian distribution, when the observation dimension is larger than one. Namely, he proposed an estimator, which is called nowadays James-Stein estimator, which displays lower MSE than the sample mean. This seminal work of Stein was so-called Stein's phenomenon and it was the foundation of shrinkage estimation. Notable contributions to the understanding of this phenomenon were [30], [31] and [59–63]. See also [28] for a thorough discussion about this topic and in general about shrinkage estimation.

The main idea behind shrinkage estimation may be summarized as follows. The bulk of error (MSE) of the sample estimators comes from their estimation variance, i.e. their bias is quite limited. Therefore, if one intends to outperform the sample estimators, a possible approach is to design methods that have larger bias than them but a lower variance such that the overall MSE is lower than the one of the sample estimators. Stein gave expression to this idea by means of an estimator consisting of a linear scaling of the sample mean. Subsequently, Stein and James generalized this concept conceiving the so-called James-Stein estimator. It is based on blending, by means of a weighted average, the sample mean, which displays much higher estimation variance than bias, with a constant estimator of the mean, which displays high bias but no variance. Thus, by optimally combining the bias variance tradeoff, the James-Stein estimator obtains a lower MSE than the sample mean. This concept may be generalized to the estimation of any parameter and it is the basis of shrinkage estimation. Next, we provide the expression of the James-Stein estimator, by means of the next proposition. Namely it is the expression exposed in [28], which is more general than the original method proposed by James and Stein.

Proposition 2.2 *(James-Stein estimator) Let* $\mathbf{x} \in \mathbb{R}^M$ *be a multivariate Gaussian random variable* $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ *and* $M > 1$*. Moreover, let* N *realizations of* $\mathbf{x}$ *be available, i.e.* $\{\mathbf{x}_n\}_{n=1}^N$*, and let denote by* $\hat{\boldsymbol{\mu}}$ *the sample estimator of* $\boldsymbol{\mu}$*, namely* $\hat{\boldsymbol{\mu}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$*. Then, the James-Stein estimator of* $\boldsymbol{\mu}$ *reads as follows,*

$$\hat{\boldsymbol{\mu}}^s = (1-\alpha)\hat{\boldsymbol{\mu}} + \alpha\mathbf{b} \tag{2.17}$$

Where α *is the shrinkage factor and it depends on the largest eigenvalue and the average of the eigenvalues of* $\boldsymbol{\Sigma}$ *denoted by* λ_1 *and* $\bar{\lambda}$*, respectively,*

$$\alpha = \frac{1}{N}\frac{M\bar{\lambda} - 2\lambda_1}{(\hat{\boldsymbol{\mu}} - \mathbf{b})^T(\hat{\boldsymbol{\mu}} - \mathbf{b})}$$

And $\mathbf{b}$ *is a constant estimator of* $\boldsymbol{\mu}$*. Namely, it can be any* M*-dimensional fixed vector stemming from a priori information of the problem at hand. Moreover, the James-Stein estimator dominates the sample mean for any choice of* $\mathbf{b}$*.*

Proof: For a proof of the dominance of the James-Stein estimator on the sample mean we refer the reader to [64, Appendix 4.5]. The definition of stochastic dominance was given in Chapter 1.

■

Indeed, there are other possible choices for $\mathbf{b}$ other than a constant vector. E.g. Jorion in [65] proposed to use an estimator based on the grand mean, namely $\mathbf{b} = \frac{\mathbf{1}^T \hat{\boldsymbol{\Sigma}}^{-1} \hat{\boldsymbol{\mu}}}{\mathbf{1}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{1}} \mathbf{1}$, where $\mathbf{1}$ is an M-dimensional vector of all ones and $\hat{\boldsymbol{\Sigma}}$ is an estimator of $\boldsymbol{\Sigma}$, e.g. the sample covariance matrix. Moreover, although for a generic constant vector $\mathbf{b}$, the James-Stein method dominates the sample mean, its choice is important. Namey, the smaller $\|\mathbf{b} - \boldsymbol{\mu}\|$ the better the performance. Regarding this method, another important remark is that other authors have shown that shrinkage type estimators dominate the sample mean for a broader class of distributions other than the Gaussian, see e.g. Evans and Stark [66].

In order to get more insights on shrinkage estimation it is worth analyzing the shrinkage factor α. Usually the constraint that $\alpha \in (0, 1)$ is imposed. Thus, on the one hand when $\alpha \to 0$, the James-Stein method tends to the sample mean $\hat{\boldsymbol{\mu}}^s \to \hat{\boldsymbol{\mu}}$. On the other hand, $\alpha \to 1$ implies that $\hat{\boldsymbol{\mu}}^s \to \mathbf{b}$, i.e. the estimator is shrunken toward "the target" $\mathbf{b}$. This is a phenomenon that happens in general in these type of estimators and this is the reason of calling them shrinkage estimators. Indeed, α controls the amount of shrinkage. Namely, according to (2.17), when the sample size N tends to infinity and the observation dimension M remains fixed, the shrinkage factor tends to vanish, i.e. $\alpha \to 0$ and the James-Stein tends to the sample mean $\hat{\boldsymbol{\mu}}^s \to \hat{\boldsymbol{\mu}}$. Which is logic as in this situation the sample mean is the optimal estimator. On the other hand, when M tends to be comparable or even higher than N, the shrinking factor α tends to increase and $\hat{\boldsymbol{\mu}}^s$ is shrunken towards $\mathbf{b}$. This supports the intuition that in the small sample size regime the performance of sample mean method is considerably degraded. Or in other words, $\hat{\boldsymbol{\mu}}^s$ is shrunken towards $\mathbf{b}$ because the information obtained from the measured samples is worse than the a priori information embedded in $\mathbf{b}$. This behavior also highlights the robustness of shrinkage estimators to the small sample size regime. Indeed, as the interpretation of the shrinkage factor has glimpsed, the James-Stein estimator and in general shrinkage estimation can be explained within the context of Bayesian estimation, e.g. Efron in [67] showed the empirical Bayes derivation of the James-Stein method.

The paradigm initiaded by the James-Stein estimator has been applied in several works among the community of signal processing. Poor in [68] applied it to adaptive filtering, other authors applied the Stein's Unbiased Risk Estimator (SURE) principle, which stems from James-Stein estimation, to obtain methods having lower MSE than the ML, see [69] and references therein. Moreover, James-Stein estimation has been applied to other fields of science such as quantitative finance [65].

2.3.2 Shrinkage estimators of the sample covariance

Although shrinkage estimation arose in the context of estimating the mean of a Gaussian distribution, it has been applied to the estimation of other parameters, e.g. to the estimation of the covariance of a given distribution. Originally it was also Stein who studied the shrinkage of the SCM in [30]. More recently, Ledoit and Wolf proposed in [25] a shrinkage estimator of the SCM $\hat{\mathbf{R}}$, consisting of shrinking $\hat{\mathbf{R}}$ towards the identity matrix by means of a linear combination. The contribution of that method is that it deals with the case where the sample size N may be lower than the observation dimension M and that does not require any assumption about the distribution of the data used for the estimation. Considering as a reference the work of Ledoit and Wolf, recently the signal processing community has proposed other shrinkage estimators of the SCM. E.g., Stoica, Guerci et al. [24], in the context of radar, extended the work of Ledoit and Wolf to complex data and to a shrinkage target consisting of a general matrix which expresses a priori information about the SCM, which is obtained from the problem at hand. In this regard, another contribution was proposed by Eldar, Hero et al. in [70], where assuming a Gaussian distribution of the data, the authors proposed two shrinkage estimators of the SCM that outperform the Ledoit and Wolf method. In order to get more insights, the Ledoit and Wolf estimator is exemplified in the next proposition.

Proposition 2.3 *(Ledoit and Wolf Shrinkage estimator of the covariance) Let* $\mathbf{X} \in \mathbb{R}^{M \times N}$ *be a matrix of* N *iid observations of* M *random variables with mean zero and covariance* $\mathbf{\Sigma}$*. Let denote by* $\hat{\mathbf{\Sigma}} = \mathbf{X}\mathbf{X}^T/N$ *the SCM. Consider the problem of estimating* $\mathbf{\Sigma}$ *based on a shrinkage estimator of the SCM towards a scaling of the identity matrix,* $\mathbf{\Sigma}_0 = \frac{\mathrm{Tr}[\hat{\mathbf{\Sigma}}]}{M}\mathbf{I}_M$*, which minimize the MSE, namely,*

$$\begin{aligned} &\min_{\rho} \mathbb{E}\left[\left\|\check{\mathbf{\Sigma}} - \mathbf{\Sigma}\right\|_F^2\right] \\ &s.t.\check{\mathbf{\Sigma}} = (1-\rho)\hat{\mathbf{\Sigma}} + \rho\mathbf{\Sigma}_0 \end{aligned} \tag{2.18}$$

Where $\|\cdot\|_F$ *denotes the Frobenius norm. Then, an (M,N)-consistent estimator of the optimal, though unrealizable, solution to (2.18) within the general asymptotics where* $M, N \to \infty$ *at a constant rate* $M/N \to c \in (0, \infty)$*, is given by the Ledoit and Wolf estimator,*

$$\begin{aligned} \check{\mathbf{\Sigma}}^{LW} &= (1-\hat{\rho}^{LW})\hat{\mathbf{\Sigma}} + \hat{\rho}^{LW}\mathbf{\Sigma}_0 \\ \hat{\rho}^{LW} &= \frac{\sum_{n=1}^{N}\left\|\mathbf{x}_n\mathbf{x}_n^T - \hat{\mathbf{\Sigma}}\right\|_F^2}{N^2\,\mathrm{Tr}\left[\hat{\mathbf{\Sigma}}^2\right] - \frac{\mathrm{Tr}^2[\hat{\mathbf{\Sigma}}]}{M}} \end{aligned} \tag{2.19}$$

Where $\mathbf{x}_n$ *is the* n*-th column of* $\mathbf{X}$*.*

Proof: For a proof of this proposition the reader is referred to [25].

■

Appendix Proof of Fundamental Results

Proof of (2.13) and (2.14)

This proof is mainly based on the proof provided in [19, Appendix I], (cf. [15, Chapter4], [20] and [26]). First, let define the next random quantities,

$$\eta_n(\alpha) \; = \mathbf{s}^H(\hat{\mathbf{R}}+\alpha\mathbf{I}_M)^{-1}\mathbf{R}(\hat{\mathbf{R}}+\alpha\mathbf{I}_M)^{-1}\mathbf{s} \tag{2.20}$$

$$\eta_d(\alpha) \; = \mathbf{s}^H(\hat{\mathbf{R}}+\alpha\mathbf{I}_M)^{-1}\mathbf{s} \tag{2.21}$$

Hence, we can express the quantities of interest $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}$ as a function of η_n and η_d, respectively,

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} = \lim_{\alpha\to 0}\eta_n(\alpha) \tag{2.22}$$

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} = \lim_{\alpha\to 0}\eta_d(\alpha) \tag{2.23}$$

Now, note that $\hat{\mathbf{R}}$ may be decomposed as a function of the true hermitian matrix $\mathbf{R}$ and a random matrix $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$ with iid entries and whose real and imaginary parts are independent, have zero mean and $1/2N$ variance,

$$\hat{\mathbf{R}} = \mathbf{R}^{1/2}\mathbf{\Phi}\mathbf{\Phi}^H\mathbf{R}^{1/2} \tag{2.24}$$

With the decomposition in (2.24), expressions $\eta_n(\alpha)$ in (2.20) and $\eta_d(\alpha)$ in (2.21) may be written in terms of another function $m(z)$,

$$\eta_d(\alpha) = m(z)_{|_{z=0}} \tag{2.25}$$

$$\eta_n(\alpha) = \frac{dm(z)}{dz}|_{z=0} \tag{2.26}$$

Where $m(z)$ is defined as follows,

$$m(z) = \mathbf{s}^H\mathbf{R}^{-1/2}(\mathbf{\Phi}\mathbf{\Phi}^H + \alpha\mathbf{R}^{-1} - z\mathbf{I}_M)^{-1}\mathbf{R}^{-1/2}\mathbf{s} \tag{2.27}$$

With this definition at hand, observe that indeed $m(z)$ has the form $\mathbf{a}^H(\mathbf{M} - z\mathbf{I}_M)^{-1}\mathbf{a}$, which as it was commented in (2.8) is the Stieljes transform of a spectral function of the type $G^{\mathbf{M}}(\lambda) = \sum_{m=1}^{M} \mathbf{a}^H\mathbf{e}_m\mathbf{e}_m^H\mathbf{a}\mathcal{I}_{\lambda_m \leq \lambda}(\lambda)$ associated to the random matrix $\mathbf{M}$, see (2.7). Therefore, the proof of the asymptotic equivalences (2.13) and (2.14) is based on studying the asymptotic convergence of spectral functions of matrices of the type $\mathbf{M} = \mathbf{\Phi}\mathbf{\Phi}^H + \alpha\mathbf{R}^{-1}$. This is provided by [19, Lemma 1 in Appendix I] and leads to obtain that $m(z)$ converges almost surely to the function $\bar{m}(z)$ when both $M, N \to \infty$ at a constant rate c, namely,

$$m(z) \asymp \bar{m}(z)$$

$$\bar{m}(z) = \sum_{m=1}^{M} \frac{\mathbf{s}^H\mathbf{R}^{-1/2}\mathbf{e}_m\mathbf{e}_m^H\mathbf{R}^{-1/2}\mathbf{s}(1 + cb(z))}{1 + (\alpha\lambda_m^{-1} - z)(1 + cb(z))} \tag{2.28}$$

With $\mathbf{e}_m$ and λ_m being the *m-th* eigenvector and eigenvalue of $\mathbf{R}$, respectively, and $b(z)$ being the positive solution to the next transcendental equation,

$$b(z) = \frac{1}{M}\sum_{m=1}^{M} \frac{(1 + cb(z))}{1 + (\lambda_m - z)(1 + cb(z))}$$

Now, recalling that according to (2.25) $\eta_d(\alpha) = m(z)_{|_{z=0}}$ and that according to (2.28) $m(z) \asymp \bar{m}(z)$, we readily obtain that $\eta_d(\alpha)$ converges in probability to $\bar{\eta}_d(\alpha)$,

$$\eta_d(\alpha) \asymp \bar{\eta}_d(\alpha)$$

$$\bar{\eta}_d(\alpha) = \sum_{m=1}^{M} \frac{\left|\mathbf{s}^H\mathbf{e}_m\right|^2 (1 + cb)}{\lambda_m + \alpha(1 + cb)} \tag{2.29}$$

With $b = b(0)$ the positive solution to the next equation,

$$b = \frac{1}{M}\sum_{m=1}^{M}\frac{\lambda_m(1+cb)}{\lambda_m+\alpha(1+cb)} \tag{2.30}$$

Moreover, recalling that according to (2.26) $\eta_n(\alpha) = \frac{dm(z)}{dz}|_{z=0}$ and again bearing in mind that according to (2.28) $m(z) \asymp \bar{m}(z)$, we readily obtain the next asymptotic equivalence for $\eta_n(\alpha)$,

$$\eta_n(\alpha) \asymp \bar{\eta}_n(\alpha)$$

$$\bar{\eta}_n(\alpha) = \sum_{m=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_m\right|^2\lambda_m(1+cb)^2+cb'}{\left(\lambda_m+\alpha(1+cb)\right)^2} \tag{2.31}$$

With $b' = \frac{db(z)}{dz}|_{z=0}$ having the next expression,

$$b' = \frac{db(z)}{dz}|_{z=0} = \left[1-\frac{1}{M}\sum_{m=1}^{M}\frac{c\lambda_m^2}{\left(\lambda_m+\alpha(1+cb)\right)^2}\right]^{-1}\left[\frac{1}{M}\sum_{m=1}^{M}\frac{\lambda_m^2(1+cb)^2}{\left(\lambda_m+\alpha(1+cb)\right)^2}\right] \tag{2.32}$$

At this point, recalling the relation between $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}$ and $\eta_d(\alpha)$ shown in (2.23) and considering the asymptotic equivalent of $\eta_d(\alpha)$ in (2.29),

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp \lim_{\alpha\to 0}\bar{\eta}_d(\alpha) \tag{2.33}$$

Now, as according to (2.30) $b \overset{\alpha\to 0}{\longrightarrow} (1-c)^{-1}$ and after operating the limit $\alpha \to 0$, we can rewrite (2.33) as follows,

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp \frac{1}{(1-c)}\sum_{m=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_m\right|^2}{\lambda_m}$$

Finally, noting that the right hand of this expression is a Stieljes transform and recalling the equivalence in (2.8), we obtain the next asymptotic equivalence, which concludes the proof for (2.13),

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

■

Moreover, recalling the relation between $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$ and $\eta_n(\alpha)$ shown in (2.22) and considering the asymptotic equivalent of $\eta_n(\alpha)$ in (2.31),

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp \lim_{\alpha\to 0}\bar{\eta}_n(\alpha) \tag{2.34}$$

According to (2.30) and (2.32) $b \xrightarrow{\alpha\to 0} (1-c)^{-1}$ and $b' \xrightarrow{\alpha\to 0} (1-c)^{-3}$, respectively. Therefore, after operating the limit $\alpha \to 0$, we can rewrite (2.34) as follows,

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp \frac{1}{(1-c)^3}\sum_{m=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_m\right|^2}{\lambda_m}$$

Finally, noting that the right hand of this expression is a Stieljes transform and recalling the equivalence in (2.8), we obtain the next asymptotic equivalence, which concludes the proof for (2.14),

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

■

Proof of (2.15)

The proof of this asymptotic equivalence is analogous to the one for (2.13), though it is based on [38, Theorem 1]. First, note that both $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}$ and $\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$ can be expressed in terms of a Stieljes transform $\hat{m}(z)$ as follows,

$$\hat{m}(z) = \mathbf{a}^H(\hat{\mathbf{R}}-z\mathbf{I}_M)^{-1}\mathbf{b} = \sum_{m=1}^{M}\frac{\mathbf{a}^H\hat{\mathbf{e}}_m\hat{\mathbf{e}}_m^H\mathbf{b}}{\hat{\lambda}_m - z}$$

$$\begin{gathered}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} = \lim_{z\to 0}\hat{m}(z) \\ \hat{m}(z) \text{ s.t. } \mathbf{a}=\mathbf{s},\ \mathbf{b}=\mathbf{R}\mathbf{s}\end{gathered} \tag{2.35}$$

$$\begin{gathered}\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} = \lim_{z\to 0}\hat{m}(z) \\ \hat{m}(z) \text{ s.t. } \mathbf{a}=\mathbf{R}\mathbf{s},\ \mathbf{b}=\mathbf{s}\end{gathered} \tag{2.36}$$

Therefore, in order to proof **(**2.15**)**, the asymptotics of $\hat{m}(z)$ must be studied. This is provided by means of [38, Theorem 1], which states that $\hat{m}(z)$ converges almost surely to a function $\bar{m}(z)$ when $M, N \to \infty$ and $M/N \to c$ with $0 < c < \infty$,

$$\hat{m}(z) \asymp \bar{m}(z)$$

$$\bar{m}(z) = \sum_{m=1}^{M} \frac{\mathbf{a}^H \mathbf{e}_m \mathbf{e}_m^H \mathbf{b}}{\lambda_m(1 - c - cz\bar{b}(z)) - z} \tag{2.37}$$

Being $\bar{b}(z)$ the positive solution to the next transcendental equation,

$$\bar{b}(z) = \frac{1}{M} \sum_{m=1}^{M} \frac{1}{\lambda_m(1 - c - cz\bar{b}(z)) - z} \tag{2.38}$$

Now, recalling the relations in (2.35) and (2.36), applying the limit $z \to 0$ in the expression of $\bar{m}(z)$ and assuming $\mathbf{s}^H\mathbf{s}$ =1, we obtain the convergence of the desired quantities $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}$ and $\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$ which concludes the proof,

$$\begin{gathered} \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \asymp \lim_{z\to 0} \bar{m}(z) = \tfrac{1}{1-c} \\ \bar{m}(z) \text{ s.t. } \mathbf{a} = \mathbf{s},\ \mathbf{b} = \mathbf{R}\mathbf{s} \end{gathered}$$

$$\begin{gathered} \mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp \lim_{z\to 0} \bar{m}(z) = \tfrac{1}{1-c} \\ \bar{m}(z) \text{ s.t. } \mathbf{a} = \mathbf{R}\mathbf{s},\ \mathbf{b} = \mathbf{s} \end{gathered}$$

■

Proof of (2.16)

The proof is based on [19], cf. also [71, Proposition 1]. First, note that according to (2.12), $\mathbf{s}^H\mathbf{R}\mathbf{s}$ can be expressed as,

$$\mathbf{s}^H\mathbf{R}\mathbf{s} = -\left[\frac{d}{dx}\mathbf{s}^H(\mathbf{I}_M + x\mathbf{R})^{-1}\mathbf{s}\right]_{|_{x=0}} \tag{2.39}$$

Now, the key in the proof is to recall the G-25 estimator of the real Stieljes transform of $\mathbf{R}$ stated in (2.9) and (2.10),

$$\mathbf{s}^H(\mathbf{I}_M + x\mathbf{R})^{-1}\mathbf{s} \asymp \mathbf{s}^H(\mathbf{I}_M + \theta(x)\hat{\mathbf{R}})^{-1}\mathbf{s} \tag{2.40}$$

Where $\theta(x)$ is the positive solution of the next equation,

$$\theta(x)\left[1-c+\frac{c}{M}Tr\left[(\mathbf{I}_M+\theta(x)\hat{\mathbf{R}})^{-1}\right]\right]=x$$

Now considering the asymptotic equivalence (2.40) in (2.39) we obtain the next relation,

$$\mathbf{s}^H\mathbf{R}\mathbf{s}\asymp-\left[\frac{d}{dx}\mathbf{s}^H(\mathbf{I}_M+\theta(x)\hat{\mathbf{R}})^{-1}\mathbf{s}\right]_{|_{x=0}}=\mathbf{s}^H(\mathbf{I}_M+\theta(x)\hat{\mathbf{R}})^{-2}\frac{d\theta(x)}{dx}\hat{\mathbf{R}}\mathbf{s}_{|_{x=0}}$$

Finally, after easy manipultaions it is easy to check that $\theta(0)=0$ and that $\frac{d\theta(x)}{dx}|_{x=0}=1$ and as a consequence we obtain the next relation, which concludes the proof,

$$\mathbf{s}^H\mathbf{R}\mathbf{s}\asymp\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}$$

■

Chapter 3

Optimal shrinkage for large sample LMMSE

3.1 Introduction

In this chapter shrinkage corrections of the sample LMMSE estimators are dealt with. The proposed methods overcome the main drawbacks of the sample LMMSE (1.9), as they are robust to the small sample size regime and consistent under the doubly asymptotic regime where both the sample size M and the observation dimension N grow towards infinity at a fixed rate c. Two type of estimators are considered, a shrinkage of the sample LMMSE $\mathbf{w}_s = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$ and a more general case $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, where the sample LMMSE is shrunk towards a Bartlett filter, which is known to give better results for small sample size regimes. The design of the shrinkage factors, which characterize the estimators, is based on the minimization of the MSE, as it was stated in (1.11). Moreover, as direct minimization of the MSE leads to unrealizable estimators, we propose to use RMT. This mathematical tool leads to obtain consistent estimates which are asymptotically optimal, in an MSE sense. Moreover, the numerical simulations will highlight that the proposed methods not only are consitent and robust to the small sample size regimes, but also outperform the sample LMMSE in any of the sample sizes considered herein, i.e. $\frac{M}{N} \in (0, 1)$.

3.2 Optimal shrinkage of the sample LMMSE estimator

We begin this chapter by introducing which are the optimal, though unrealizable filters for direct shrinkage of the sample LMMSE filter and shrinkage of the sample LMMSE towards a Bartlett filter. They are presented in the next two lemmas and they serve as a benchmark to develop the proposed estimators in the upcoming sections.

Lemma 3.1 *Assume that a set of observations* $\{\mathbf{y}(n)\}_{n=1}^{N}$ *fulfilling the model in (1.1), with assumptions (a)-(e) is available. Given* $\{\mathbf{y}(n)\}_{n=1}^{N}$*, consider the problem of estimating the unknown* $x(n)$ *in (1.1), based on minimizing the MSE in (1.5), when the estimator* $\hat{x}_{lb,s}(n) = \mathbf{w}_{lb,s}^{H}\mathbf{y}(n)$ *is a linear shrinkage of the sample LMMSE towards a Bartlett filter, i.e.* $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}+\alpha_2\mathbf{s}$*. This problem is mathematically formulated as follows,*

$$\hat{x}_{lb,s}(n) = \mathbf{w}_{lb,s}^{H}\mathbf{y}(n); \qquad \begin{aligned} \mathbf{w}_{lb,s} = &\arg\min_{\mathbf{w}} \ \mathrm{MSE}(\mathbf{w}) \\ &s.t.\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}+\alpha_2\mathbf{s} \end{aligned} \tag{3.1}$$

Then, defining $\boldsymbol{\alpha}_{lb} \triangleq (\alpha_1 \ , \ \alpha_2)^T$*, the optimal solution for this problem is given by the next shrinkage factors,*

$$\boldsymbol{\alpha}_{lb} = \frac{\gamma\begin{pmatrix} \mathbf{s}^H\mathbf{R}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \\ \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \end{pmatrix}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H\mathbf{R}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}} \tag{3.2}$$

Proof: See section 3.4.

■

Next we present which is the optimal shrinkage factor when a direct shrinkage of the LMMSE method is considered. Indeed, this is a particular case of the filter presented in Lemma 3.1.

Lemma 3.2 *Assume that a set of observations* $\{\mathbf{y}(n)\}_{n=1}^{N}$ *fulfilling the model in (1.1), with assumptions (a)-(e) is available. Given* $\{\mathbf{y}(n)\}_{n=1}^{N}$*, consider the problem of estimating the unknown* $x(n)$ *in (1.1), based on minimizing the MSE in (1.5), when the estimator*

$\hat{x}_{l,s}(n) = \mathbf{w}_{l,s}^H \mathbf{y}(n)$ is a linear shrinkage of the sample LMMSE, i.e. $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$. This problem is mathematically formulated as follows,

$$\hat{x}_{l,s}(n) = \mathbf{w}_{l,s}^H \mathbf{y}(n); \qquad \begin{array}{c} \mathbf{w}_{l,s} = \arg\min\limits_{\mathbf{w}} \ \mathrm{MSE}\left(\mathbf{w}\right) \\ s.t. \mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s} \end{array} \tag{3.3}$$

Then, the optimal solution for this problem is given by the next shrinkage factor,

$$\alpha_l = \frac{\gamma \mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}} \tag{3.4}$$

Proof: See section 3.4.

■

Interesting enough (3.4) highlights that the sample LMMSE is not in general an optimal estimator in the MSE sense. Indeed, it is obtained when substituting $\mathbf{R}$ for the SCM in the optimal shrinkage factor (3.4) and as a consequence is only optimal in the large sample size regime. Indeed, the sample LMMSE neither is in general a consistent estimate of the LMMSE estimator within the general asymptotics framework considered herein. Namely, as it can be easily derived from [19, Section V.A] $\mathbf{R}^{-1}\mathbf{s} \asymp (1-c)\hat{\mathbf{R}}^{-1}\mathbf{s}$, where $\asymp$ denotes almost surely convergence. I.e. the consistent estimator of $\mathbf{R}^{-1}\mathbf{s}$ within the general asymptotics where $M, N \to \infty$ with $M/N \to c \in (0,1)$ is $(1-c)\hat{\mathbf{R}}^{-1}\mathbf{s}$.

The expressions of the optimal shrinkage LMMSE estimators in (3.2) and (3.4) highlight the dependance on the unknown $\mathbf{R}$. As a consequence they are not realizable. A possible approach to circumvent this problem is to substitute the unknown $\mathbf{R}$ for its sample estimate. This point of view is proposed by some authors dealing with analogous shrinkage estimation problems, e.g. [33] in the context of optimal portfolio allocation in quantitative finance. Nonetheless, that approach entails an estimation risk that may lead to a performance degradation. Indeed, applying this strategy to any of the proposed shrinkage estimators of the sample LMMSE, i.e. (3.2) and (3.4), leads to the conventional sample LMMSE method (1.9), as $\alpha_{l|\mathbf{R}=\hat{\mathbf{R}}} = \gamma$ and $\boldsymbol{\alpha}_{lb|\mathbf{R}=\hat{\mathbf{R}}} = (\gamma, 0)^T$, and as a consequence the potencial benefits of the shrinkage approach are lost.

In the next section, in order to tackle this problem another strategy based on RMT is proposed. This strategy, not only leads to counteract the small sample size degradation and to obtain a realizable estimator, but also yields consistent estimates of the optimal, though unrealizable, shrinkage estimators.

3.3 Optimal shrinkage for large sample LMMSE

Next, let introduce by means of two theorems the main contributions of this chapter. Namely, we propose methods which are consistent estimates of the optimal, though unrealizable, shrinkage methods introduced in the last section by means of expressions (3.1) to (3.4). The tool enabling these results is RMT, as it provides a suitable framework to study the stochastic convergence of the involved random matrices, under the general asymptotic regime where both the observation dimension M and the sample size N tend to infinity at a fixed rate $M/N \to c \in (0,1)$.

Theorem 3.1 *Let define $\check{\boldsymbol{\alpha}}_{lb} \triangleq \left(\check{\alpha}_{lb,1}\ ,\ \check{\alpha}_{lb,2}\right)^T$, then a realizable and consistent estimate of the optimal shrinkage of the sample LMMSE towards a Bartlett filter (3.1), within the general asymptotics where $M, N \to \infty$, $M/N \to c \in (0,1)$, reads as follows,*

$$\check{x}_{lb,s}(n) = \check{\mathbf{w}}_{lb,s}^H \mathbf{y}(n); \check{\mathbf{w}}_{lb,s} = \check{\alpha}_{lb,1} \hat{\mathbf{R}}^{-1}\mathbf{s} + \check{\alpha}_{lb,2}\mathbf{s}$$

$$\check{\boldsymbol{\alpha}}_{lb} = \frac{\gamma \begin{pmatrix} (1-c)^2 \mathbf{s}^H \hat{\mathbf{R}} \mathbf{s}\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s} - (1-c) \\ c\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s} \end{pmatrix}}{\mathbf{s}^H \hat{\mathbf{R}} \mathbf{s}\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s} - 1} \tag{3.5}$$

Proof: See section 3.4.

■

Remark: The proposed estimator in (3.5) incorporates implicitly the robustness to small sample size regimes, i.e. $M \approx N$, as it is a shrinkage estimator and it is based on the RMT approach, that implicitly considers this scenario through $M/N \to 1$. In the numerical results section, more insights about the robustness to the small sample size regime will be given. Moreover, in that section, it will be demonstrated that (3.5) outperforms the traditional sample LMMSE estimator (1.9) in all the sample size regimes dealt with herein, i.e. $M/N \in (0,1)$. This makes sense as the design of the proposed estimator, based on minimizing the MSE, embraces both the large and small sample size regimes, i.e. $N >> M$ and $M \approx N$, respectively. More specifically it is valid for any ratio $M/N \in (0,1)$. On the contrary, the conventional sample LMMSE is a rather ad hoc method, as it does not consider the minimization of the MSE when having the SCM instead of the true correlation $\mathbf{R}$ in the expression of the LMMSE method. Indeed, the presence of the SCM entails an estimation risk, which leads to a performance degradation in any sample size regime. Obviously, this performance degradation is more evident when we approach the small sample size regime.

In order to gain more insights about how the shrinkage estimation framework affects the method proposed in Theorem 3.1 it is interesting to study the asymptotic values of the shrinkage filter when c tends to its extreme values, i.e. $c \to 1$ and $c \to 0$ which denote a small and large sample size regime, respectively. Thus, when the sample dimension is much larger than the observation dimension, i.e. $c \to 0$, the shrinkage filter tends to the traditional LMMSE implementation.

$$c \to 0 \Rightarrow \check{\mathbf{w}}_{lb,s} \to \gamma \hat{\mathbf{R}}^{-1}\mathbf{s} \tag{3.6}$$

This behavior makes sense as in this situation, $\hat{\mathbf{R}}$ is the optimal estimator of $\mathbf{R}$ and as a consequence $\gamma \hat{\mathbf{R}}^{-1}\mathbf{s} \to \gamma \mathbf{R}^{-1}\mathbf{s}$, i.e. the traditional implementation of the LMMSE tends to the optimal LMMSE filter in (1.4). With regard to the case where $c \to 1$, i.e. in the small sample size regime, it is easy to check the following relation for the shrinkage filter holds,

$$c \to 1 \Rightarrow \check{\mathbf{w}}_{lb,s} \to \left(\frac{\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H \hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s} - 1} \right) \mathbf{s} \tag{3.7}$$

I.e. in the small sample size regime, the shrinkage filter tends to a scaling of a Bartlett type filter an it disregards the contribution of the sample LMMSE, as it has in general worse performance than a Bartlett type filter. The expressions (3.6) and (3.7) highlight the rationale behind the shrinkage estimation paradigm, which optimally combines, by means of a weighted average, a sample based estimator with an estimator based on available a priori information. Thus, on the one hand, in the large sample size regime, as the sample LMMSE is optimal, the proposed shrinkage method tends to it. On the other hand, as in the small sample size regime an estimator based on a Bartlett type filter may behave better than the sample LMMSE, the proposed shrinkage filter tends to a scaling of a Bartlett type filter. Another interpretation is in terms of bias variance tradeoff. I.e. the proposed shrinkage estimator aims to optimize the bias variance tradeoff to approach to the minimum MSE. In this regard, the bias is mainly coming from the Bartlett type filter, whereas the variance is coming from the sample based method.

Next, by means of a new theorem, we present the consistent estimate for the optimal shrinkage of the sample LMMSE method in (3.3).

Theorem 3.2 *A realizable and consistent estimate of the optimal shrinkage of the sample LMMSE filter (3.3), within the general asymptotics where $M, N \to \infty$, $M/N \to c \in (0,1)$, reads as follows,*

$$\begin{array}{c} \breve{x}_{l,s}(n) = \breve{\mathbf{w}}_{l,s}^H \mathbf{y}(n); \breve{\mathbf{w}}_{l,s}{=}\breve{\alpha}_l \hat{\mathbf{R}}^{-1}\mathbf{s} \\ \breve{\alpha}_l = \gamma(1{-}c)^2 \end{array} \tag{3.8}$$

Proof: See section 3.4.

■

As it was reasoned out for the more general case exposed in Theorem 3.1, the proposed estimator in Theorem 3.2 not only is realizable and consistent, but also robust to the small sample size regime. This is due to its shrinkage structure and to rely on the RMT approach, as it was discussed previously. Moreover, the numerical simulations section, will highlight that is outperforms the conventional sample LMMSE in any of the sample size regimes considered herein, i.e. $M/N \in (0,1)$. It is also worth observing that the method proposed in Theorem 3.2 is likely to behave worse than the one in Theorem 3.1, especially in the small sample size regime. This will be confirmed in the numerical simulations section. The reason for that behavior is that the method in Theorem 3.1 incorporates more a priori information in the structure of the estimator, through the presence of the Bartlett filter $\mathbf{w} = \mathbf{s}$, which is known to behave better than the sample LMMSE in the small sample size regime.

3.4 Proofs

In this section, the main results of this chapter will be proven. Namely, the statements in question are Lemma 3.1, Lemma 3.2, Theorem 3.1 and Theorem 3.2.

Proof of Lemma 3.1

Let define $\boldsymbol{\alpha} \triangleq (\alpha_1\ ,\ \alpha_2)^T$, $\boldsymbol{\Omega} \triangleq \left(\hat{\mathbf{R}}^{-1}\mathbf{s}\ ,\ \mathbf{s}\right)$ and $\boldsymbol{\alpha}_o$ the optimal shrinkage factors, then the MSE optimization problem in (3.1) may be reformulated as,

$$\boldsymbol{\alpha}_o{=}\arg\min_{\boldsymbol{\alpha}}\ \mathrm{MSE}\left(\mathbf{w} = \boldsymbol{\Omega}\boldsymbol{\alpha}\right)$$

Bearing in mind the expression of the MSE in (1.5) and recalling that according to assumption *(b)* in our data model (1.1), $\mathbf{R} = \gamma \mathbf{s}\mathbf{s}^H + \mathbf{R}_\mathbf{n}$, this problem can be rewritten as follows,

$$\boldsymbol{\alpha}_o = \arg\min_{\boldsymbol{\alpha}} \; \boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{R} \boldsymbol{\Omega} \boldsymbol{\alpha} + \gamma (1 - \boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{s} - \mathbf{s}^H \boldsymbol{\Omega} \boldsymbol{\alpha}) \tag{3.9}$$

Notice that in (3.9) a real scalar function is optimized with respect to a complex vector. Therefore, in order to find the optimal solution $\frac{\partial \, \mathrm{MSE}(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}^H} = 0$ must be solved, [1]. Indeed, the optimization problem in (3.9) is analogous to the one involved in the theoretical LMMSE method, see (1.3), (1.5) and recall that $\mathbf{R} = \gamma \mathbf{s}\mathbf{s}^H + \mathbf{R}_\mathbf{n}$. Therefore, bearing in mind these statements and after easy manipulations, it is easy to check that the optimal shrinkage factors read,

$$\boldsymbol{\alpha}_o = \left(\boldsymbol{\Omega}^H \mathbf{R} \boldsymbol{\Omega} \right)^{-1} \gamma \boldsymbol{\Omega}^H \mathbf{s} \tag{3.10}$$

Now, recalling that $\boldsymbol{\Omega} \triangleq \left(\hat{\mathbf{R}}^{-1}\mathbf{s} \; , \; \mathbf{s} \right)$ and taking into account the property of multiplication of partitioned matrices [1], the expression (3.10) yields,

$$\boldsymbol{\alpha}_o = \gamma \begin{pmatrix} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} & \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s} \\ \mathbf{s}^H \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} & \mathbf{s}^H \mathbf{R} \mathbf{s} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} \\ 1 \end{pmatrix}$$

At this point, applying the definition of the inverse of a matrix and again applying the property of multiplication of partitioned matrices we obtain that the optimal shrinkage factors read,

$$\boldsymbol{\alpha}_o = \frac{\gamma \begin{pmatrix} \mathbf{s}^H \mathbf{R} \mathbf{s}\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s} \\ \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}\mathbf{s}^H \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} \end{pmatrix}}{\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s}\mathbf{s}^H \mathbf{R} \mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s}\mathbf{s}^H \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s}} \tag{3.11}$$

Which concludes the proof as (3.11) coincides with (3.2).

Proof of Lemma 3.2

The proof of this lemma is analogous to the one in Lemma 3.1, as the shrinkage filter $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$ in (3.3) is a particular case of the shrinkage filter $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$ in Lemma

3.1. Namely, let define $\boldsymbol{\psi} = \hat{\mathbf{R}}^{-1}\mathbf{s}$, then the MSE optimization problem in (3.3) may be reformulated as,

$$\alpha_o = \arg\min_{\alpha} \ \mathrm{MSE}\left(\mathbf{w} = \boldsymbol{\psi}\alpha\right)$$

Bearing in mind the expression of the MSE in (1.5) and recalling that according to assumption *(b)* in our data model (1.1), $\mathbf{R} = \gamma\mathbf{s}\mathbf{s}^H + \mathbf{R_n}$, this problem can be rewritten as,

$$\alpha_o = \arg\min_{\alpha} \ \alpha^H\boldsymbol{\psi}^H\mathbf{R}\boldsymbol{\psi}\alpha + \gamma(1 - \alpha^H\boldsymbol{\psi}^H\mathbf{s} - \mathbf{s}^H\boldsymbol{\psi}\alpha) \tag{3.12}$$

At this point, the optimal solution is found as in (3.9), i.e. it is the solution to $\frac{\partial\,\mathrm{MSE}(\boldsymbol{\alpha})}{\partial\boldsymbol{\alpha}^*} = 0$, which yields,

$$\alpha_o = \left(\boldsymbol{\psi}^H\mathbf{R}\boldsymbol{\psi}\right)^{-1}\gamma\boldsymbol{\psi}^H\mathbf{s} \tag{3.13}$$

Now, recalling that $\boldsymbol{\psi} = \hat{\mathbf{R}}^{-1}\mathbf{s}$, (3.13) leads to the next expression,

$$\alpha_o = \frac{\gamma\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}} \tag{3.14}$$

Which concludes the proof as (3.14) coincides with (3.4).

Proof of Theorem 3.1

The proof for Theorem 3.1 is readily obtained from Lemma 3.1 and Lemma 2.1. Namely, the claim that the estimator is realizable is evident from the expression of its shrinkage factors, (3.5). With regard to the consistency of the estimator it suffices to proof that the shrinkage factor $\breve{\boldsymbol{\alpha}}_{lb}$ in (3.5) is a consistent estimate of the optimal shrinkage vector $\boldsymbol{\alpha}_{lb}$ in (3.2), i.e. $\breve{\boldsymbol{\alpha}}_{lb} \asymp \boldsymbol{\alpha}_{lb}$ within the general asymptotics framework where $M, N \to \infty$ at a constant rate $M/N \to c \in (0, 1)$.

In order to proof that $\breve{\boldsymbol{\alpha}}_{lb} \asymp \boldsymbol{\alpha}_{lb}$, let use the asymptotic equivalences presented in Lemma 2.1, in the theoretical LMMSE shrinkage vector (3.2). This yields the next asymptotic equivalent expression for $\breve{\boldsymbol{\alpha}}_{lb}$,

$$\boldsymbol{\alpha}_{lb} \asymp \frac{\gamma\begin{pmatrix} \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}-(1-c)^{-1} \\ (1-c)^{-2}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}-\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(1-c)^{-1} \end{pmatrix}}{(1-c)^{-2}\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}-(1-c)^{-2}} \tag{3.15}$$

And, after straightforward manipulations, expression (3.15) may be rewritten as follows,

$$\boldsymbol{\alpha}_{lb} \asymp \frac{\gamma\begin{pmatrix} (1-c)^2\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}-(1-c) \\ c\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} \end{pmatrix}}{\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}-1} = \breve{\boldsymbol{\alpha}}_{lb} \tag{3.16}$$

Which concludes the proof, as it highlights that the shrinkage factor in (3.5) is a consistent estimate of the optimal shrinkage factor in (3.2) within the general asymptotics where $M, N \to \infty$ and $M/N \to c \in (0,1)$.

Proof of Theorem 3.2

The proof of this theorem follows the same guideline than the one for Theorem 3.1. Namely, the statement that the estimator is realizable is clear from the expression of its shrinkage factors (3.8), as they do not depend on any unknown parameter. With regard to the consistency, it suffices to proof that the shrinkage factor $\breve{\alpha}_l$ in (3.8) is a consistent estimate of the theoretical factors α_l in (3.4), i.e. we have to demonstrate that $\breve{\alpha}_l \asymp \alpha_l$ within the general asymptotics framework where $M, N \to \infty$ at a constant rate $M/N \to c \in (0,1)$. To this end, let use the asymptotic equivalences presented in Lemma 2.1, in the theoretical LMMSE shrinkage vector (3.4). This yields the next asymptotic deterministic equivalent expression for α_l,

$$\alpha_l \asymp \gamma(1-c)^2 \tag{3.17}$$

Which concludes the proof as according to (3.8) $\breve{\alpha}_l = \gamma(1-c)^2$ and as a consequence $\alpha_l \asymp \breve{\alpha}_l$.

3.5 Numerical simulations

This section is devoted to study the performance of the consistent shrinkage estimators proposed in (3.5) and (3.8) by means of numerical simulations. Namely, the performance of the proposed estimators, in terms of MSE, is compared to that of the optimal LMMSE and its traditional sample implementation in (1.4) and (1.9), respectively. According to the expressions of the estimators and the MSE in (1.5), the parameters controlling the simulations are c, $\frac{M}{N}$, $\hat{\mathbf{R}}$, $\mathbf{R}$, γ and $\mathbf{s}$. In order to specify the models for these parameters an array signal processing application is considered. Nonetheless, they are flexible enough to be applied to other fields of signal processing, e.g. in spectrum analysis. According to (1.1),

$$\mathbf{R} = \gamma \mathbf{s}\mathbf{s}^H + \mathbf{R_n} \tag{3.18}$$

Without loss of generality $\gamma \triangleq \mathbb{E}\left[|x(n)|^2\right]$ is set to 1 in all the simulations. Regarding the steering vector $\mathbf{s}$ associated to the parameter of interest $x(n)$, a uniform linear array (ULA) is assumed, i.e.,

$$[\mathbf{s}]_m = \frac{e^{j\pi \sin\theta_0 m}}{\sqrt{M}} \tag{3.19}$$

Where θ_0 is the Direction of Arrival (DOA) of the signal of interest and $\sqrt{M}$ is just a normalization factor yielding $\|\mathbf{s}\|^2 = 1$, see [1]. Moreover, for the simulation purposes θ_0 is set to 0. With regard to $\mathbf{R_n}$, a model that has been extensively used in signal processing, e.g. in spectrum estimation and array signal processing see [1] and [2], is,

$$\mathbf{R_n} = \mathbf{S}\mathbf{P}\mathbf{S}^H + \sigma^2\mathbf{I} \tag{3.20}$$

Where, $[\mathbf{S}]_{m,k} = \frac{e^{j\pi\sin\theta_k m}}{\sqrt{M}}$, $m = 0, \ldots, M-1$ is the antenna index, $k = 1, \ldots, K$ defines a set of interferers and θ_k is the DOA of the k-th interferer. $\mathbf{P}$ is the covariance matrix of the interferers and σ^2 is the power of an AWGN. For the simulations we consider, $\theta_k = (2 + 10(k-1))\frac{\pi}{180}$ with $k = 1, \ldots, K$; $K = M - 1$. $\mathbf{P}$ is considered to be diagonal and the elements of the diagonal are set according to $\sigma_k^2 = \gamma 10^{-SIR_k/10}$ $\forall k = 1, \ldots, K$. Where SIR_k is the ratio, in dB, between the power of the signal of interest and the power of the k-th interferer. The value for SIR_k depends on the simulation and will be specified below. With regard to σ^2 it is set to $\sigma^2 = \gamma 10^{-SNR/10}$, where SNR is the signal to noise ratio in dB and it will be specified below depending on the simulation. The parameter c

is set to $c = \frac{M}{N}$. Moreover, different values of M are considered herein $M = 6$, $M = 10$ and $M = 50$ corresponding to low, intermediate and high M. With regard to N it varies fulfilling that $M/N \in (0, 1)$. Finally, the sample correlation is

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{y}(n)\mathbf{y}^H(n) \tag{3.21}$$

And $\mathbf{y}(n)$ is generated according to the data model in (1.1) and taking into account the comments of the last paragraph, i.e.,

$$\mathbf{y}(n) = x(n)\mathbf{s} + \sum_{k=1}^{K} x_k(n)\mathbf{s}_k + \boldsymbol{\eta}(n),\ 1 \leq n \leq N \tag{3.22}$$

Where $\mathbf{s}_k$ is the k-th column of $\mathbf{S}$, $x_k(n)$ is the signal associated to the $k - th$ interferer and $\boldsymbol{\eta}(n)$ is the noise vector. Moreover, $\mathbf{y}(n)$ is assumed to be iid among samples and distributed according to a multivariate complex gaussian, namely,

$$\begin{gathered} \mathbf{y}(n) \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}) \\ x(n) \sim \mathcal{CN}(0, \gamma), x_k(n) \sim \mathcal{CN}(0, \sigma_k^2), \boldsymbol{\eta}(n) \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}). \end{gathered} \tag{3.23}$$

Next, with the simulation conditions at hand, the performance of the estimators proposed in this chapter is assessed. Thus, the MSE of the proposed shrinkage LMMSE estimators in (3.5) and (3.8) is compared to the MSE of the theoretical LMMSE estimator (1.4) and its sample implementation (1.9), when N varies to simulate different sample size regimes, i.e. within the range where $\frac{M}{N} \in (0, 1)$. Namely, the MSE of the estimators is computed by substituting the expression of the filter of each estimator in (1.5). In figure 3.1 this simulation is carried out for a high observation dimension, namely M is set to 50. In this situation the proposed shrinkage methods in (3.5) and (3.8) tend to be the optimal shrinkage estimators, i.e. they are (M, N)-consistent estimates of the optimal though unrealizable methods in (3.2) and (3.4), respectively. Moreover the SNR is set to 5 dB and the SIR_k to 10 dB. Observing the plot one can see that the performance of the shrinkage of the sample LMMSE towards the Bartlett fitler is extraordinary. Indeed it is almost the same than the performance of the theoretical LMMSE, which is the lower bound. Moreover, this figure highlights that the proposed methods outperform the sample LMMSE for any of the sample sizes considered herein i.e. $\frac{M}{N} \in (0, 1)$, specially in the small sample size regime where the improvement is huge. Moreover, they proof to be robust to small sample size situations, i.e. when $\frac{M}{N} \to 1$. It is also interesting to observe that for $\frac{M}{N} \to 0$ the shrinkage, the theoretical and the sample LMMSE estimators tend to converge. This is

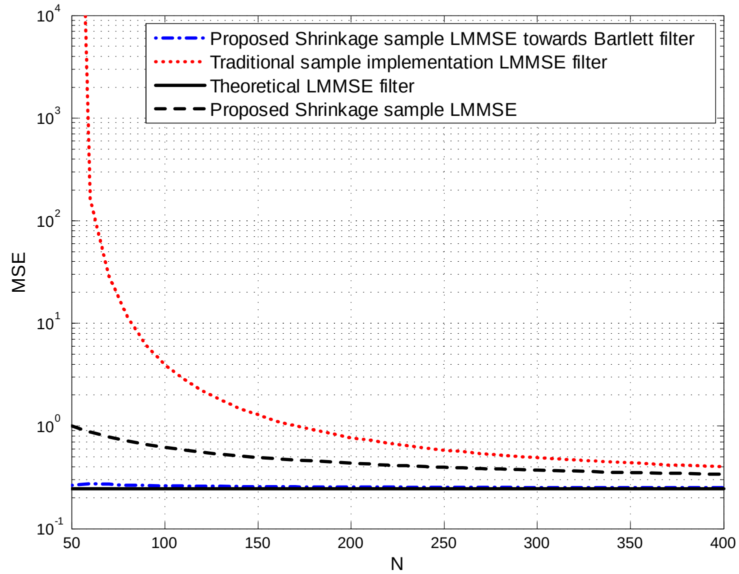

Figure 3.1: Performance comparison between proposed Shrinkage LMMSE estimators in (3.5) and (3.8), theoretical LMMSE estimator (1.4) and its sample based implementation (1.9), when SNR= 5 dB, M=50 and SIR_i=10 dB.

because in this case the SCM is the MVUE of $\mathbf{R}$ and it is well conditioned and as consequence the sample LMMSE tends to the optimal LMMSE estimator. The shrinkage estimators are aware of this situation and reflect it by means of the shrinkage factors, which lead to obtain the sample LMMSE, as it was noted in (3.7) and as it can be readily inferred from (3.8). Moreover observe that among the proposed shrinkage estimators, the shrinkage of the sample LMMSE towards the Bartlett filter behaves better than direct shrinking the sample LMMSE. This is because the former incorporates more information about the problem. Namely, as in (3.8), it incorporates through the sample LMMSE information obtained from the measures and also a priori information consisting of the power and the DOA of the parameter of interest. Nonetheless, it also incorporates more a priori information consisting of the knowledge that for small sample size regime the Bartlett filter may give better performance than the sample LMMSE. I.e. (3.5) optimally combines the sample LMMSE and the Bartlett filter by giving more weight to the sample LMMSE in large sample size situations and more weight to the Bartlett filter in small sample size regimes. This behavior will be further investigated in figure 3.6.

Next in figure 3.2 the same type of simulation than for figure 3.1 is carried out, but for an intermediate observation dimension. Namely, M is set to 10, the SNR is set to 5 dB and the SIR_k to 0 dB. The aim is to assess the impact in the performance of the methods of lowering M to an intermediate value, this study is completed in the upcoming figures, where M is further reduced. One can see that the proposed methods are still robust to the small sample size regime and that outperform the sample LMMSE in any sample size regime, specially in the small sample size situation. Moreover, one can see that the proposed shrinkage of the sample LMMSE towards the Bartlett filter still clearly outpeforms the other proposed method, consisting of a direct shrinkage of the sample LMMSE, though the difference is reduced in the small sample size regime. The reason for this behavior is that the latter shrinks the sample LMMSE towards zero and as a consequence its performance for $\frac{M}{N} \to 1$ is almost the same whatever the value of M is. On the other hand, the other proposed method shrinks the sample LMMSE towards the Bartlett filter, cf. see (3.7), and therefore its performance for $\frac{M}{N} \to 1$ depends on the value of M, and also SIR_k as it will be shown in the upcoming figures.

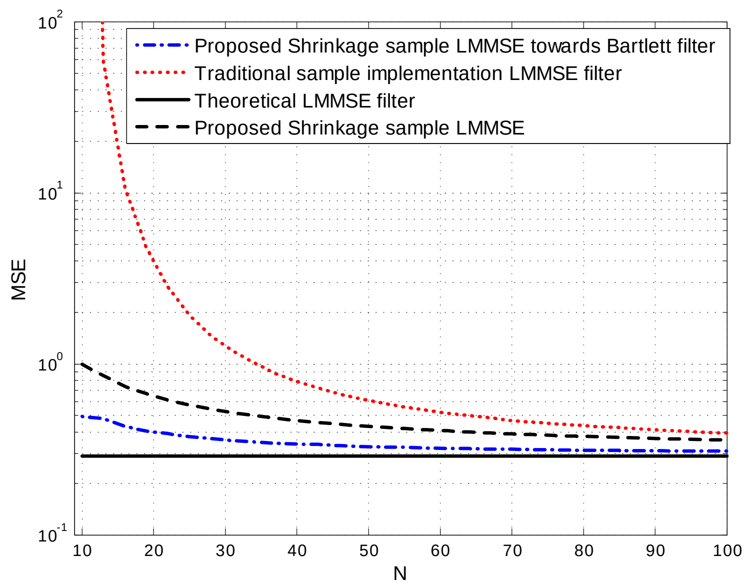

Figure 3.2: Performance comparison between proposed Shrinkage LMMSE estimators in (3.5) and (3.8), theoretical LMMSE estimator (1.4) and its sample based implementation (1.9), when SNR= 5 dB, M=10 and SIR$_i$=10 dB.

In figure 3.3 the influence of the power of the interferers on the performance of the estimators is studied. Thus, the same simulation than for figure 3.2 is carried out, but now for a stronger interference power, namely now SIR_k is set to 0 dB. Observe that in this situation, the improvement in performance when considering the shrinkage of the sample LMMSE towards the Bartlett filter (3.5) instead of only the shrinkage of the sample LMMSE (3.8) is reduced. This is because in the previous figure the additional a priori information incorporated in the former, by means of the Bartlett filter, was more valuable. Namely, the Bartlett filter is optimum when considering that only the signal of interest and additive white noise are present in the scenario, which is equivalent to say that SIR_k is large. And the previous figure was closer to that situation as the SIR_k was set to 10 dB and now to 0 dB. Moreover, regarding the performance comparison between the proposed shrinkage estimators, the sample and the theoretical LMMSE the same comments than for figure 3.2 can be extracted.

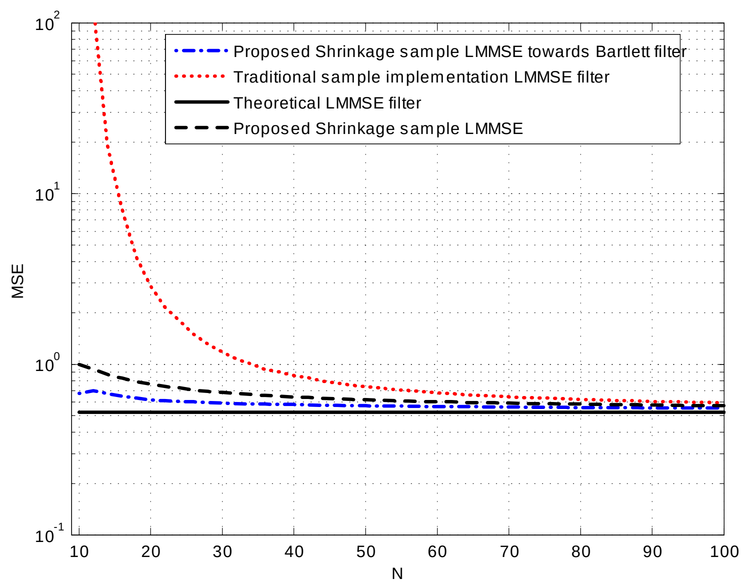

Figure 3.3: Performance comparison between proposed Shrinkage LMMSE estimators in (3.5) and (3.8), theoretical LMMSE estimator (1.4) and its sample based implementation (1.9), when SNR= 5 dB, M=10 and SIR_i=0 dB.

Next, in figures 3.4 and 3.5 the effect of a low observation dimension, i.e. low M, on the performance of the proposed shrinkage methods is studied, namely M is set to 6. This study is important because these estimators were obtained based on the asymptotic equivalences in Lemma 2.1, that are obtained within the asymptotic regime were $M, N \to \infty$ at a constant rate c. The rest of simulation parameters for figures 3.4 and 3.5 are the same than for figures 3.2 and 3.3, respectively. The simulations results highlight that the proposed shrinkage estimators are still robust to the small sample size regime and that outperform the sample LMMSE in all the sample size regimes, specially for $\frac{M}{N} \to 1$ where the improvement is huge. Moreover, one can see that the proposed method based on shrinking the sample

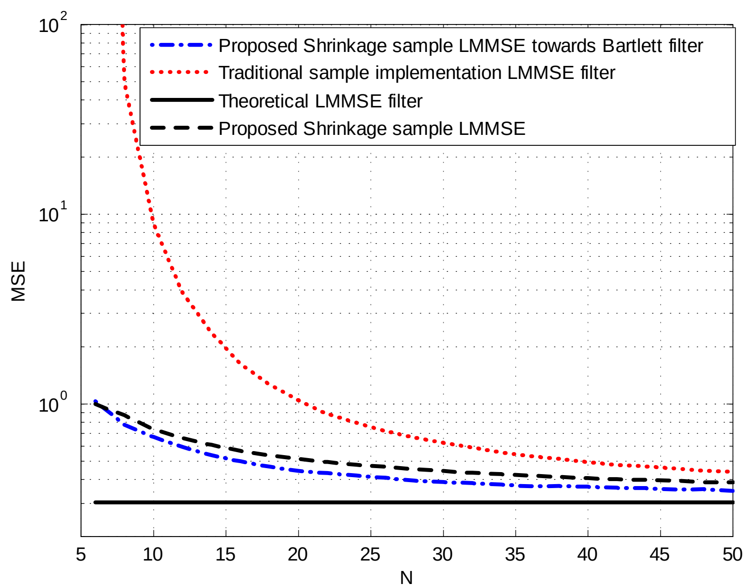

Figure 3.4: Performance comparison between proposed Shrinkage LMMSE estimators in (3.5) and (3.8), theoretical LMMSE estimator (1.4) and its sample based implementation (1.9), when SNR= 5 dB, M=6 and SIR$_i$=10 dB.

LMMSE towards the Bartlett filter slightly outperforms the other proposed method, based on direct shrinking the sample LMMSE, though the gain in performance is reduced compared to previous figures. The rationale for this behavior is the same than for figures 3.3.and 3.2. I.e. the latter shrinks the sample LMMSE towards zero and thus its performance for $\frac{M}{N} \to 1$ is almost the same whatever the value of M is, whereas the former shrinks the sample LMMSE towards the Bartlett filter and as a consequence its performance for $\frac{M}{N} \to 1$ depends on the value of M and SIR_k.

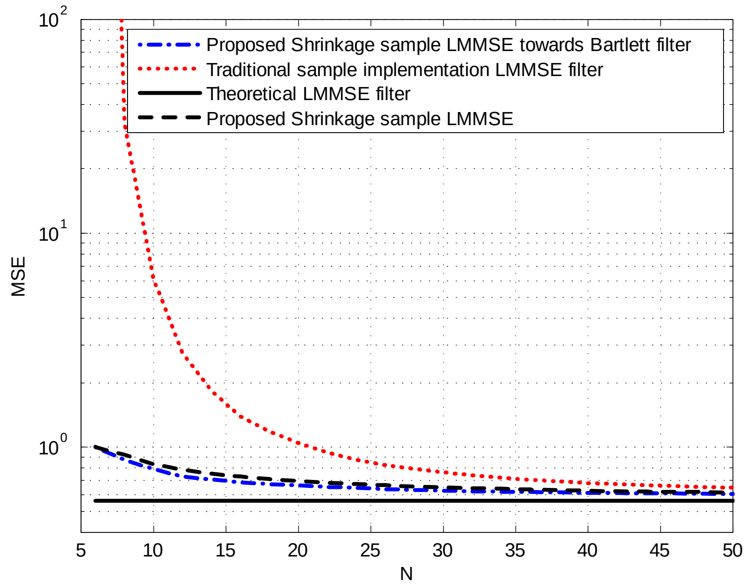

Figure 3.5: Performance comparison between proposed Shrinkage LMMSE estimators in (3.5) and (3.8), theoretical LMMSE estimator (1.4) and its sample based implementation (1.9), when SNR= 5 dB, M=6 and SIR$_i$=0 dB.

Finally, in figure 3.6, the shrinkage effect is exemplified. Namely, we run a montecarlo simulation to plot $|\breve{\alpha}_{lb,1}|^2$ and $|\breve{\alpha}_{lb,2}|^2$ of the proposed method in (3.5), consisting of shrinking the sample LMMSE towards the Bartlett filter. In this case the simulation conditions are the same than for figure 3.3, i.e. SNR= 5 dB, $M = 10$ and SIR$_i$ $= 0$ dB and N varies fulfilling $\frac{M}{N} \in (0, 1)$. Recall that the proposed shrinkage filter reads $\breve{\mathbf{w}}_{lb,s}{=}\breve{\alpha}_{lb,1}\hat{\mathbf{R}}^{-1}\mathbf{s}{+}\breve{\alpha}_{lb,2}\mathbf{s}$. Moreover, recall that the behavior of this filter is as follows. On the one hand when the sample size increases, i.e. $\frac{M}{N}$ decreases, $\breve{\mathbf{w}}_{lb,s}$ tends to give more weight to the sample LMMSE than to the Bartlett filter. Indeed when $\frac{M}{N} \to 0$ the proposed filter $\breve{\mathbf{w}}_{lb,s}$ tends

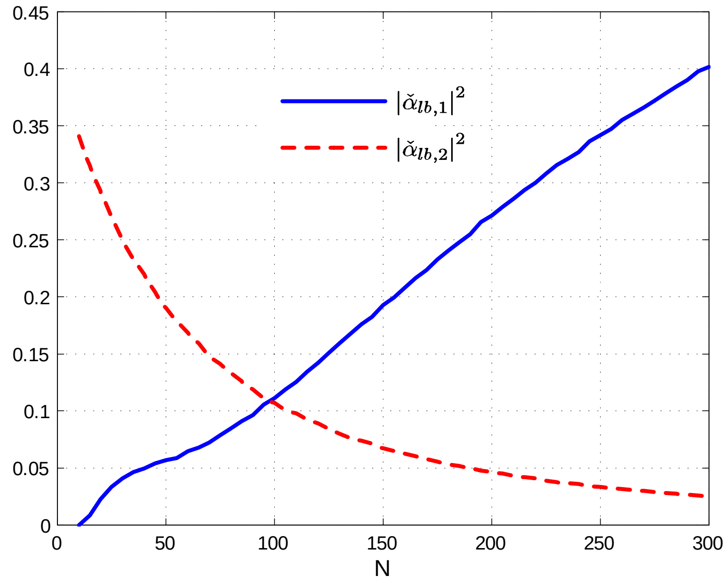

Figure 3.6: Shrinkage Factors of the proposed shrinkage method in (3.5) when $M = 10$, SNR= 5 dB and SIR$_i$=0 dB.

to disregard the Bartlett filter and give most of the weight to the sample LMMSE. This is because the sample LMMSE is the optimal filter for the large sample size regime, see also (3.6). And effectively, figure 3.6 highlights this behavior, as $\frac{M}{N}$ decreases $|\breve{\alpha}_{lb,1}|^2$ tends to increase whereas $|\breve{\alpha}_{lb,2}|^2$ tends to decrease. On the other hand, as in general in the small sample size regime the Bartlett filter yields better performance than the sample LMMSE, $\breve{\mathbf{w}}_{lb,s}$ has the next behavior. As $\frac{M}{N}$ increases, $\breve{\mathbf{w}}_{lb,s}$ tends to give more weight to the Bartlett filter than to the sample LMMSE. Indeed in the extreme case where $\frac{M}{N} \to 1$, the proposed filter $\breve{\mathbf{w}}_{lb,s}$ tends to disregard the sample LMMSE and give most of the weight to the Bartlett filter. And effectively figure 3.6 highlights this behavior as well. Namely, As $\frac{M}{N}$ increases, $|\breve{\alpha}_{lb,2}|^2$ tends to increase whereas $|\breve{\alpha}_{lb,1}|^2$ tends to decrease.

Chapter 4

Optimal shrinkage for large sample MVDR

4.1 Introduction

The aim of this chapter is to design an estimator that overcomes the drawbacks of the sample MVDR method exposed in (1.9). Namely, the aim is to obtain an estimator which is robust to the small sample size regime and that preserves the optimality of the sample MVDR in the large sample size. To this end, we have at our disposal two powerfool tools, shrinkage estimation and random matrix theory, namely G-estimation. More specifically, the approach is analogous to the design of the estimators proposed in chapter 3 to overcome the drawbacks of the sample LMMSE, and can be summarized as follows. First, in section 4.2 we propose to use a shrinkage of the sample MVDR estimator towards a Bartlett filter. This structure permits to combine the optimality of the sample MVDR for large sample size and avoid its performance degradation in the small sample size as in this situation the estimator is shrunk towards the Bartlett filter $\mathbf{w} = \mathbf{s}$. Moreover, the optimal shrinkage factor is obtained as the result of optimizing the MSE subject to the common MVDR constraint, i.e. $\mathbf{w}^H\mathbf{s} = 1$. Unfortunately, this optimal shrinkage factor depends on the unknown correlation matrix. In order to circumvent this problem, in section 4.3, we propose to apply random matrix theory to obtain a consistent estimate of that shrinkage factor in the doubly asymptotic regime where both the sample size N and the observation dimension M grow at a fixed rate. Thus, by means of the use of random matrix theory not only we are obtaining a consistent and realizable estimator, but also we are implicitly tackling the situation where M and N may be comparable. I.e. the robustness of the designed estimator to the small sample size regime is due to both the shrinkage and the RMT tools. Indeed, random matrix theory permits to obtain an optimal shrinkage factor when both M and N tend to infinity at a fixed rate. Moreover, the numerical simulations highlight that the proposed shrinkage method outperforms the conventional sample MVDR method in any of the sample sizes considered herein, namely $M/N \to c \in (0,1)$, being the improvement dramatic when we approach the small sample size regime, i.e. when $M/N \to 1$.

4.2 Optimal shrinkage of the sample MVDR estimator

As it was pointed out in chapter 1, the performance of the sample MVDR (1.9) is rapidly degraded when the sample size N is compared to the observation dimension M. This is due to its strategy of replacing $\mathbf{R}^{-1}$ by its sample estimate $\hat{\mathbf{R}}^{-1}$ relying on the fact that the sample correlation is the ML estimator of the correlation. Nonetheless, when N is comparable to M, the sample estimate $\hat{\mathbf{R}}^{-1}$is no longer a good estimate. Indeed, it is an ill conditioned estimator, see e.g. [25], this means that when $N \approx M$ and $M \geqslant N$ inverting $\hat{\mathbf{R}}$ dramatically amplifies the estimation error. In fact, when $M < N$ the sample correlation matrix is not even invertible. On the other hand, when $M \gg N$ the sample MVDR is optimal as $\hat{\mathbf{R}}$ is the MVUE estimator of $\mathbf{R}$. Therefore in order to overcome the small sample size degradation of the sample MVDR and mantain its optimality in the large sample size we propose a shrinkage of the sample MVDR filter towards a Bartlett filter,

$$\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$$

Note that the original expression of the sample MVDR is $\mathbf{w} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}}$, but the denominator is just a normalization quantity. Therefore for the shrinkage structure purposes, we can consider $\alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s}$ or $\alpha \frac{\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}}$ and we prefer to use the former to avoid cumbersome calculations. According to the shrinkage estimation theory, in order to improve a sample estimator, whose estimation error mostly comes from the estimation variance, we may introduce a bias such that the overall estimation error is diminished. Note that this role is played by the term $\alpha_2 \mathbf{s}$. I.e. we are shrinking the sample MVDR filter $\mathbf{w} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}}$ towards $\mathbf{w} = \mathbf{s}$. Indeed, this latter filter is obtained by substituting in the theoretical MVDR the correlation matrix by an estimate consisting of the identity matrix, which is a biased estimation of the theoretical MVDR, $\mathbf{w} = \frac{\mathbf{R}^{-1}\mathbf{s}}{\mathbf{s}^H \mathbf{R}^{-1}\mathbf{s}}$, but which has not estimation variance.

With the filter structure at hand, $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, we may formulate the problem that permits the design of the optimal linear estimator of $x(n)$ in (1.1), when there is not knowledge about its second moment. I.e. the aim is to solve the problem stated in (1.11), when we impose the constraint $\mathbf{w} \in \mathcal{C} \Leftrightarrow \mathbf{w}^H\mathbf{s} = 1$ in the filter to avoid the dependence on the second moment of the parameter to estimate, when $f(\mathrm{MSE}(\mathbf{w})) = \mathrm{MSE}(\mathbf{w})$ and when the assumptions about the linear model of the observed signal (1.1) are *(a)-(d)*. This problem and its solution are formalized in the next lemma.

Lemma 4.1 *Assume that a set of observations $\{\mathbf{y}(n)\}_{n=1}^{N}$ fulfilling the model in (1.1), with assumptions (a)-(d) is available. Given $\{\mathbf{y}(n)\}_{n=1}^{N}$, consider the problem of estimating the unknown $x(n)$ in (1.1), based on minimizing the MSE in (1.5), when the estimator $\hat{x}_{c,s}(n) = \mathbf{w}_{c,s}^H\mathbf{y}(n)$ is a linear shrinkage of the sample MVDR towards a Bartlett filter, i.e. $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2\mathbf{s}$, and when the constraint $\mathbf{w}^H\mathbf{s} = 1$ is imposed in the filter to avoid the lack of knowledge about the second moment of $x(n)$. This problem is mathematically formulated as follows,*

$$\hat{x}_{c,s}(n) = \mathbf{w}_{c,s}^H\mathbf{y}(n); \qquad \begin{array}{c} \mathbf{w}_{c,s} = \arg\min\limits_{\mathbf{w}} \ \mathrm{MSE}(\mathbf{w}) \\ s.t. \mathbf{w}^H\mathbf{s} = 1, \mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2\mathbf{s} \end{array} \tag{4.1}$$

Then, defining $\boldsymbol{\alpha}_c \triangleq (\alpha_1 \ , \ \alpha_2)^T$, the optimal solution for this problem is given by the next shrinkage factors,

$$\boldsymbol{\alpha}_c = \frac{\begin{pmatrix} \mathbf{s}^H\mathbf{R}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \\ \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \end{pmatrix}}{G} \tag{4.2}$$

Where, $G \triangleq \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(\mathbf{s}^H\mathbf{R}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}) - \mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} + \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$

Proof: See section 4.4.

■

As it was carried out for the shrinkage of the sample LMMSE, it is interesting to study the particular case of direct shrinkage of the sample MVDR. I.e. $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}$ in lemma 4.1. Following the same procedure than for the proof of lemma 4.1, this leads to obtain the next optimal shrinkage factor, in an MSE sense,

$$\alpha_{c,p} = \frac{1}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}} \tag{4.3}$$

Thus, the optimal shrinkage filter reads $\mathbf{w} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}}$, i.e. coincides with the sample MVDR. Therefore, direct shrinkage of the sample MVDR does not help to improve its performance. Moving on to something else, the expression for the optimal shrinkage MVDR estimator in (4.2) highlights the dependance on the unknown $\mathbf{R}$ and as a consequence that it is not realizable. At this point, in other contexts dealing with shrinkage estimation and facing an analogous problem the authors propose to substitute the unknown $\mathbf{R}$ for its sample estimate [33]. Nonetheless, that approach entails an estimation risk that may lead to a performance degradation. Indeed, if one applies this strategy to (4.2), it turns out that obtains the sample MVDR, i.e. $\boldsymbol{\alpha}_{c|\mathbf{R}=\hat{\mathbf{R}}} = \left(\frac{1}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}}, 0\right)^T$, and as a consequence the potential benefits of shrinkage estimation disappear. Herein, in order to tackle this problem and obtain a realizable method, another strategy is proposed. We propose to use random matrix theory, or more specifically G-estimation, to obtain an *(M,N)*-consistent estimate of the optimal shrinkage factor in (4.2). Namely, the general asymptotics framework where $M\ ,N \to \infty$ at a constant rate $M/N \to c \in (0,1)$ is adopted as it is enough general to study the consistency for different sample sizes. E.g. it deals with the situations where M may be comparable to N, i.e. small sample size and it embraces the classical large sample size assumption for obtaining consistent estimators, where the observation dimension M is fixed and the sample size N is assumed to tend to infinity. This powerful approach is presented in the next section.

4.3 Optimal shrinkage for large sample MVDR

In this section an *(M,N)*-consistent estimate of the optimal, though unrealizable, shrinkage factors of the method proposed in (4.2) is exposed and relevant comments are discussed. This method, based on recent results from random matrix theory, is presented in the next theorem,

Theorem 4.1 *Let define $\check{\boldsymbol{\alpha}}_c \triangleq (\check{\alpha}_{c,1}\ ,\ \check{\alpha}_{c,2})^T$, and let assume the normalization $\|\mathbf{s}\|^2 = 1$, then a realizable and (M,N)-consistent estimate of the optimal shrinkage MVDR estimator (4.2), within the general asymptotics framework where $M, N \to \infty$ at a constant rate $M/N \to c \in (0,1)$, reads as follows,*

$$\check{x}_{c,s}(n) = \check{\mathbf{w}}_{c,s}^H\mathbf{y}(n); \check{\mathbf{w}}_{c,s} = \check{\alpha}_{c,1}\hat{\mathbf{R}}^{-1}\mathbf{s} + \check{\alpha}_{c,2}\mathbf{s}$$

$$\check{\boldsymbol{\alpha}}_c = \frac{\left(\begin{array}{c}(1-c)(\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(1-c)-1)\\ c\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\end{array}\right)}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(1-c)^2-2(1-c)+1)} \tag{4.4}$$

Proof: See section 4.4.

■

This *(M,N)*-consistent estimator, shows implicitly the robustness to the small sample size regime. On the one hand, its underlying structure corresponds to a shrinkage estimator, which are known to be robust to the small sample size, see chapter 2. On the other hand, as we will see in the proof of this theorem, it relies on random matrix theory, i.e. the consistency is obtained within the framework of general asymptotics, that embraces the scenario $M/N \to 1$. Moreover, in the numerical results section, the robustness to the small sample size will be studied in more detail. Furthermore, in that section it will be demonstrated that (4.4) outperforms the conventional sample MVDR estimator (1.9) in all the sample size regimes dealt with herein, i.e $M/N \in (0,1)$ and that the improvement in performance is dramatic when $M/N \to 1$.

At this point, it is interesting to study the asymptotic expression of the consistent shrinkage MVDR filter in Theorem 4.1 when $c \to 1$ and $c \to 0$, as it gives more insights about the designed estimator. In the former case, after straightforward manipulations the next asymptotic expression is obtained,

$$c \to 1 \Rightarrow \check{\mathbf{w}}_{c,s} \to \mathbf{s} \tag{4.5}$$

This is a meaningful result as in the small sample size regime the sample implementation of the MVDR is no longer a good estimate and in general may display worse performance than an estimator based on a Bartlett filter, which does not use any information about the available samples. On the other hand, in the large sample size regime, i.e. when $c \to 0$, it is easy to obtain that the consistent shrinkage MVDR filter tends to the traditional sample implementation,

$$c \to 0 \Rightarrow \check{\mathbf{w}}_{c,s} \to \frac{\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}} \tag{4.6}$$

This is a meaningful result as in this case we are in the framework of classical asymptotics that is commonly assumed to obtain the sample MVDR, i.e. M fixed and N tending to infinity. I.e in this situation $\hat{\mathbf{R}}$ is the MVUE of $\mathbf{R}$, it is well conditioned and it is consistent and as a consequence the sample implementation of the MVDR tends to the theoretical filter (1.7). Another interesting approach is the bayesian point of view, that usuallly is given in shrinkage estimation. As $c \to 1$ the amount of information obtained from the measured samples is lower and it is more convenient that the shrinkage filter tend to the "a priori" information about the filter represented by $\mathbf{w} \propto \mathbf{s}$. On the other hand, as $c \to 0$, the amount of information obtained from the measured samples is much more relevant than the a priori information and therefore it is logic that the shrinkage filter tend to $\mathbf{w} \propto \frac{\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}}$.

4.4 Proofs

In this section we provide the proofs of the main results of this chapter, namely Lemma 4.1 and Theorem 4.1.

Proof of Lemma 4.1

First, let substitute the expression (1.5) of the MSE in the optimization problem (4.1) of Lemma 4.1. This leads to obtain the next expression,

$$\begin{aligned} \mathbf{w}_o = \arg\min_{\mathbf{w}} \ & \mathbf{w}^H\mathbf{R}_\mathbf{n}\mathbf{w}+\gamma\left|1-\mathbf{w}^H\mathbf{s}\right|^2 \\ s.t. \mathbf{w}^H\mathbf{s} =&1, \mathbf{w} =\alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}+\alpha_2\mathbf{s} \end{aligned} \tag{4.7}$$

Now, let apply the constraint $\mathbf{w}^H\mathbf{s} =1$ to the objective function, which reduces it to $\mathbf{w}^H\mathbf{R}_\mathbf{n}\mathbf{w}$. After this, observe that as $\mathbf{R} =\gamma\mathbf{s}\mathbf{s}^H + \mathbf{R}_\mathbf{n}$, the resulting problem is not affected if the considered objective function is $\mathbf{w}^H\mathbf{R}\mathbf{w}$. Therefore, (4.7) can be reformulated as,

$$\begin{aligned} \mathbf{w}_o = \arg\min_{\mathbf{w}} \ & \mathbf{w}^H\mathbf{R}\mathbf{w} \\ s.t. \mathbf{w}^H\mathbf{s} =&1, \mathbf{w} =\alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}+\alpha_2\mathbf{s} \end{aligned} \tag{4.8}$$

At this point, let define $\boldsymbol{\alpha} \triangleq (\alpha_1 \ , \ \alpha_2)$, $\boldsymbol{\Omega} \triangleq \left(\hat{\mathbf{R}}^{-1}\mathbf{s} \ , \ \mathbf{s}\right)$ and let $\boldsymbol{\alpha}_o$ denote the optimal shrinkage factors. Then, the optimization problem (4.8) can be rewritten as a function of $\boldsymbol{\alpha}$,

$$\begin{aligned} \boldsymbol{\alpha}_o = \arg\min_{\boldsymbol{\alpha}} \ & \boldsymbol{\alpha}^H\boldsymbol{\Omega}^H\mathbf{R}\boldsymbol{\Omega}\boldsymbol{\alpha} \\ s.t. \boldsymbol{\alpha}^H\boldsymbol{\Omega}^H\mathbf{s} =&1 \end{aligned} \tag{4.9}$$

Observe that this optimization problem is analogous to the one involved in the MVDR estimator (1.6). Therefore, following the same procedure, i.e. using the method of Lagrange multipliers [72], we readily obtain that the optimum shrinkage factors are,

$$\boldsymbol{\alpha}_o = \frac{\left(\boldsymbol{\Omega}^H \mathbf{R} \boldsymbol{\Omega}\right)^{-1} \boldsymbol{\Omega}^H \mathbf{s}}{\left(\boldsymbol{\Omega}^H \mathbf{s}\right)^H \left(\boldsymbol{\Omega}^H \mathbf{R} \boldsymbol{\Omega}\right)^{-1} \boldsymbol{\Omega}^H \mathbf{s}} \tag{4.10}$$

At this point, applying to (4.10) the property of multiplication of partitioned matrices [1], bearing in mind that $\boldsymbol{\Omega} \triangleq \left(\hat{\mathbf{R}}^{-1}\mathbf{s}\ ,\ \mathbf{s}\right)$ and after straightforward manipulations, the next expression is obtained for the optimal shrinkage factors,

$$\boldsymbol{\alpha}_o = \frac{\begin{pmatrix} \mathbf{s}^H \mathbf{R} \mathbf{s} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s} \\ \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} \mathbf{s}^H \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} \end{pmatrix}}{G} \tag{4.11}$$

Being $G = \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} (\mathbf{s}^H \mathbf{R} \mathbf{s} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s}) - \mathbf{s}^H \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} + \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s}$. Which concludes the proof, as (4.11) is equal to expression (4.2) in Lemma 4.1.

Proof of Theorem 4.1

The claim that (4.4) is a realizable estimator follows from its expression. With regard to the consistency, the proof is readily obtained from Lemma 4.1, which provides the optimal shrinkage of the sample MVDR towards a Bartlett filter, and Lemma 2.1, which are a set of results from random matrix theory that pave the way to study the consistency of that optimal filter within the general asymptotics framework. Namely, in order to prove Theorem 4.1 it must be shown that $\check{\boldsymbol{\alpha}}_c$ in (4.4) is a consistent estimate of the theoretical shrinkage factor $\boldsymbol{\alpha}_c$ in (4.2). In order to attain this aim let use the RMT results in Lemma 2.1 in the theoretical MVDR shrinkage vector (4.2). This leads to obtain the next asymptotic equivalence for $\boldsymbol{\alpha}_c$,

$$\boldsymbol{\alpha}_c \asymp \frac{\begin{pmatrix} \mathbf{s}^H \hat{\mathbf{R}} \mathbf{s} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} - (1-c)^{-1} \\ (1-c)^{-2} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} (1-c)^{-1} \end{pmatrix}}{\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} (\mathbf{s}^H \hat{\mathbf{R}} \mathbf{s} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} - 2(1-c)^{-1} + (1-c)^{-2})} \tag{4.12}$$

Now, after straightforwards manipulations, one obtains that the quantitiy in (4.12) is asymptotically equivalent to the next expression, within the general asymptotics where $M, N \to \infty$ and $M/N \to c \in (0, 1)$,

$$\boldsymbol{\alpha}_c \asymp \frac{\left(\begin{array}{c}(1-c)(\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(1-c)-1)\\ c\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\end{array}\right)}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(1-c)^2-2(1-c)+1)} = \check{\boldsymbol{\alpha}}_c \tag{4.13}$$

This highlights that $\check{\boldsymbol{\alpha}}_c$ is asymptotically equivalent to $\boldsymbol{\alpha}_c$. I.e. $\check{\boldsymbol{\alpha}}_c$ in Theorem 4.1 is an *(M,N)*-consistent estimator of $\boldsymbol{\alpha}_c$ in Lemma 2.1, within the general asymptotics where $M, N \to \infty$ at a constant rate $M/N \to c \in (0,1)$, and as a consequence the proof is concluded.

4.5 Numerical simulations

In this section the performance of the estimator proposed in this chapter, based on shrinking the sample MVDR (4.4), is compared to the conventional sample MVDR estimator (1.9) and the theoretical MVDR method (1.7) by means of numerical simulations. More specifically, the performance is analyzed in terms of the MSE of each method. The aim is to gain more insights about the performance of the proposed shrinkage method for any of the sample size situations considered herein, i.e. $M/N \in (0,1)$. Namely, it will be interesting to observe its robustness to the small sample size regime, whether it dominates the sample MVDR within the set $M/N \in (0,1)$, the shrinkage effect and whether it remains close to the performance of the theoretical MVDR for any sample size regime. In order to conduct the simulations the same simulation environment than in chapter 3 is considered. I.e. an array signal processing is considered to specify the value of the simulation parameters. These are c, $\frac{M}{N}$, $\hat{\mathbf{R}}$, $\mathbf{R}$, and $\mathbf{s}$. The model for $\hat{\mathbf{R}}$, $\mathbf{R}$, and $\mathbf{s}$ is specified in equations (3.18) to (3.23). Moreover, M is considered to be fixed and the sample size N to be variable to emulate any of the sample size regimes considered herein, i.e. $M/N \in (0,1)$. Three values of M are considered herein 50, 10 and 6, correponding to a high, intermediate and low observation dimension, respectively. Next the performance comparison of the proposed shrinkage of the sample MVDR, the sample MVDR and the theoretical MVDR is exposed. To this end, the expressions of their filters (4.4), (1.9) and (1.7), respectively, have been substituted in the expression of the MSE (1.5).

In figure 4.1 the performance of the proposed shrinkage MVDR method is compared to the theoretical and the sample MVDR for a high observation dimension, namely M is set to 50. This simulation is interesting because it paves the way to get more insights about the performance upper bound of the proposed method compared to the theoretical and the

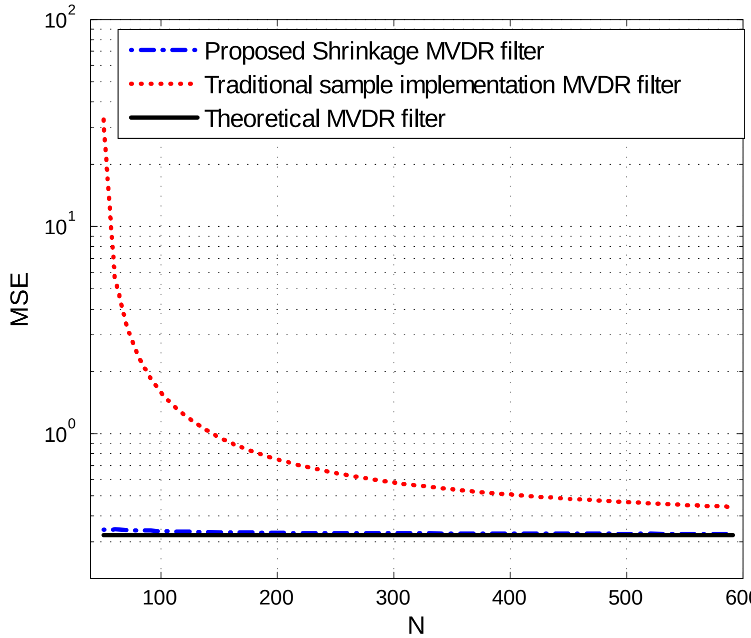

Figure 4.1: Performance comparison between proposed Shrinkage MVDR estimator in (4.4) , theoretical MVDR estimator (1.7) and its sample based implementation (1.9), when SNR= 5 dB, M=50 and SIR$_i$=10 dB.

sample MVDR. The reason is that for a high observation dimension M, the proposed method tends to be the optimal shrinkage estimator as it is an (M, N)-consistent estimate of the optimal though unrealizable method in (4.2). Moreover the SNR is set to 5 dB and the SIR_k to 10 dB. Observing this figure one can see the overwhelming superiority of the proposed shrinkage MVDR method compared to the sample MVDR. The improvement in performance is huge for small and intermediate sample size regimes. Indeed the proposed method outperforms the sample MVDR in any of the sample size regimes considered herein, i.e. $M/N \in (0, 1)$. Moreover, in figure 4.1 one can see that the proposed method remains very close to the theoretical MVDR method for any sample size situation and that it is robust to the small sample regime. Figure 4.1 also suggests that for $\frac{M}{N} \to 0$ the shrinkage, the theoretical and the sample LMMSE estimators tend to converge, this behavior will be even more clearer in the upcoming figures. This behavior is due to the optimality of the sample MVDR for large sample size regimes. Namely, in this situation the SCM is the

MVUE of **R** and it is a well conditioned estimator of **R** and as consequence the sample MVDR tends to the theoretical MVDR estimator. The shrinkage estimators are aware of this situation and reflect it by means of the shrinkage factors, which lead to obtain the sample MVDR, as it was noted in (4.6).

In figure 4.2, the observation dimension is fixed to an intermediate value, namely $M = 10$ antennas. The effect of reducing M respect to figure 4.1 is important, as the proposed method was envisaged for a large M, i.e. it is an (M, N)-consistent estimate of (4.2) and as a consequence is optimal for large M. Therefore, it is important to get more insights about whether the proposed method is robust to reduce M. This study will be completed with figures 4.4 and 4.5, where M is even lower. Moreover, in the current figure rather strong interferences are considered to be present in the scenario. Namely the ratio between the power of the signal of interest and the power of the i-th interferer is set to SIR_i=0 dB for any i. Furthermore, the SNR is fixed to 5 dB. This figure highlights that although M is reduced until an intermediate value of 10, respect to the previous figure, the proposed method still outperforms clearly the sample MVDR, specially in the small sample size

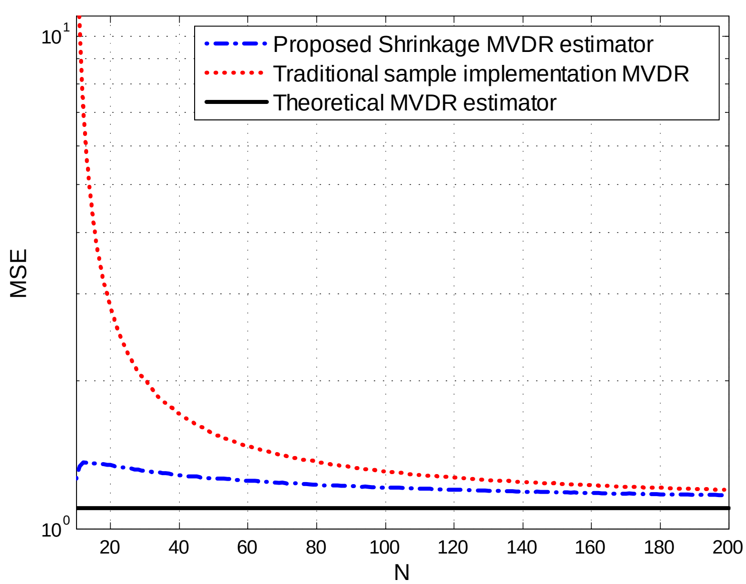

Figure 4.2: Performance comparison between proposed Shrinkage MVDR estimator in (4.4) , theoretical MVDR estimator (1.7) and its sample based implementation (1.9), when SNR= 5 dB, M=10 and SIR_i=0 dB.

regime and it remains close to the theoretical MVDR. Moreover, it is interesting to observe how as N increases, all the methods tend to converge. Recall that on the one hand this is due to the optimality of the sample MVDR, in this situation. On the other hand, when N is large compared to M, the proposed method is shrunk towards the sample MVDR, see (4.6). More insights about the shrinkage effect will be given in figure 4.6.

In figure 4.3 the same type of simulation than in figure 4.2 is carried out. The only difference is that now a weaker interterference is considered, namely SIR_i=10 dB for any i to assess the influence of the interferers power on the performance of the estimators. Figure 4.3 highlights that the improvement in performance of the proposed shrinkage method with respect to the sample MVDR is even more clear for any sample size regime. Namely, the convergence of the proposed method to the theoretical MVDR is faster than in the sample MVDR. Moreover, the improvement in the small sample size regime of the proposed method with respect to the sample MVDR is even higher. The reason for this behavior is that in

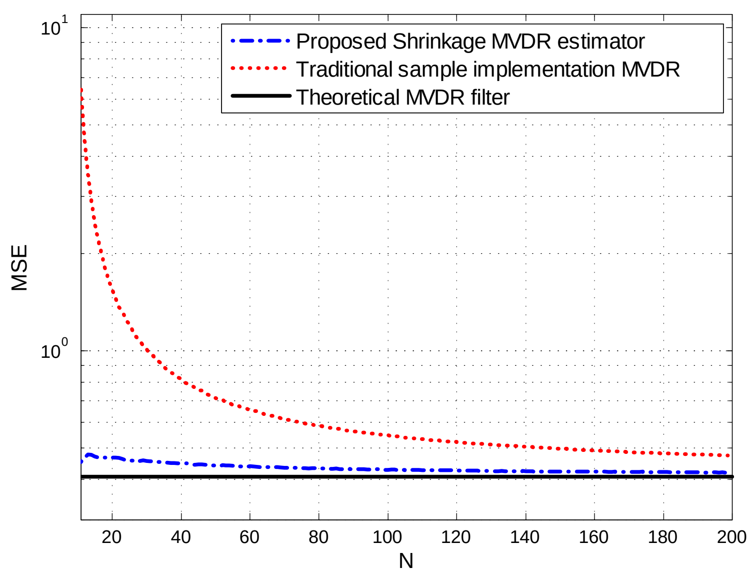

Figure 4.3: Performance comparison between proposed Shrinkage MVDR estimator in (4.4) , theoretical MVDR estimator (1.7) and its sample based implementation (1.9), when SNR= 5 dB, M=10 and SIR_i=10 dB.

the proposed method we are shrinking the sample MVDR towards the Bartlett filter and it is well known that the Bartlett filter is optimal when only the signal of interest and additive white noise are present in the scenario. This is equivalent to say that SIR_k is large and in the current figure we are closer to this situation than in the previous one. Moreover, the rest of the comments inferred from figure 4.2 apply for 4.3 as well. I.e. the robustness of the proposed shrinkage method to the small sample size situation, the statistical dominance of it compared to the sample MVDR and the convergence of all the methods for large sample size regimes.

Next, in figures 4.4 and 4.5 the same type of simulation than in figures 4.2 and 4.3 is carried out, respectively. The difference is that now M is reduced to a relatively low observation dimension, namely M is set to 6 antennas. This simulation completes the

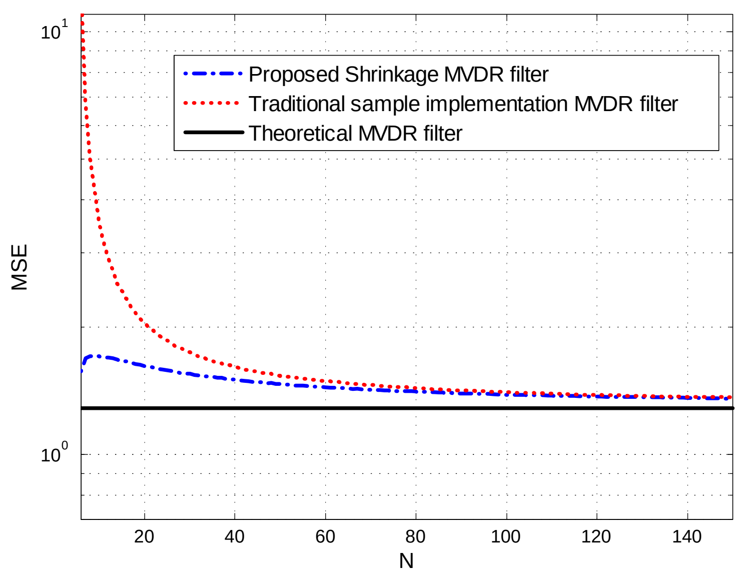

Figure 4.4: Performance comparison between proposed Shrinkage MVDR estimator in (4.4) , theoretical MVDR estimator (1.7) and its sample based implementation (1.9), when SNR= 5 dB, M=6 and SIR$_i$=0 dB.

study of the effect of decreasing M in the performance of the proposed estimator, which was started in figures 4.1 to 4.3. Interestingly enough, figures 4.4 and 4.5 highlight that the performance of the proposed method is robust to this situation. Altough it is not so close to the theoretical MVDR as in the previous figures, it still outperforms the sample

MVDR in any of the sample size regimes considered herein, i.e. $M/N \in (0,1)$. Indeed, the improvement of performance is still huge in the small sample size regime.

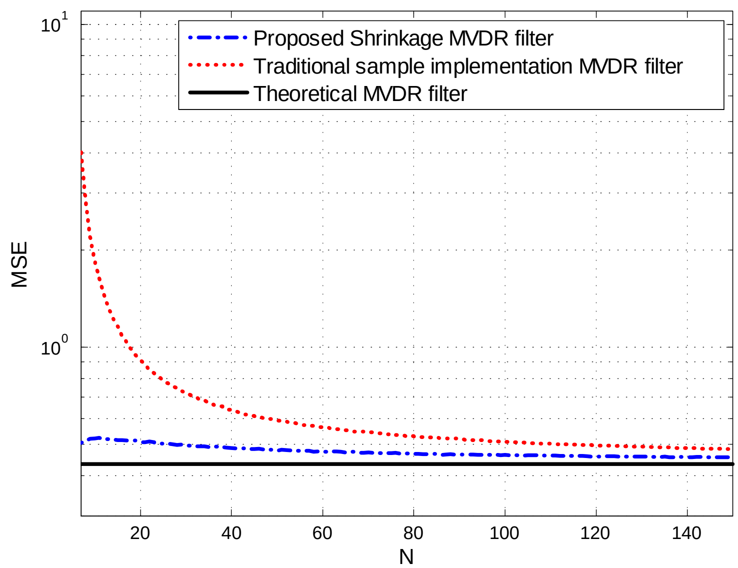

Figure 4.5: Performance comparison between proposed Shrinkage MVDR estimator in (4.4) , theoretical MVDR estimator (1.7) and its sample based implementation (1.9), when SNR= 5 dB, M=6 and SIR$_i$=10 dB.

In figure 4.6 more insights about the shrinkage effect are given. Namely, we run a montecarlo simulation to plot $|\breve{\alpha}_{c,1}|^2$ and $|\breve{\alpha}_{c,2}|^2$ of the proposed method in (4.4), consisting of shrinking the sample MVDR towards the Bartlett filter. Moreover, the simulation conditions are the same than for figure 4.2, i.e. SNR= 5 dB, $M = 10$ and SIR$_i$ $= 0$ dB and N varies fulfilling $\frac{M}{N} \in (0,1)$. Recall that the proposed shrinkage filter reads $\breve{\mathbf{w}}_{c,s}$=$\breve{\alpha}_{c,1}\hat{\mathbf{R}}^{-1}\mathbf{s}$+$\breve{\alpha}_{c,2}\mathbf{s}$ and that its behavior is as follows. On the one hand when the sample size increases, i.e. $\frac{M}{N}$ decreases, $\breve{\mathbf{w}}_{c,s}$ tends to give more weight to the sample MVDR than to the Bartlett filter. Indeed when $\frac{M}{N} \rightarrow 0$ the proposed filter $\breve{\mathbf{w}}_{c,s}$ tends to disregard the Bartlett filter and give most of the weight to the sample MVDR. This is because the sample MVDR is the optimal filter for the large sample size regime, see also (4.6). And effectively, figure 4.6 highlights this behavior, as $\frac{M}{N}$ decreases $|\breve{\alpha}_{c,1}|^2$ tends to increase whereas $|\breve{\alpha}_{c,2}|^2$ tends to decrease. On the other hand, as in general in the small sample size regime the

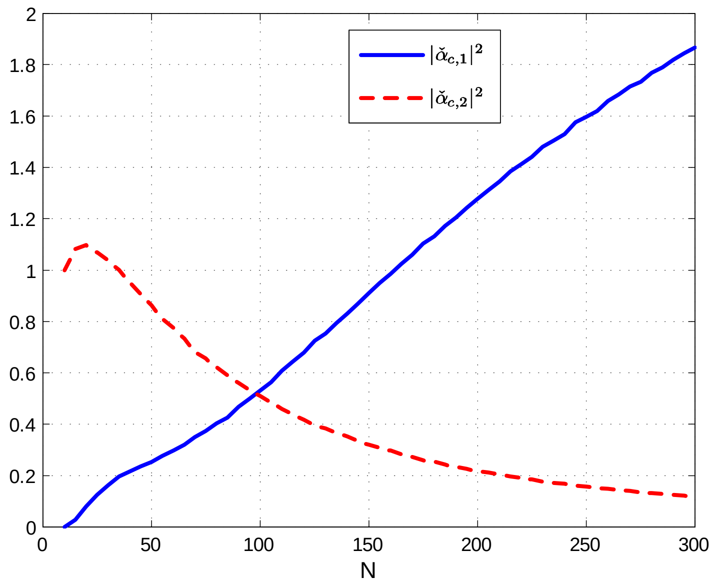

Figure 4.6: Shrinkage Factors of the proposed shrinkage method in (4.4) when $M = 10$, SNR= 5 dB and SIR$_i$=0 dB.

Bartlett filter yields better performance than the sample MVDR, $\check{\mathbf{w}}_{c,s}$ has the next behavior. As $\frac{M}{N}$ increases, $\check{\mathbf{w}}_{c,s}$ tends to give more weight to the Bartlett filter than to the sample MVDR. Indeed in the extreme case where $\frac{M}{N} \to 1$, the proposed filter $\check{\mathbf{w}}_{c,s}$ tends to disregard the sample MVDR and give most of the weight to the Bartlett filter. And effectively figure 4.6 highlights this behavior as well. Namely, As $\frac{M}{N}$ increases, $|\check{\alpha}_{c,2}|^2$ tends to increase whereas $|\check{\alpha}_{c,1}|^2$ tends to decrease.

Chapter 5

Optimal shrinkage of the sample LMMSE using summary statistics

5.1 Introduction

In this chapter, we present a method that overcomes the drawbacks of the conventional sample LMMSE that were discussed in section 1.3. Moreover, it is an alternative to the shrinkage method, based on a random matrix theory approach, exposed in (3.8). Namely, the method proposed in this chapter is also based on shrinking the sample LMMSE as in (3.8). Nonetheless, in order to obtain the optimal shrinking factor a different approach is proposed. Namely, instead of direct minimization of the MSE, we suggest to minimize the average MSE. Then, assuming that the observed data is Gaussian distributed, it turns out that the optimal shrinkage factor depends on the summary statistics of a complex inverse Wishart distribution, namely on the first two moments. Therefore, as this information is known we come up with a shrinkage estimator of the sample LMMSE that is optimal when considering that the observed data is Gaussian and the average of the MSE as a cost function.

This chapter is organized as follows, in section 5.2 the design criterion for the shrinkage LMMSE estimator to be designed is presented. Then, in section 5.3 the derivation of the optimal shrinkage LMMSE is exposed. Finally, in section 5.4 numerical simulations that assess the performance of the proposed estimator are presented.

5.2 Optimal Shrinkage based on minimizing the average MSE

Let begin this section by recalling that the MSE in the estimation of $x(n)$, when considering the linear model in (1.1) for the observed signal $\mathbf{y}(n)$ and a linear estimator $\hat{x}(n) = \mathbf{w}^H\mathbf{y}(n)$ of $x(n)$, reads as follows for the generic filter $\mathbf{w}$,

$$\begin{aligned}&\mathrm{MSE}\left(\mathbf{w}\right) \triangleq \mathbb{E}\left[\left|x(n) - \mathbf{w}^H\mathbf{y}(n)\right|^2\right] = \\ &\mathbf{w}^H\mathbf{R}\mathbf{w}+\gamma(1 - \mathbf{w}^H\mathbf{s} - \mathbf{s}^H\mathbf{w})\end{aligned}$$

The approach followed in Chapters 3 and 4 to design the estimators is based on a constrained optimization of this MSE with respect to $\mathbf{w}$, when considering a shrinkage filter of the form $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}+\alpha_2\mathbf{s}$, and then applying RMT results to obtain a realizable and consistent estimate of the optimal filter. In this chapter we consider also a shrinkage filter, namely a shrinkage of the sample LMMSE, i.e. $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$ as in (3.8). Nevertheless, herein we propose a different approach, instead of the design based on direct optimization of the MSE and the application of RMT results to obtain a realizable and asymptotically optimum estimator. First, observe that when considering a filter based on shrinking the sample LMMSE the MSE is indeed a conditional expectation,

$$\mathrm{MSE}\left(\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}\right) \triangleq \mathbb{E}\left[\left|x(n) - \mathbf{w}^H\mathbf{y}(n)\right|^2 \mid \hat{\mathbf{R}}\right] \tag{5.1}$$

As a consequence, the MSE in this case is a random quantity. Thus, by taking the expectation of (5.1) over $\hat{\mathbf{R}}$, we obtain a statistical average of the outcomes of the MSE, which arise from the values of the support of the random matrix $\hat{\mathbf{R}}$. I.e. we obtain an average MSE. Therefore, we propose, in the next statement, to design a shrinkage filter $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$, which optimizes the average MSE.

Problem statement:
Consider that a set of N observations $\{\mathbf{y}(n)\}_{n=1}^{N}$ are available and that are modeled according to (1.1) when assumptions (a)-(f) hold. Then, obtain an estimate of $x(n)$ in (1.1) by solving the next optimization problem,

$$\begin{aligned}&\hat{x}_{ls,s}(n) = \hat{\mathbf{w}}_{ls,s}^H\mathbf{y}(n) \\ &\hat{\mathbf{w}}_{ls,s} = \arg\min_{\mathbf{w}} \mathbb{E}_{\hat{\mathbf{R}}}\left[\mathbb{E}_{x,\mathbf{n}}\left[\left|x(n) - \mathbf{w}^H\mathbf{y}(n)\right|^2 \mid \hat{\mathbf{R}}\right]\right] \\ &s.t.\ \mathbf{w} = \hat{\alpha}_{ls,s}\hat{\mathbf{R}}^{-1}\mathbf{s}\end{aligned} \tag{5.2}$$

Observe that this problem formulation is a particular case of the general problem posed in (1.11) which aims to summarize all the problems dealt with in this master thesis. Namely, the functional $f(\cdot)$ corresponds to the expectation operator $\mathbb{E}[\cdot]$, which is taken over $\hat{\mathbf{R}}$.

5.3 Optimal shrinkage of the sample LMMSE based on summary statistics

In this section we propose an estimator, based on shrinking the sample LMMSE, which solves the problem stated in (5.2). I.e. a method relying on a filter $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$ which optimizes the average MSE. To this end, first we expose the result in the next theorem and then the derivation and the corresponding comments are exposed.

Theorem 5.1 *Let consider that a set of N observations $\{\mathbf{y}(n)\}_{n=1}^{N}$, modeled according to (1.1) with assumptions (a)-(f), are available. Moreover, let consider a linear estimator of the parameter $x(n)$ in (1.1), based on shrinking the sample LMMSE and let define $c_f = M/N$. Then, the estimator that optimizes the average MSE, i.e. that solves the problem stated in (5.2), reads as follows,*

$$\begin{array}{c} \hat{x}_{ls,s}(n) = \hat{\mathbf{w}}_{ls,s}^{H}\mathbf{y}(n); \hat{\mathbf{w}}_{ls,s} = \hat{\alpha}_{ls,s}\hat{\mathbf{R}}^{-1}\mathbf{s} \\ \hat{\alpha}_{ls,s} = \gamma\left((1-c_f)^2 - \frac{1}{N^2}\right) \end{array} \tag{5.3}$$

Proof: The proof is provided below in this section.

■

Remark: As the estimator proposed in Chapter 3 in (3.8), the method proposed herein in (5.3) relies on a shrinkage of the sample LMMSE, i.e. on a filter of the type $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$. Nonetheless, the shrinkage factor α is obtained following a different procedure. On the one hand, the method in (3.8) was obtained by direct optimization of the MSE and then using an asymptotic approximation, relying on RMT results, of the optimal though unrealizable shrinkage factor (3.4). On the other hand, the method proposed in this chapter in (5.3), optimizes the average MSE and does not require any asymptotic approximation. Nevertheless, the price to pay is that the observed data is assumed to be Gaussian distributed, as assumption *(f)* in *(1.1)* is presumed to hold. On the contrary, the method (3.8), based on RMT, does not require any assumption about the distribution of the observations.

Proof of Theorem 5.1

Next, we prove that the proposed shrinkage LMMSE method in (5.3) solves the problem stated in (5.2). To this end, observe that (5.2) may be rewritten as follows, after introducing the signal model for $\mathbf{y}(n)$ in (1.1) into (5.2), after solving the inner conditional expectation, which is operated upon the joint pdf of $x(n)$ and $\mathbf{n}(n)$ and bearing in mind the model of $\mathbf{R}$ in assumption *(b)* of (1.1),

$$\begin{aligned}\hat{\alpha}_{ls,s} = \arg\min_{\alpha_s} |\alpha_s|^2 \,\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right] + \\ \gamma - \gamma(\alpha_s^* + \alpha_s)\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right]\end{aligned} \tag{5.4}$$

Where with some abuse of notation, the subindex in the expectation, indicating that it is operated upon $\hat{\mathbf{R}}$ has been dropped. The solution to (5.4) is found after setting the first derivate of the cost function equal to zero and after straigthforward manipulations,

$$\hat{\alpha}_{ls,s} = \frac{\gamma\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right]}{\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right]} \tag{5.5}$$

Therefore, in order to obtain the proposed shrinkage LMMSE estimator (5.2), the summary statistics, namely the first moment, of the random quantities $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$ must be specified. To this end, let define $\mathbf{Y}$ as the juxtaposition of the available realizations of $\mathbf{y}(n)$ in (1.1), i.e. $\mathbf{y}(n)$ is the *n-th* column of $\mathbf{Y}$. Moreover, let $\mathbf{X} \in \mathbb{C}^{M\times N}$ be a random matrix, whose columns are iid according to a standard complex Gaussian distribution, namely $[\mathbf{X}]_{:,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)\ \forall k$. Then, we can write the available data as a function of $\mathbf{X}$, indeed $\mathbf{Y} \overset{d}{=} \mathbf{R}^{1/2}\mathbf{X}$, where $\overset{d}{=}$ denotes equality in distribution. Moreover, as $\hat{\mathbf{R}} = 1/N\mathbf{Y}\mathbf{Y}^H$ we can rewritte the SCM as a function of $\mathbf{X}$. Namely, applying the property of the inverse of a multiplication of matrices we obtain,

$$\hat{\mathbf{R}}^{-1} \overset{d}{=} N\mathbf{R}^{-1/2}(\mathbf{X}\mathbf{X}^H)^{-1}\mathbf{R}^{-1/2} \tag{5.6}$$

Substituting (5.6) in (5.5) the next equalities can be readily verified,

$$\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right] = N\mathbf{s}^H\mathbf{R}^{-1/2}\mathbb{E}\left[(\mathbf{X}\mathbf{X}^H)^{-1}\right]\mathbf{R}^{-1/2}\mathbf{s} \tag{5.7}$$

$$\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right] = \\ N^2\mathbf{s}^H\mathbf{R}^{-1/2}\mathbb{E}\left[(\mathbf{X}\mathbf{X}^H)^{-1}(\mathbf{X}\mathbf{X}^H)^{-1}\right]\mathbf{R}^{-1/2}\mathbf{s} \tag{5.8}$$

Now, let $\mathbf{\Omega} \triangleq \mathbf{X}\mathbf{X}^H$, then as $[\mathbf{X}]_{:,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)\ \forall k$, $\mathbf{\Omega}^{-1}$ is distributed according to a complex inverse Wishart distributions with N degrees of freedom and scale parameter $\mathbf{I}_M$, see [73], i.e. $\mathbf{\Omega}^{-1} \sim \mathcal{CW}_M^{-1}(N, \mathbf{I}_M)$. Therefore, it turns out that in order to obtain the optimal shrinkage factor (5.5) the first and second moments of a complex inverse Wishart distribution must be found. Namely, in (5.7) the first moment is needed, which according to [73, eq. 39] reads component-wise $\mathbb{E}[\,[\mathbf{\Omega}^{-1}]_{i,j}] = 1/(N-M)$ if $i = j\ \forall i \in \{1, \ldots, M\}$ and 0 otherwise. I.e. $\mathbb{E}[\mathbf{\Omega}^{-1}] = 1/(N-M)\mathbf{I}_M$, which substituted in (5.7) yields,

$$\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right] = \frac{N}{N-M}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} \tag{5.9}$$

In order to obtain the proposed estimator it remains to obtain an expression for (5.8), namely for $\mathbb{E}[\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}]$. To this end, the second moment of the complex inverse Wishart is needed, which according to ([73, eq. 41]) reads component-wise as follows,

$$\mathbb{E}[[\mathbf{\Omega}^{-1}]_{i,j}[\mathbf{\Omega}^{-1}]_{l,k}] = \frac{[\mathbf{I}_M]_{i,j}\,[\mathbf{I}_M]_{l,k} + \frac{1}{N-M}\,[\mathbf{I}_M]_{l,j}\,[\mathbf{I}_M]_{i,k}}{(N-M)^2-1} \tag{5.10}$$

Now, observe that the *p-th* element of the main diagonal of $\mathbb{E}[\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}]$ reads $\sum_{i=1}^{M}\mathbb{E}[[\mathbf{\Omega}^{-1}]_{p,i}[\mathbf{\Omega}^{-1}]_{i,p}]$ $\forall p$. Moreover, according to (5.10),

$$\mathbb{E}[[\mathbf{\Omega}^{-1}]_{p,i}[\mathbf{\Omega}^{-1}]_{i,p}] = \begin{cases} \frac{1}{(N-M)((N-M)^2-1)}, & p \neq i \\ \frac{N-M+1}{(N-M)((N-M)^2-1)}, & p = i \end{cases}$$

Therefore the elements of the main diagonal of $\mathbb{E}[\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}]$ read $\forall p \in \{1, \ldots, M\}$,

$$\sum_{i=1}^{M}\mathbb{E}[[\mathbf{\Omega}^{-1}]_{p,i}[\mathbf{\Omega}^{-1}]_{i,p}] = \frac{N}{(N-M)((N-M)^2-1)} \tag{5.11}$$

With regard to the elements of $\mathbb{E}[\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}]$ out of the main diagonal, they are characterized by the expression $\sum_{j=1}^{M}\mathbb{E}[[\mathbf{\Omega}^{-1}]_{i,j}[\mathbf{\Omega}^{-1}]_{j,k}]$ with $i \neq k\ \forall i, k \in \{1, \ldots, M\}$. Therefore according to (5.10) we can conclude that $\forall i, k \in \{1, \ldots, M\}$ with $i \neq k$ the next equality holds,

$$\sum_{j=1}^{M} \mathbb{E}[[\boldsymbol{\Omega}^{-1}]_{i,j}[\boldsymbol{\Omega}^{-1}]_{j,k}] = 0 \tag{5.12}$$

As a consequence, considering (5.11) and (5.12), $\mathbb{E}[\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}^{-1}]$ is given by,

$$\mathbb{E}[\boldsymbol{\Omega}^{-1}\boldsymbol{\Omega}^{-1}] = \frac{N}{(N-M)((N-M)^2-1)}\mathbf{I}_M \tag{5.13}$$

Recalling that $\boldsymbol{\Omega} \triangleq \mathbf{X}\mathbf{X}^H$ and inserting (5.13) in (5.8) we obtain the desired expression for the denominator of the optimal shrinkage factor (5.5),

$$\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right] = \frac{N^3}{(N-M)((N-M)^2-1)}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} \tag{5.14}$$

Finally, substituting (5.9) and (5.14) in the expression of the optimal shrinkage factor in (5.5) and after straightforward manipulations we obtain the optimal shrinkage LMMSE estimator for our problem statement in (5.2).

$$\begin{gathered}\hat{x}_{l,s}(n) = \mathbf{w}_{l,s}^H\mathbf{y}(n) \\ \mathbf{w}_{l,s} = \gamma\left((1-c_f)^2 - \tfrac{1}{N^2}\right)\hat{\mathbf{R}}^{-1}\mathbf{s}\end{gathered} \tag{5.15}$$

This concludes the proof as (5.15) coincides with the proposed shrinkage estimator in (5.3).

■

5.4 Numerical simulations

In this section the performance of the proposed estimator in this chapter in (5.3), based on a shrinkage of the sample LMMSE, which optimizes the average MSE and which uses summary statistics of the complex inverse Wishart, is assessed by means of numerical simulations. As this method intends to overcome the performance degradation of the sample LMMSE in the small sample size regime, we compare its performance with that of the sample and the theoretical LMMSE methods in (1.9) and (1.4), respectively. Moreover, we compare it also with the shrinkage of the sample LMMSE based on optimizing the MSE and using RMT tools, i.e. with (3.8), as it intends to be an alternative to that method. The performance comparison is conducted in terms of the MSE.

According to the expressions of the estimators and the MSE in (1.5), the parameters controlling the simulations are c, $\frac{M}{N}$, $\hat{\mathbf{R}}$, $\mathbf{R}$, γ and $\mathbf{s}$. As in chapter 3, in order to specify the models for these parameters an array signal processing application is considered, though the models are flexible enough to be applied to other fields of signal processing, e.g. in spectrum analysis. The models for these simulation parameters are the same than the ones in chapter 3, namely we consider the models specified in equations (3.18) to (3.23). The only difference is that the parameter M will be specified depending on the simulations. Moreover in all the figures the SNR is set to 5 dB and the SIR_i to 10 dB for all the interferers and N varies fulfilling that $\frac{M}{N} \in (0, 1)$ for a fixed M.

The first simulation, exposed in figure 5.1, compares the sample LMMSE method, the theoretical LMMSE and the proposed Shrinkage LMMSE method in this chapter when $M = 5$. It can be observed that the proposed method dramatically outperforms the sample LMMSE in the small sample size regime, i.e. when $\frac{M}{N} \to 1$. This behavior is due to the robustness of the shrinkage methods to the small sample size regimes. Moreover, figure 5.1 shows that the proposed method outperforms the conventional sample LMMSE for any of the sample sizes considered herein, i.e. $\frac{M}{N} \in (0, 1)$ and remains closer to the theoretical LMMSE estimator. Another comment worth mentioning is the evolution as the sample size tends to be large, i.e. as $\frac{M}{N} \to 0$. It can be observed that in this situation all the estimators tend to converge to the optimal method, i.e. the theoretical LMMSE. The rationale for that behavior in the case of the sample LMMSE is that in this situation the SCM is the MVUE of $\mathbf{R}$ and it is well conditioned. That is to say in that case $\gamma\hat{\mathbf{R}}^{-1}\mathbf{s} \to \gamma\mathbf{R}^{-1}\mathbf{s}$. With regard to the proposed method based on shrinking the sample LMMSE, when $c_f = \frac{M}{N} \to 0$ implies that N is large compared to M and that $\hat{\mathbf{R}}^{-1} \to \mathbf{R}^{-1}$, and as a consequence the proposed shrinkage LMMSE in (5.3) tends to the theoretical LMMSE in (1.4),

$$\hat{\mathbf{w}}_{ls,s}=\gamma((1-c_f)^2 - \frac{1}{N^2})\hat{\mathbf{R}}^{-1}\mathbf{s} \xrightarrow{c_f \to 0} \mathbf{w}_l = \gamma\mathbf{R}^{-1}\mathbf{s}$$

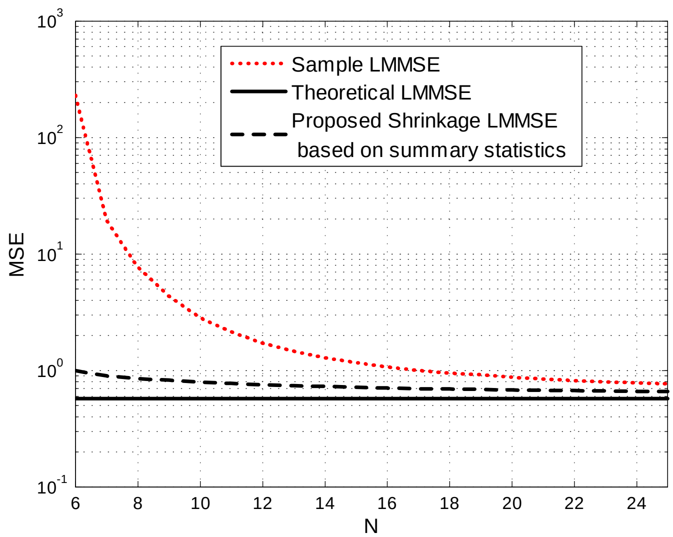

Figure 5.1: Performance comparison between proposed Shrinkage LMMSE based on summary statistics (5.3), theoretical LMMSE estimator (1.4) and sample LMMSE (1.9) when $M = 5$, $SNR = 5$ dB and $SIR_i = 10$ dB.

The next set of simulations is presented in figures 5.2 to 5.4. They compare, for $M = 3$, $M = 5$ and $M = 10$ respectively, the performance of the two proposed methods based on shrinking the sample LMMSE. The one proposed in this chapter in (5.3) and the one, based on RMT, which was proposed in chapter 3 in (3.8). For the simulation purposes, the theoretical LMMSE is also considered and all the simulation parameters, except M, are the same than in figure 5.1. This set of simulations highlights that both shrinkage methods are robust to the small sample size regime, i.e. $\frac{M}{N} \to 0$, though the method proposed in this chapter in (5.3), is slightly better. This behavior is due to the fact that the shrinkage method based on RMT is optimum when $M, N \to \infty$ at a constant rate c, see Theorem 3.2. That is to say, that method is obtained by finding an asymptotic equivalent of the

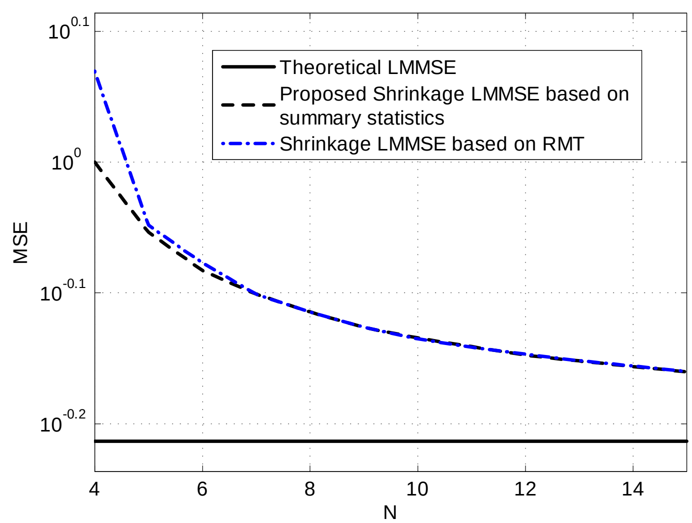

Figure 5.2: Performance comparison between proposed Shrinkage LMMSE based on summary statistics (5.3), theoretical LMMSE estimator (1.4) and Shrinkage LMMSE based on RMT (3.8) when M=3 , $SNR = 5$ dB and $SIR_i = 10$ dB.

optimal though unrealizable method (3.4) by means of RMT tools. As a consequence as in these figures M and N are finite, a degradation in performance may arise. On the contrary, the shrinkage method (5.3), based on averaging the MSE and using the summary statistics of the Complex Inverse Wishart, does not need any asymptotic approximation. Therefore, it makes sense that it behaves better than the one based on RMT. The same rationale explains why as M becomes larger both methods tend to have the same performance, e.g. see figure 5.4. Anyway, it is important to observe that the performance degradation of the method based on RMT, due to the fact of having a finite M and N, is rather small, compared to an alternative shrinkage estimator as (5.3) that does not assume any asymptotic approximation.

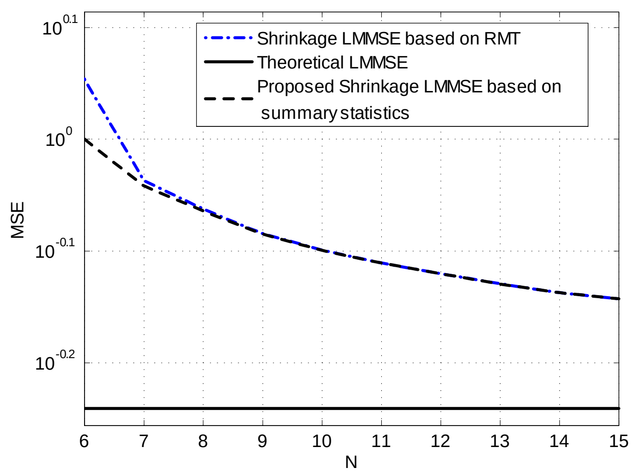

Figure 5.3: Performance comparison between proposed Shrinkage LMMSE based on summary statistics (5.3), theoretical LMMSE estimator (1.4) and Shrinkage LMMSE based on RMT (3.8) when M=5.when $SNR = 5$ dB and $SIR_i = 10$ dB.

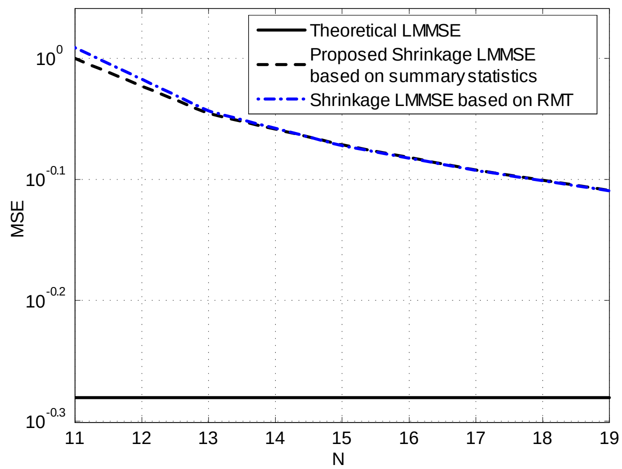

Figure 5.4: Performance comparison between proposed Shrinkage LMMSE based on summary statistics (5.3), theoretical LMMSE estimator (1.4) and Shrinkage LMMSE based on RMT (3.8) when M=10, $SNR = 5$ dB and $SIR_i = 10$ dB.

Chapter 6

Conclusions and Future Research Topics

This master thesis has dealt with the performance degradation, in the small sample size regime, of the sample LMMSE and sample MVDR. These are the reference methods in linear estimation of a parameter as they are optimal, in an MSE sense, provided that the number of available realizations of the observed data is large compared to its dimension. Thus, these sample methods are used in a wide range of applications in signal processing such as beamforming in array signal processing or spectrum analysis. In order to counteract the peformance degradation of the sample LMMSE and the sample MVDR methods, in the small sample size regime, corrections based on affine transformations of these methods have been considered.

This approach, stemming from shrinkage estimation theory, permits to combine the optimal properties of the sample methods for large sample size regimes and to counteract the performance degradation in small sample size situations. The idea is that the affine transformation is introducing a bias in the original sample estimators with the aim that the overall estimation error is reduced. I.e. a bias variance tradeoff is trying to be optimized. Thus, the shrinkage filters proposed herein consist of a scaling of the sample LMMSE and the more general case where both the sample LMMSE and the sample MVDR are combined, by means of a weighted average, with a Bartlett filter. The rationale behind this more general method is that the Bartlett filter is a constant estimator of the theoretical LMMSE or MVDR filters, as it is obtained when substituting the unknown correlation matrix by the identity matrix. As a consequence, the Bartlett filter displays certain bias but zero variance in the estimation of the theoretical filter, as it has a deterministic expression. In terms of shrinkage estimation this is called to shrink the sample methods towards the

Bartlett filter and it is introducing a bias in the estimation with the aim of diminishing the overall estimation error when combining it with the sample LMMSE or MVDR. Worth mentioning is also the fact that for small sample size regime the shrinkage filter tends to be a weighted version of the Bartlett filter and tends to disregard the contribution of the sample methods. This is a desirable behavior as the Bartlett filter perform in general better than the sample estimators in the small sample size regime. Thus, we obtain a robust estimator to the small sample size regime. On the other hand, for a large sample size situation, the shrinkage methods tend to the sample methods, which is a desirable property as in this situation they are optimal.

As direct optimization of the MSE leads to unrealizable shrinkage methods, random matrix theory or more specifically G-estimation has been considered to obtain consistent estimators of the optimal shrinkage LMMSE and MVDR methods in chapters 3 and 4, respectively. Namely, the consistency is studied within the general asymptotic analysis where both the observation dimension and the sample size tend to infinity but at a fixed rate. This generalizes the classical concept of consistency where only the sample size tends to infinity while the observation dimension remains fixed. Indeed, general asymptotic analysis naturally deals with both the small, intermedium and large sample size regimes. Thus, the random matrix based approach leads to obtain shrinkage methods that are realizable, consistent and robust to the small sample size regime.

Numerical simulations have shown a huge improvement of the proposed shrinkage methods compared to their conventional counterparts, i.e. the sample LMMSE and MVDR, in the small sample size regimes. Indeed, the statistical dominance of the proposed methods with respect to their sample couterparts has been highlighted in any of the sample size regimes considered herein. I.e. recalling that M denotes the observation dimension and N the sample size, the shrinkage methods outperform both the sample LMMSE and MVDR whenever $M/N \to c \in (0, 1)$. Moreover, another advantage of this RMT based approach is that it does not rely on any assumption about the distribution of the observations.

In this regard, future research should extend the proposed estimators to the general case where M may be even lower than N, i.e. to $M/N \to c \in (0, \infty)$. It would be also interesting to extend these shrinkage methods, based on RMT, to the more general case where one aims to estimate a vector of parameters observed through a linear model. Moreover, as the strategy based on using shrinkage estimation and random matrix theory has shown to be so powerful, we could try to apply it to other classical methods such as Maximum Likelihood (ML) or Least Squares (LS). In this regard other shrinkage estimation techniques such as the LASSO and ridge regression could be further explored.

In chapter 5, an alternative to RMT based on multivariate analysis has been proposed. Namely, this is an alternative for the particular case of gaussian observations and for a shrinkage of the sample LMMSE based on a weighted version of it. The proposed approach is based on minimizing the average MSE. This leads to an optimal shrinkage filter that depends on the moments of an inverse complex Wishart distribution. Therefore, as this statistical information is known we can obtain an optimal shrinkage estimator that minimizes the average MSE. Numerical simulations have shown that this approach dramatically outperforms the sample LMMSE and slightly outperforms the shrinkage LMMSE based on RMT for a low observations dimension as it does not rely on any asymptotic approximation. Related to this approach, a future topic of research could be to study the applicability of this multivariate analysis based method to other type of distributions. Namely, a possible candidate is the elliptical distribution, which generalizes several distributons such as the Gaussian, the Cauchy or the Student-t.

References

[1] H.L. Van Trees, *Optimum Array Processing*, John Wiley and Sons, New York, NY, USA, 2002.

[2] P. Stoica and R. Moses, *Introduction to Spectral Analysis*, Prentice-Hal, NJ, 2000.

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

[4] A.E. Hoerl and R.W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, February 1970.

[5] A.N. Tikhonov and V.Y. Arsenin, *Solution of Ill-Posed Problems*, V. H. Winston and Sons, New York, 1977.

[6] L.S. Mayer and T.A. Willke, "On biased estimation in linear models," *Technometrics*, vol. 15, pp. 497–508, August 1973.

[7] Y.C. Eldar and A.V. Oppenheim, "Covariance shaping least-squares estimation," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 686–697, March 2003.

[8] M.S. Pinsker, "Optimal filtering of square-integrable signals in gaussian noise," *Problems Inform. Trans.*, vol. 16, pp. 120–133, 1980.

[9] J. Pilz, "Minimax linear regression estimation with symmetric parameter restrictions," *J. Stat. Planning Inference*, vol. 13, pp. 297–318, 1986.

[10] Y.C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

[11] Y.C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation of deterministic parameters with bounded data uncertainties," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2177–2188, Aug. 2004.

[12] A. Beck, A. Ben-Tal, and Y.C. Eldar, "Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties," *Math. Progr., ser. B,Springer-Verlag*, Dec. 2005.

[13] Z. Ben-Haim and Y.C. Eldar, "Maximum set estimators with bounded estimation error," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3172–3182, Aug. 2005.

[14] N.R. Goodman, "Statistical analysis based on a certain multivariate complex gaussian distribution," *Ann. Math. Stat.*, vol. 34, pp. 152–177, March 1963.

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

[16] X. Mestre and M. Lagunas, "Modified subspace algorithms for DOA estimation with large arrays," *IEEE Trans. Signal Process.*, vol. 56, no. 2, pp. 598–614, February 2008.

[17] T.L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Comm.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[18] R. Couillet, J. W. Silverstein, Z. Bai, and M. Debbah, "Eigen-inference for energy estimation of multiple sources," *IEEE Trans. Info. Theory*, vol. 57, no. 4, pp. 2420–2439, March 2011.

[19] X. Mestre and M.A. 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.

[20] 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.

[21] Y.I. Abramovich and A.I. Nevrev, "An analysis of effectiveness of adaptive maximisation of the signal-to-noise ratio which utilises the inversion of the estimated correlation matrix.," *Radio Engineering and Electronic Physics. Translated from: Radiotekhnika i Elektronika*, vol. 26, no. 12, pp. 67–74, December 1981.

[22] R.A. Monzingo and T.W. Miller, *Introduction to Adaptive Arrays*, John Wiley and Sons, New York, 1980.

[23] W.W. Piegorsch and G. Casella, "The early use of matrix diagonal increments in statistical problems," *SIAM review*, vol. 31, no. 3, pp. 428–434, 1989.

[24] P. Stoica, J. Li, X. Zhu, and J.R. Guerci, "On using a priori knowledge in space-time processing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2598–2602, June 2008.

[25] O. Ledoit and M. Wolf, "A well conditioned estimator for large-dimensional covariance matrices," *Journal of Multivariate Analysis*, vol. 88, pp. 365–411, 2004.

[26] F. Rubio, *Generalized Consistent Estimation in Arbitrarily High Dimensional Signal Processing*, Ph.D. thesis, Universitat Politècnica de Catalunya, Barcelona, April 2008.

[27] Y. Eldar, "Comparing between estimation approaches: Admissible and dominating linear estimators," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1689–1702, May 2006.

[28] E.L. Lehmann and G. Casella, *Theory of Point Estimation*, Springer, New York, 2nd edition, 1999.

[29] 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*, 1956, vol. 1, pp. 197–206.

[30] W. James and C. Stein, "Estimation with quadratic loss," in *Proc. of 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1961, vol. 1, pp. 311–319.

[31] L.D. Brown, "On the admissibility of invariant estimators of one or more location parameters," *Annals of Mathematical Statistics*, , no. 37, pp. 1087–1136, August 1966.

[32] J. Jacod and Ph. Protter, *Probability essentials*, Springer, New York, 2nd edition, 2003.

[33] J. Tu and G. Zhou, "Markowitz meets talmud: A combination of sophisticated and naive diversification strategies," *Journal of Financial Economics*, vol. 99, pp. 204–215, August 2010.

[34] E. Wigner, "Characteristic vectors of bordered matrices with infinite dimensions," *The Annal of Mathematics*, vol. 62, no. 3, pp. 548–564, November 1955.

[35] V.A. Marcenko and L.A. Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR - Sbornik*, vol. 1, no. 4, pp. 457–483, 1967.

[36] V.L. Girko, *Statistical Analysis of Observations of Increasing Dimension*, vol. 28, Kluwer, Dordrecht, The Netherlands, 1995.

[37] V.L. Girko, *An Introduction to Statistical Analysis of Random Arrays*, VSP, The Netherlands, 1998.

[38] X. Mestre, "On the asymptotic behavior of the sample estimates of eigenvalues and eigenvectors of covariance matrices," *IEEE Transactions on Signal Processing*, vol. 56, no. 11, pp. 5353–5368, November 2008.

[39] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim, "On the capacity achieving covariance matrix for rician mimo channels: An asymptotic approach," *IEEE Trans. Information Theory*, vol. 56, no. 3, pp. 1048–1069, March 2010.

[40] A. Tulino and S. Verdú, "Random matrix theory and wireless communications," in *Foundations and Trends in Communications and Information Theory*, S. Verdú, Ed., vol. 1. New York: NOW Publishers, June 2004.

[41] J. Capon, "High resolution frequency wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, August 1969.

[42] J. Wishart, "The generalized product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20, no. 1–2, pp. 32–52, 1928.

[43] A.T. James, "Distributions of matrix variates and latent roots derived from normal samples," *The Annals of Mathematical Statistics*, vol. 35, no. 2, pp. 475–501, 1964.

[44] T. Ratnarajah, R. Vaillancourt, and M. Alvo, "Eigenvalues and condition numbers of complex random matrices," *SIAM Journal on Matrix Analysis and Applications*, vol. 26, no. 2, pp. 441–456, 2005.

[45] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy - Mathematical surveys and monographs No. 77*, vol. 77, American Mathematical Society, Providence, RI, USA, 2006.

[46] P. Biane, "Free probability for probabilists," *Quantum probability communications*, vol. 11, pp. 55–71, 2003.

[47] A. Masucci, Ø. Ryan, and S. Yang M. Debbah, "Finite dimensional statistical inference," *IEEE Trans. Info. Theory*, vol. 57, pp. 2457–2473, April 2011.

[48] N.R. Rao and A. Edelman, "The polynomial method for random matrices," *Foundations of Computational Mathematics*, vol. 8, no. 6, pp. 649–702, Dec. 2008.

[49] J.W. Silverstein and Z.D. Bai, "On the empirical distribution of eigenvalues of a class of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, February 1995.

[50] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Transactions on Information Theory*, vol. 54, no. 11, pp. 5113–5129, November 2008.

[51] Z. D. Bai, "Methodologies in spectral analysis of large dimensional random matrices, a review," *Statistica Sinica*, vol. 9, pp. 611–677, 1999.

[52] T. Guhr, A. Müller-Groeling, and H.A. Weidenmüller, "Random matrix theories in quantum physics: common concepts," *Phy. Rep.*, vol. 299, pp. 189–425, 1998.

[53] L. Laloux, P. Cizeaux, M. Potters, and J.P. Bouchaud, "Random matrix theory and financial correlations," *International Journal of Theoretical and applied finance*, vol. 3, no. 3, pp. 391–397, July 2000.

[54] N. Hansen and A. Ostermeier, "Adapting arbitrary normal mutation distributions in evolution strategies: the covariance matrix adaptation," in *Proceedings of IEEE International conference in evolutionary computation*, July 1996, pp. 312–317.

[55] R. Couillet and M. Debbah, *Random Matrix Methods for Wireless Communications*, Cambridge University Press, 2011.

[56] F. Dupuy and P. Loubaton, "On the capacity achieving covariance matrix for frequency selective mimo channels using the asymptotic approach," *IEEE Trans. Info. Theory*, vol. 57, pp. 5737–5753, September 2011.

[57] J.W. Silverstein, "Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices," *Journal of Multivariate Analysis*, vol. 55, pp. 331–339, August 1995.

[58] V. Girko, "Strong law for the eigenvalues and eigenvectors of empirical covariance matrices," *Random Operators and Stochastic Equations*, vol. 4, no. 2, pp. 179–204, 1996.

[59] A.J. Baranchik, "A family of minimax estimators of the mean of a multivariate normal distribution," *Ann. Math. Statist.*, vol. 41, pp. 642–645, September 1970.

[60] W.E. Strawderman, "On the existence of proper bayes minimax estimators of the mean of a multivariate normal distribution," in *Proc. Sixth Berkeley Symp. Math. Stats. Prob.*, 1972, pp. 51–55.

[61] B. Efron and C. Morris, "Stein's estimation rule and its competitors-an empirical bayes approach," *J. Amer. Statist. Assoc.*, vol. 68, pp. 117–130, 1973.

[62] C. Stein, "Estimation of the mean of a normal distribution," *Ann. Statist.*, vol. 9, pp. 1135–1151, 1981.

[63] J. Berger and R. Wolpert, "Estimating the mean function of a gaussian process and the stein effect," *J. Multivariate Anal.*, vol. 13, pp. 401–424, 1983.

[64] A. Meucci, *Risk and Asset Allocation*, Springer, New York, 2005.

[65] P. Jorion, "Bayes-stein estimation for portfolio analysis," *Journal of Financial and Quantitative Analysis*, vol. 21, pp. 279–291, 1986.

[66] S.N. Evans and P.B. Stark, "Shrinkage estimators, skhorokhod's problem and stochastic integration by parts," *Annals of Statistics*, vol. 24, pp. 809–815, 1996.

[67] B. Efron and C. Morris, "Limiting the risk of bayes and empirical bayes estimators-Part ii: The empirical bayes case.," *J. Amer. Statist. Assoc.*, vol. 67, pp. 130–139, 1972.

[68] J.H. Manton, V. Krishnamurthy, and H.V. Poor, "James-Stein state filtering algorithms," *IEEE Trans. Signal Proc.*, vol. 46, no. 9, pp. 2431–2447, Sept. 1998.

[69] Y.C. Eldar, "Generalized sure for exponential families:applications to regularization," *IEEE Trans. Signal Proc.*, vol. 46, no. 9, pp. 2431–2447, Sept. 1998.

[70] Y. Chen, A. Wiesel, Y.C. Eldar, and A.O. Hero, "Shrinkage algorithms for mmse covariance estimation," *IEEE Trans. Signal Proc.*, vol. 58, no. 10, pp. 5016–5029, Oct. 2010.

[71] F. Rubio and X. Mestre, "Generalized consistent estimation on low-rank Krylov subspaces of arbitrarily high dimension," *IEEE Trans. Signal Process.*, vol. 57, no. 10, October 2009.

[72] L. Vandenberghe and S. Boyd, *Convex Optimization*, Cambridge University Press, New York, 6th edition, 2008.

[73] D. Maiwald and D. Kraus, "Calculation of moments of complex Wishart and complex inverse Wishart distributed matrices," *IEE Proc. - Radar, Sonar and Navigation*, vol. 147, no. 4, pp. 162–168, August 2000.