UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Shrinkage corrections of sample linear estimators in the small sample size regime

Jorge Serra Puertas

Shrinkage corrections of sample linear estimators in the small sample size regime

Ph.D. Dissertation

Author: Jorge Serra Puertas

Advisor: Dra. Montse Nájar Martón

Department of Signal Theory and Communications
Universitat Politècnica de Catalunya (UPC)

Barcelona, Spain, September 2016

Abstract

We are living in a data deluge era where the dimensionality of the data gathered by inexpensive sensors is growing at a fast pace, whereas the availability of independent realizations of this observed data is rather limited. Therefore, classical statistical inference methods relying on the assumption that the sample size is large, compared to the observation dimension, are suffering a severe performance degradation.

Within this general framework, this thesis focus on a problem that appears in a vast number of signal processing applications. That 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. For instance, 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. The design of the linear filters is based on the optimization of a measure of performance. In signal processing and in general in statistical inference the Mean Square Error (MSE) and the Signal to Interference plus Noise Ratio (SINR) are widely accepted as possible performance measures. Thus, the optimal estimators are obtained by means of the optimization of these metrics 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 thesis. On the one hand, when there is information about the first two moments of the parameter of interest, the optimization of the MSE leads to obtain the Linear Minimum Mean Square Error (LMMSE). On the other hand, when such statistical information is not available one may force a no distortion constraint towards the signal of interest in the optimization of the MSE, which is equivalent to maximize the SINR. This leads to obtain the Capon or Minimum Variance Distortionless Response (MVDR) method. This is the Best Linear Unbiased Estimator (BLUE) indeed.

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 in the small sample size regime. The small sample size regime may be due to short stationarity constraints or due to a system with a high observation dimension. These

situations have appeared traditionally in applications such as radar or adaptive beamforming, e.g. in radioastronomy or over-the-horizon-radar there are large arrays with hundreds or thousands of sensors. Another example is environmental monitoring in wireless sensor networks. In this case there may be a large number of nodes measuring a physical parameter, though the number of available measures for inference purposes may be rather limited as they are battery powered in general. That is, as it was mentioned above, we are living in a data deluge era where a variety of sensors are generating high dimensionality data, though the number of independent realizations of these random processes is rather limited.

Therefore, the aim of this thesis is to propose corrections of sample estimators, such as the sample LMMSE and MVDR, that allow to circumvent their performance degradation in the small sample size regime. To this end, we are equipped with two powerful tools, shrinkage estimation and random matrix theory (RMT). On the one hand, shrinkage estimation introduces a structure on the estimators that permits to take profit of the optimality of the sample methods in the large sample size regime and force some corrections in small sample size situations. In fact, historically since the time of Stein, shrinkage methods have shown to be a mean to improve sample based estimators by optimizing a bias variance tradeoff. On the other hand, 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 grow without bound, but at a fixed rate. That is, RMT is used to obtain consistent estimates within an asymptotic regime that deals naturally with small sample size situations. Moreover, another advantage is that this approach based on shrinkage estimation and RMT does not rely on any assumptions about the distribution of the observations.

The proposed shrinkage filters deal directly with the estimation of the signal of interest (SOI), which leads to performance gains compared to some related work methods based on either optimizing a metric related to the data covariance estimate or proposing rather ad-hoc regularizations of the sample covariance. Moreover, the next benefits are observed, compared to state-of-the-art methods which also treat directly the estimation of the SOI and which are based on a shrinkage of the sample covariance or diagonal loading. The proposed shrinkage filter structure is more general as it contemplates corrections of the inverse of the sample covariance and considers the related work methods as particular cases. This leads to performance gains which are notable when there is a mismatch in the signature vector related to the SOI. This mismatch and the finite sample size are indeed two important sources of degradation of the sample LMMSE and MVDR methods. Thereby, in the last part of this thesis, unlike the previous proposed filters and the related work, we propose a shrinkage filter which treats directly both sources of degradation.

Resumen

Estamos viviendo en una era en la que la dimensión de los datos, recogidos por sensores de bajo precio, está creciendo a un ritmo elevado. Sin embargo, la disponibilidad de observaciones estadísticamente independientes de estos datos es relativamente limitada. Por lo tanto, los métodos clásicos de inferencia estadística sufren una degradación importante en sus prestaciones, ya que asumen que el tamaño muestral es grande en comparación con la dimensión de los datos.

En este marco general, esta tesis se centra en un problema que aparece en un gran número de aplicaciones en el procesado estadísitco de señales. Éste es la estimación de un parámetro, observado a través de un modelo lineal. La inferencia de este parámetro se basa normalmente en una transformación lineal de los datos, es decir en un filtrado lineal. Por ejemplo, el objetivo de la conformación de haz en procesado de agrupaciones de antenas es enfocar el haz de la agrupación hacia una dirección para obtener la señal asociada a una fuente de interés. Esto se consigue mediante un filtrado espacial. El diseño de dichos filtros lineales se basa en la optimización de una medida de prestación. En procesado de señal y en general en inferencia estadística el error cuadrático medio (MSE) y la relación señal a ruido más interferente (SINR) son medidas de pretaciones ampliamente aceptadas. Esto lleva a dos estimadores notables que servirán como referencia durante toda la tesis. Por una parte, cuando hay información sobre los momentos de segundo orden del parámetro a estimar, la optimización del MSE lleva a obtener el estimador lineal de mínimo error cuadrático medio (LMMSE). Por otra parte, cuando esa información estadística no está disponible, se puede forzar la restricción de no distorsión de la señal de interés en la optimización del MSE, lo que es equivalente a maximizar la SINR. Esto conduce a obtener el estimador de Capon (MVDR).

Aunque el LMMSE y el MVDR son los métodos óptimos no son realizables, ya que dependen de la inversa de la matriz de correlación de los datos, que no es conocida. El procedimiento habitual para solventar este problema es sustituirla por la inversa de la correlación muestral (SCM), esto lleva al LMMSE y al MVDR muestral. Este procedimiento es óptimo cuando el tamaño muestral tiende a infinito o es muy grande en comparación con la dimensión de los datos. Sin embargo, en la práctica este tamaño muestral elevado no suele producirse y los métodos LMMSE y MVDR muestrales sufren una degradación importante ya que la covarianza muestral no está bien condicionada en este régimen de tamaño muestral pequeño. Este escenario se puede deber a periodos cortos de estacionariedad estadística o a sistemas cuya dimensión sea elevada. Estas situaciones han aparecido tradicionalmente en aplicaciones como radar o conformación de haz adaptativo, por ejemplo en radioastronomía o "over-the-horizon-radar" hay agrupaciones de antenas con cientos

o miles de elementos. Otro ejemplo es en redes de sensores inalámbricas, donde un gran número de nodos pueden tomar medidas de un parámetro físico a estimar, aunque el tamaño muestral para la inferencia es limitado, debido a las restricciones de batería de los nodos. Es decir, vivimos en la era de la inundación de los datos, donde sensores generan datos de dimensión elevada, pero el tamaño muestral para el análisis de los datos es limitado.

Por lo tanto, el objetivo de esta tesis es proponer correcciones de los estimadores muestrales de los métodos LMMSE y MVDR que permitan combatir la degradación en el régimen de tamaño muestral pequeño. Para conseguir este objetivo se utilizan dos herramientas potentes como son la teoría de las matrices aletorias (RMT) y la estimación shrinkage. Por una parte, la estimación shrinkage introduce una estructura de los estimadores que permite forzar ciertas correcciones cuando el tamaño muestral es pequeño. De hecho históricamente, promovidos por Stein, los métodos shrinkage han demostrado mejorar los estimadores muestrales mediante la optimización del compromiso entre media y varianza del estimador. Por otra parte, la optimización directa de los métodos shrinkage lleva a métodos no realizables. Por eso, luego se propone obtener una estimación consistente de ellos en el régimen asintótico en el que tanto la dimensión de los datos como el tamaño muestral tienden a infinito, pero manteniendo un ratio constante. Es decir RMT se usa para obtener estimaciones consistentes en un régimen asintótico que trata naturalmente las situaciones de tamaño muestral pequeño. Además otra ventaja de esta metodología basada en RMT y estimación shrinkage es que no asume ninguna distribución de probabilidad concreta sobre las observaciones.

Los filtros shrinkage que se proponen tratan directamente la estimación del parámetro de interés. Esto lleva a ganancias de prestaciones en comparación con otros métodos basados en optimizar una métrica relacionada con la estimación de la covarianza de los datos. También lleva a mejoras en comparación a métodos que proponen regularizaciones ad hoc de la SCM. Además las siguientes ventajas se observan en comparación con metódos que también tratan directamente la estimación del parámetro de interés y que se basan en una regularización de tipo shrinkage de la SCM. La estructura de filtro shrinkage propuesta es más general ya que contempla correcciones de la inversa de la covarianza muestral y contempla los métodos del estado del arte como casos particulares. Esto lleva a ganancias en las prestaciones que son notables cuando hay una incertidumbre en el signature vector que se presume asociado a la señal de interés. Esta incertidumbre en el signature vector y el tamaño muestral pequeño son de hecho las dos degradaciones más importantes del LMMSE y MVDR muestrales. Así, en la última parte de la tesis se propone un filtro shrinkage que trata esas dos degradaciones de forma directa, a diferencia de los filtros propuestos con anterioridad en la tesis y a diferencia de los métodos del estado del arte mencionados anteriormente.

A mi familia.

Agradecimientos

El camino que me ha llevado hasta este momento ha requerido un esfuerzo notable, pero las situaciones vividas durante este trayecto lo compensan. Enseñanzas que llevaré en mi mochila y que con toda seguridad me ayudarán a saber afrontar los senderos que depare el futuro con más claridad y seguridad. Este periodo ha estado enmarcado en una especie de dualidad onda partícula, en sentido metafórico. En efecto, compaginar mi trabajo como ingeniero de investigación y el desarrollo de esta tesis como estudiante de doctorado ha sido una tarea ardua. Por ello tengo mucho que agradecer a las personas que en algún momento u otro me he ido encontrando durante este camino o que han caminado conmigo. El primer agradecimiento es para mi directora de tesis, muchas gracias Montse por la paciencia que siempre has demostrado. Por supuesto otro reconocimiento y agradecimiento está dedicado a Luis, quien también ha experimentado los efectos de la dualidad onda partícula durante estos años. Recuerdo por ejemplo esas noches sin dormir haciendo ejercicios para las asignaturas de doctorado o mejor dicho el máster MERIT. Prosiguiendo con los agradecimientos, me gustaría dedicar unas palabras a todas las personas con las que he convivido y convivo en el trabajo. Gracias a todos los que me habeis brindado la oportunidad de trabajar en diversos proyectos de investigación y a todos los compañeros y compañeras con los que he desarrollado dichos proyectos. Gracias a esto mi formación como investigador es mucho más completa y me ha permitido afrontar la tesis con mayor perspectiva. También quiero agradecer a Fran el haberme introducido en la teoría de las matrices aleatorias. Por otra parte, las conversaciones diarias sobre cualquier tema con Marc Portolés y con Ciprian siempre fueron muy motivadoras, intelectualmente hablando, e indudablemente me han ayudado a forjar un pensamiento crítico que en mi opinión es fundamental tener no sólo como investigador sino también como persona. Mi trabajo como ingeniero de investigación también me ha permitido estar en contacto con un ambiente multicultural muy enriquecedor que me ha permitido abrir la mente y conocer otras realidades, a menudo rompiendo ideas preconcebidas que en nada se asemejan a la realidad. Los agradecimientos no estarían completos si no mencionara a mis amigos y a mi familia. Vuestro apoyo incondicional es fundamental para mí. Me gustaría mencionar a mi padre y a mi abuela, físicamente ya no están, pero valores como la humildad, el esfuerzo en el trabajo, el pensamiento crítico o el ayudar a los demás perduran en mí. Otra persona fundamental de mi familia es mi madre, quien siempre ha estado a mi lado. Y para el final he dejado la mejor parte, el agradecimiento a mi pareja, Shakira, que tiene una magia que hace que cada día sea especial y que los problemas desaparezcan cuando estamos juntos.

Table of Contents

Notation

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

$\mathbf{A}^T, \mathbf{A}^*, \mathbf{A}^H$	Transpose, complex conjugate and complex conjugate transpose 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 $\mathbf{A}$.
$\left\|\mathbf{A}\right\|_F$	Frobenius norm of a matrix $\mathbf{A}$, $\left\|\mathbf{A}\right\|_F = \left(\mathrm{Tr}\left[\mathbf{A}^H\mathbf{A}\right]\right)^{1/2}$.
$\left\|\mathbf{a}\right\|$, $\left\|\mathbf{a}\right\|_2$	Euclidean norm of a vector $\mathbf{a}$, $\left\|\mathbf{a}\right\| \triangleq \left\|\mathbf{a}\right\|_2 = \left(\mathbf{a}^H\mathbf{a}\right)^{1/2}$.
$\left[\mathbf{a}\right]_i, \mathbf{a}_i$	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 $\boldsymbol{\Sigma}$.
$\lvert a \rvert$	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. $\lvert a - b \rvert \rightarrow 0$, where the convergence is almost surely unless otherwise stated.
$\mathrm{Re}\{a\}$ and $\mathrm{Im}\{a\}$	Denote the real and imaginary part of the complex number a, respectively.

Acronyms

AWGN Additive White Gaussian Noise

BLUE Best Linear Unbiased Estimator

DL Diagonal Loading

DOA Direction of Arrival

ESD Empirical Spectral Distribution

GLRT Generalized Likelihood Ratio Test

GSA General Statistical Analysis

LMMSE Linear Minimum Mean Square Error

LS Least Squares

LSD Limiting Spectral Distribution

LW Ledoit and Wolf

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

SOCP Second-Order Cone Program

SOI Signal Of Interest

ULA Uniform Linear Array

WSN Wireless Sensor Network

Chapter 1

Introduction: Research motivation, State of the Art and Objectives

1.1 Introduction

1.1.1 Research motivation and introduction to the problem

As Donoho said in [1], this is the century of data. That is, we are gathering and processing more data every day. From a signal processing point of view, one of the consequences of this paradigm is the trend of analyzing high dimensional signals for different inference purposes. As an illustration of this framework consider the next examples. The first one is the data collected in the context of large sensor networks. These data are employed for large dimensional covariance estimation in [2] or for failure and anomaly detection of the sensor nodes in [3]. Another interesting example may be found in future wireless communications systems, where large arrays of antennas are considered as a means of potentially increasing, by orders of magnitude, the spectral and energy efficiency of current wireless cellular systems using relatively simple linear processing, see e.g. [4] or [5]. Other examples of inference problems involving high dimensional signals are covariance matrix estimation from genomic data in bionformatics [6] or classification of hyperspectral images in remote sensing [7]. Moreover, the number of available statistical samples to analyze high dimensional signals is in general rather limited [8], this is the case of most of the examples introduced above. This is a framework that in adaptive beamforming, in the context of array signal processing, is well known and it is due to either arrays of large antennas or to short stationarity properties of the underlying signal [9], [10]. In other

words, the dimension of an antenna array may be comparable to the number of available snapshots or even larger, e.g. in radioastronomy there may be large arrays with hundreds of elements. Another example, in the context of cognitive wireless networks is [11], where 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. This statistical framework where the sample size is comparable to the observation dimension provokes that classical estimation methods relying on sample moments, such as the sample covariance matrix (SCM), undergo a severe performance degradation and that new methodologies be envisaged, see e.g. [9] or [12]. In fact, even for a moderate dimension of the observed signal, if the sample size is comparable to it then the sample methods suffer a performance degradation. Within this framework, the aim of this thesis is to design estimation methods that counteract the degradation of methods relying on the SCM.

More specifically, herein the problem of linear estimation of an unknown parameter observed through a linear model is considered, see section 1.2 for more details. This problem is ubiquitous in signal processing [13], e.g. beamforming in array signal processing [14]. In this case, given a set of observations at the output of the antenna array, one may seek to estimate the signal associated to a given direction of arrival by means of a spatial filter [15]. In order to design the estimator the aim is to optimize a measure of performance. In this regard, among the statistical signal processing community, the mean square error (MSE) is a widely accepted metric to measure the quality of an estimator, see e.g. [13]. Besides the MSE, another popular metric, which is widely used in the communications literature, is the Signal to Interference plus Noise Ratio (SINR), see e.g. [16]. In fact, when the parameters of the model are perfectly known, the estimator obtained from optimizing the MSE leads to maximize the SINR as well, though the converse is not true, see e.g. [17]. Nonetheless under parameter model uncertainty the optimization of the SINR and the MSE cannot be simultaneously attained see [18]. A myriad of estimators have been designed in the literature with the common aim of obtaining a good MSE performance, see [13], [19] and references therein. 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, see e.g. [13]. This is indeed 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, one can resort to the MVDR or Capon method [20] [14], which imposes a constraint that removes the dependence on the unknown second moment of the parameter of interest. This

constraint is interpreted as a distortionless constraint in the direction of interest in array processing, and as an unbiasedness constraint when the parameter of interest is modeled as deterministic [17]. Thereby, assuming the unbiasedness constraint, the MVDR optimizes the variance of the estimation and corresponds to the well known Best Linear Unbiassed Estimator (BLUE) in the statistical literature [13]. 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. For instance, the Tikhonov regularizer [21] [22], the shrunken estimator [23], the covariance shaping least-squares estimator [24] and minimax MSE estimators [25–30].

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, see section 1.3 for further details. 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 [14]. 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 [31, 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, as it was mentioned above, in practical scenario conditions the assumption that the sample size is large compared to the observation dimension, i.e. $N \gg M$, does hardly hold. In fact N may be comparable to M or even lower, leading to the so-called small sample size regime. Unfortunately, when the sample size is comparable to the observation dimension, the traditional implementation of the optimal estimators based on the SCM leads to a severe performance degradation, see [9], [10], [14], [32] and references therein. In fact, they may display worse performance than a matched filter, also called conventional beamformer in array processing, which is a naive strategy based on directly replacing the theoretical covariance matrix by a scaled identity matrix. 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 [33]. In fact, random matrix theory (RMT) characterizes the asymptotic eigenvalue density of the SCM and its relation to the eigenvalues of the true covariance matrix, see [34], [35] or [36]. That characterization relies on the asymptotic regime where $M, N \to \infty$ with $M/N \to c \in (0, \infty)$, though it is a good approximation in a finite regime as well. Namely, RMT states that when N increases, compared to M, then the eigenvalue density tends to concentrate around the true eigenvalues in the form of a narrow cluster. In fact, for $N \to \infty$ and a finite M, the eigenvalue density tends to the true eigenvalues. However, when N decreases (compared to M) then the eigenvalue density suffers a widening effect. That is to say, the clusters of eigenvalues tend to widen and if N keeps drecreasing then a single cluster in the eigenvalue density of the SCM is observed. Therefore, in small sample size situations, as the sample LMMSE and MVDR methods rely on the SCM, they are no longer optimal and require a calibration that counteract their severe performance degradation.

Another important source of degradation of the LMMSE and MVDR methods is an imprecise knowledge of the vector that performs the linear transformation between the signal of interest and the observations, that is the steering vector in array signal processing. This imprecise knowledge can be viewed as a mismatch between the presumed and the actual steering vector. In practice, in array processing, these mismatches may be due to array calibration errors, distorted antenna shape, pointing errors towards the signal of interest or source wavefront distortions among other reasons, see e.g. [37]. The consequence of this mismatch is that the LMMSE and MVDR methods may interpret the signal of interest as an interference and may tend to cancel it, which leads to an important degradation.

1.1.2 Related work

Several approaches have been suggested in the literature to deal with the severe performance degradation that undergo the methods relying on the SCM, in the small sample size regime. Most of them seek to obtain an estimation of $\mathbf{R}$ that is better than the SCM or that regularizes it. Perhaps, one of the first was a regularization technique applied in the field of array signal processing. It is called diagonal loading (DL) and consists of 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 [9], [10], [38–40] and references therein. The rationale motivating this method may be analyzed from different viewpoints. First, DL is a type of shrinkage estimation, see e.g. [33] or [41], where the term $\alpha\mathbf{I}$ is introducing a bias in the estimation of the covariance. The aim is to reduce the overall estimation error, compared to the SCM, whose estimation error comes from the

estimation variance. The family of shrinkage estimators of the covariance will be analyzed in more detail below. In fact, the term $\alpha\mathbf{I}$ permits to invert the estimation of the covariance even when $M > N$ and it paves the way to obtain a well-conditioned estimator of the covariance when M and N are comparable. That is, $\check{\mathbf{R}}$ can be inverted without severely amplifying the estimation error, though to achieve this aim α must be carefully selected to avoid an excessive bias. The other interesting interpretation of DL comes from the field of array signal processing, where this estimator has been thoroughly studied. The pioneering papers [42–44] focused on the case where the signal of interest is not present in the training set. They analyzed the SINR of the DL implementation of the MVDR normalized by the SINR of the theoretical MVDR and they concluded that better performance than the SMI technique can be achieved. However, due to several approximations done in that analysis, they could not specify how to compute the optimal loading factor α. When the signal of interest is present in the training set, other works, e.g. [45], concluded that DL offers better performance than the SMI technique as well, though still the optimal value of α was not clear. Thereby, the correct choice of the loading factor has been historically controversial and usually rather ad hoc methods depending on the practical setting of the application have been used. For instance, in [14, p. 748] it was proposed $\alpha = 10\hat{\lambda}_{\min}$, where $\hat{\lambda}_{\min}$ is the minimum eigenvalue of the SCM. However, more recently, the work of Mestre et al. [9] shed light on how to obtain an approximation to the optimal loading factor. Namely, they considered the Kolmogorov asymptotics where $M, N \to \infty$ at a constant rate, i.e. $M/N \to c \in (0, \infty)$, which considers implicitly the small sample size regime. Thereby, relying on RMT results, they obtained the asymptotic expression of the SINR for the DL implementation of the MVDR. This result paved the way to obtain the asymptotically optimal loading factor as the one which maximizes the asymptotic SINR through a grid search. Moreover, as that asymptotic loading factor depended on the unknown covariance, they found a consistent estimation of it relying on RMT and the general asymptotics just mentioned above. This consistent estimation of the asymptotically optimal loading factor requires a grid search optimization procedure as well.

In order to complement the discussion about the DL methods, it is worth to mention some algorithms which appeared within the field of robust beamforming in array signal processing [37, 46, 47]. These methods arise from considering some type of mismatch between the presumed and the actual steering vector of the signal of interest. To face this problem they assume that the actual steering vector lies within a region of uncertainty which is incorporated as a constraint in the optimization problem of the MVDR. The uncertainty region is modelled in a different manner in each of those works. In [37] a spherical uncertainty set is considered for the error vector between the actual and the presumed steering vector. On the other hand, in [46] and [47] ellipsoidal uncertainty sets

where considered for the steering vector. Interestingly enough, all these works shown that they can be interpreted as a DL technique were the loading factor depends clearly on the parameters that model the uncertainty region, see [37, 46, 47] for further details. That feature was important, as the pioneering works dealing with DL, see e.g. [48], shown that adding a quadratic constraint in the MVDR, a.k.a. white noise gain constraint, permits to deal with both the finite sample size effect on the SCM and the distortions in the steering vector of the signal of interest. Nonetheless, the optimal loading factor to deal with those effects was not clear. Moreover, it is important to stress that as [37, 46, 47] can be interpreted as a DL technique, they offer some robustness to the finite sample size effect that undergoes the SCM. However, unlike in [9], it is not clear that the loading factor selected by [37, 46, 47] is the optimal value to combat the finite sample size effect.

Regarding the last discussion, there are other interesting techniques that deal with uncertainties in the steering vector. Several works [18, 49–51] modelled the steering vector of the signal of interest as random with some known distribution. For instance, in [18] the steering vector was assumed to follow a Gaussian distribution, and then they designed the filter by maximizing a performance metric (average SINR or MSE). This average SINR or MSE is obtained after performing the expected value of the SINR or MSE with respect to the random steering vector. Another interesting class of techniques are eigenspace-based beamformers, which rely on using the projection of the presumed steering vector onto the signal-plus-interference subspace [52] [53], instead of just the presumed steering vector. The main shortcoming of eigenspace beamforming methods is that they perform poorly when the dimension of the signal plus interference subspace is high and at low SNR. This is mainly because at low SNR the estimation of the projection matrix onto the signal plus interference subspace breaks down because of a high probability of subspace swaps, see [54] and [55]. A different approach to the previous techniques is based on steering vector estimation, see e.g. [56] [57]. Namely, the rationale behind these works follows the next procedure. Estimate the actual steering vector as the one that maximizes the power at the output of the beamformer subject to the constraint that the estimate does not converge to any interference steering vector. The techniques [56] [57] require the a priori knowledge of the angular sector where the signal of interest lies as well as the array geometry. Moreover, they do not tackle the finite sample size degradation in the beamformer performance due to the use of the SCM. In [58], a method based on steering vector estimation is presented, which deals with the finite sample size through interference covariance matrix reconstruction. However, the computational cost is significantly higher than the worst case optimization techniques explained above [37, 46, 47].

Let us now continue with the discussion of robust methods to the finite sample size. Recently, motivated by the data deluge framework mentioned above, estimation of high

dimensional covariance matrices under small sample size regime has attracted the research community. These works may be classified in several categories. The first is shrinkage type estimators of the covariance, whose aim is to reduce the estimation error based on a linear combination of the SCM with an a priori information or guess of the covariance, denoted as $\mathbf{R}_0$. Where $\mathbf{R}_0$ may be obtained from a priori knowledge stemming from the problem at hand. Namely, the shrinkage estimator, $\check{\mathbf{R}}$, of the covariance of the observations, $\mathbf{R}$, reads $\check{\mathbf{R}} = \alpha_1 \hat{\mathbf{R}} + \alpha_2 \mathbf{R}_0$. Thereby, DL techniques can be regarded as a particular case of linear shrinkage estimators. The idea of these estimators is to blend the SCM, whose estimation error mostly comes from an estimation variance, with an estimator displaying certain amount of bias but zero or very low variance, that is a constant estimator or other type of 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 estimation error is diminished. This was the approach proposed by Ledoit and Wolf (LW) in [33], where $\mathbf{R}_0$ is a scaling of the identity matrix and the shrinkage factors α_1, α_2 are obtained as the ones that asymptotically optimize the MSE in the estimation of $\mathbf{R}$. This result was valid for any distribution of the observed data. Following this idea, [59] improved LW for Gaussian distributed observations. Namely, they proposed two strategies. The first one, uses the fact that LW is not a sufficient statistic for Gaussian data. Then, according to the Rao-Blackwell theorem, by conditioning LW to a sufficient statistic they can obtain an estimator that outperforms LW. The second one, approximates the optimal though unrealizable shrinkage estimate of the covariance, i.e. the oracle, by using an iterative procedure. In [60] the same authors study the shrinkage estimation of high dimensional covariances for elliptically distributed samples, which include as a particular case the Gaussian distribution. In the context of radar or space time adaptive processing, in [41] the authors consider a shrinkage of the SCM with a general a priori covariance $\mathbf{R}_0$ obtained from prior knowledge of the terrain probed by the radar. Unlike the previous references [41] considers complex valued data for their derivations. Also in [6] a shrinkage of the SCM is considered with a focus on bioinformatics.

A second class of estimators that regularize large covariance matrices are based on incorporating a priori knowledge on the structure of the covariance or its inverse in the form of sparsity. Also, when this a priori information is not available, one may force the sparsity structure. That is they force some entries of the covariance or its inverse to be zero and as a consequence reduce the effective number of parameters to be estimated. In other words, one may take advantage of the structure to perform the estimation in a reduced dimension. The precursor of this technique was probably Dempster in [61]. Since then, a lot of research has been devoted to select the sparsity model and to estimate the covariance. In this regard, operators such as thresholding [62], banding [63] and tapering [64] are worth to

be mentioned. Banding sets the entries far away from the main diagonal to zero and keeps the entries within a band unchanged. Tapering is similar to banding, the only difference is that the off diagonal elements within the band are gradually shrunken to zero. Thresholding, as the two previous techniques introduces sparsity in the covariance though it does not require a special structure of the covariance. These kind of estimators are consistent if certain sparsity holds and the dimensionality grows at a subexponential rate of the sample size. Nonetheless, if the sparsity assumption does not hold they are suboptimal. Moreover, tapering and thresholding are minimax estimators of the covariance, see e.g. [64] or [65]. Other techniques that have been applied for both sparse model selection and covariance estimation are the penalization techniques that enforce sparsity in gaussian graphical models (GGM). GGM represent the observed variables as nodes in a graph and their conditional independence results in a zero in the inverse covariance. Examples of these penalization techniques in GGM are [66] which uses the lasso technique [67]. Another example is [68], which considered the Dantzig selector [69] as a penalization method. Finally, it is worth to be mentioned [70], which uses the graphical lasso method [71]. Assuming that the sparsity structure is known, [72] and [2] dealt with the estimation of the covariance in GGM obtaining the MVUE in the former and a method based on pseudolikelihood estimation in the latter.

Finally, it is worth mentioning that random matrix theory (RMT) has been used in several works as a tool for dealing with an estimation problem that is constrained by a small sample size support. RMT studies, among other aspects, the asymptotic behavior of spectral functions defined from random matrices. That is, the convergence of certain functions depending on the eigenvalues and eigenvectors associated to a given random matrix. An example is the empirical distribution of the eigenvalues of a random matrix. In fact the asymptotic regime considered in RMT generalizes classical asymptotics as it considers the regime where both $M, N \to \infty$ at a constant rate $M/N \to c \in (0, \infty)$. Therefore, RMT deals naturally with the small sample size regime and the high dimensionality of the observations. What is more, it generalizes classical consistent estimation and paves the way to obtain consistent estimates within the regime where $M, N \to \infty$ at a constant rate $M/N \to c \in (0, \infty)$. In the type of linear estimation problems involving the data covariance and characterized by a limited sample size, the aim is to apply certain corrections to the classical methods which rely directly on the SCM. The design of those corrections leads usually to functions which depend on the SCM, the unknown covariance and certain design parameters. In order to obtain a realizable estimator which is asymptotically optimal and robust to the finite sample size, a RMT approach can be applied. To this end, the next procedure is usually followed. First, the asymptotic behavior of the functions involving the SCM, the unknown covariance and some design parameters, are studied. Under certain

assumptions this implies the convergence toward deterministic functions depending on the unknown covariance $\mathbf{R}$, and the design parameters. Then, relying on RMT results, the final estimator is obtained as the one that asymptotically converges to that deterministic function. Examples which follow this RMT procedure are the papers [33] and [9], mentioned above. Namely, [33] builds on RMT to obtain an estimate of the covariance, based on a shrinkage estimation of the SCM, that is asymptotically optimal in an MSE sense. On the other hand, recall that the DL method proposed in [9], builds on RMT to obtain the loading factor which asymptotically maximizes the SINR. There are other methods, which rely on RMT results, whose aim is to estimate functions involving $\mathbf{R}$, though they do not follow exactly the procedure described above. Among them, it is worth mentioning the works in [34] and [73], which study the asymptotic behavior of the classical sample estimates of eigenvalues of the covariance and the associated eigenvectors. They show that these classical sample estimates are not consistent within the asymptotic regime where $M, N \to \infty$ at a constant rate $M/N \to c \in (0, \infty)$ and propose improved estimators which show to be consistent or asymptotically optimal. In [8] an estimate of covariance matrices when $M > N$ is proposed. It is based on a dimensionality reduction through an ensemble of random unitary matrices. Finally, it is worth mentioning that reduced rank techniques building on RMT have been applied to LMMSE estimation for a limited sample size N per observation dimension M, see [32], [74] and references therein.

1.2 Signal Model

Next, we present the general model of the observed data that will be considered throughout all the 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 deterministic vector with uniformly bounded norm[1], $\mathbf{n}(n) \in \mathbb{C}^M$ is a stochastic process and N is the number of available measurements. For instance, in the context of array signal processing $\mathbf{y}(n)$ is the output of an antenna array, $\mathbf{s}$ is the steering vector, $\mathbf{n}(n)$ contains the noise plus interference signals [14] and N is the sample size or the number of available snapshots. In the sequel $x(n)$ is considered to be a random process, though for the estimators dealing with the MVDR the results will hold when $x(n)$

[1] In chapters 3 and 4 $\mathbf{s}$ is assumed to be precisely known. In chapter 5 an uncertainty in $\mathbf{s}$ is considered, i.e. the actual signature vector $\tilde{\mathbf{s}}$ differs from $\mathbf{s}$.

is modeled as a deterministic parameter. The next model assumptions are supposed to hold for any of the designed estimators throughout all this thesis,

(a) *$x(n)$ and $\mathbf{n}(n)$ are independent. Moreover, $\mathbb{E}[\mathbf{n}(n)] = \mathbf{0}$, $\mathbb{E}\left[\mathbf{n}(n)\mathbf{n}(n)^H\right] = \mathbf{R_n}$. Both the entries of $\mathbf{n}(n)$ and $x(n)$ have independent real and imaginary parts and bounded moments and $\mathbf{R_n}$ is strictly positive definite.*

(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 $\|\mathbf{s}\|^2 = 1$. Moreover it is assumed that the eigenvalues of $\mathbf{R}$ are uniformly bounded from below and above for all M and they have a limiting distribution as $M \to \infty$.*

(c) *The set of observations $\{\mathbf{y}(n)\}_{n=1}^N$ are iid.*

(d) *All the estimators, except the ones in sections 3.4, 3.6 and 4.5, assume that the number of samples is higher than the observation dimension, i.e. $N > M$ or in other words $M/N \in (0, 1)$. The ones in sections 3.4, 3.6 and 4.5 accept $M/N \in (0, \infty)$.*

Moreover, for the estimators in chapter 3 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 section 3.3,

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

Remark: In general it will be assumed that the training data set $\{\mathbf{y}(n)\}_{n=1}^N$, which is used to build the proposed filter, is statistically independent of the data to be processed.

1.3 Optimal Linear Estimators and their practical implementation issues

In this 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,

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

The MSE can be taken into account, as a measure of performance, to design the optimal linear estimator. In fact, the MSE has been widely used in signal processing to obtain linear estimators that guarantee a good estimate of the unknown parameter [13]. Examples may be found in subband beamforming in the context of array signal processing [10, Ch.5] [17] or in wireless communications for the estimation of the transmitted symbols in MIMO receivers [75] and references therein. Thereby, considering the optimization of the MSE for the family of linear estimators in (1.2) leads to the next optimization problem,

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

Furthermore, assume that the data model for the observed signal $\mathbf{y}(n)$ in (1.1) with the assumptions *(a)-(e)* holds. Then, one obtains the well known LMMSE method [13],

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

Being $\gamma \triangleq \mathbb{E}\left[|x(n)|^2\right]$ the power 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, the 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 information is not available the LMMSE estimator is not realizable.

In order to circumvent the lack of knowledge about γ one may apply the popular Capon method, also known as MVDR in the array signal processing literature, see e.g. [14] or [20]. The rationale behind this method can be interpreted from different viewpoints. First, to establish the connection with the LMMSE, consider the design of the filter based on the optimization of the MSE. After some manipulations, the expression of the MSE in (1.3) reads

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

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

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

Interestingly enough, the latter equation connects with the second viewpoint in the design of the MVDR, which is actually the most common interpretation in the literature. Namely, the MVDR arises from the optimization of the SINR at the output of the linear filtering process subject to a constraint of no distortion of the signal of interest. This statement is easily understandable by inspecting the expression of the SINR at the output of the filter $\mathbf{w}$, whose expression is given by,

$$\text{SINR} = \frac{\gamma \left|\mathbf{w}^H \mathbf{s}\right|^2}{\mathbf{w}^H \mathbf{R_n} \mathbf{w}}.$$

In order to proceed, observe that under assumptions *(a)-(e)* exposed in (1.1) the optimization in (1.6) is equivalent to the one where $\mathbf{R_n}$ 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 [14],

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

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. [13].

Before proceeding, it is important to remark that the LMMSE and MVDR methods, exposed above, have been used interchangeably in several works, as they are equivalent when the SINR is considered as the measure of performance, see e.g. [37] in the context of beamforming in array signal processing. The rationale is that they differ in a scaling which does not impact in the SINR. However, other authors have shown that this scaling is important in applications where the aim is to obtain an estimate of the signal amplitude and thereby they focused on the MSE as a measure of performance to guarantee a good estimate, this is the case of e.g. subband beamforming [17]. Moreover, within the context of wireless communications, this scaling can be interpreted as an automatic gain control which is necessary in any real MIMO system [76]. In fact, in [18], under slightly different model assumptions than herein, it was shown that optimizing the MSE leads to beamformers that optimize the SINR as well, though the converse is not true. Moreover, [18] shown that when there is an uncertainty in the steering vector $\mathbf{s}$, then the maximum SINR and minimum MSE cannot be attained simultaneously.

Therefore, the 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, re-

spectively. Moreover, they are both the optimal methos from a SINR perspective. 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, $\hat{\mathbf{R}}$, which is defined by means of the next expression,

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

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,

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

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 it is also its MVUE for a sufficiently large number of samples N compared to the observation dimension M, [31, 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. The reason for that behavior is that the SCM is not a well conditioned estimator in the small sample size regime, see e.g. [33]. This means that even small estimation errors in the SCM can lead to large errors in the inverse of the SCM, which are translated in large estimation errors in the final estimate of the parameter of interest. In fact, RMT results shed light on why the SCM is not a well conditioned estimator in the small sample size regime. Namely, relying on [35], see also [73], it can be shown that for finite sample size situations the empirical eigenvalue density of the SCM undergoes a widening effect. The consequence of this impairment is that the condition number of the SCM worsens. Indeed, in the low sample size regime the traditional implementations of the LMMSE and MVDR may display worse performance than the matched filter $\mathbf{w} = \mathbf{s}$ [9], which does not take profit of the available statistical samples $\{\mathbf{y}(n)\}_{n=1}^{N}$. Therefore, the aim of this thesis is largely devoted to propose methods that counteract the degradations of the sample LMMSE and MVDR methods in the small sample size regime.

Another impairment that the LMMSE and MVDR methods may undergo, in practice, is an uncertainty in the vector $\mathbf{s}$ that performs the linear tranformation in (1.1), i.e. the steering vector in array signal processing. Namely, in practice a model mismatch between the presumed steering vector $\mathbf{s}$ and the actual steering vector $\tilde{\mathbf{s}} = \mathbf{s} + \boldsymbol{\Delta}$ may happen. Where $\boldsymbol{\Delta}$ is an unknown distortion which may arise from look direction errors, imperfect array calibration, distorted antenna shape or source wavefront distortions among other effects, see [37]. The consequence of this model mismatch is a performance degradation of the LMMSE and MVDR methods. In fact, both the mismatches in the steering vector and in the covariance can lead the LMMSE and MVDR methods to the signal cancellation effect. This, implies that the practical implementation of the optimal methods confuse the signal of interest with an interference and try to cancel it, which clearly leads to an important performance degradation, see [9] [10] [37] and references therein. In chapter 5, the mismatches on $\mathbf{s}$ (as well as those arising from the estimation of $\mathbf{R}$) will be tackled by proposing a robust method to those uncertainties.

1.4 Robust methods to the small sample size regime and to model mismatches

In this section some of the related work presented in section 1.1.2 is explained in more detail. These are methods that will be used later on for comparison purposes with the shrinkage estimators proposed in this thesis. Two subsections are considered, the first one is devoted to explain the state-of-the-art (SoA) methods whose main aim is to tackle the small sample size degradation of the conventional sample LMMSE and MVDR techniques. As it was explained above, this is the main objective of this thesis and SoA methods based on a DL or shrinkage of the SCM are selected because these are the ones which allow a more fair comparison to the type of shrinkage methods proposed in this thesis. The second subsection presents a worst-case optimization method whose objective is to deal with uncertainties in the model, namely the vector $\mathbf{s}$ in (1.1) a.k.a the steering vector in array signal processing applications. This method also shows some robustness to the finite sample size regime as it can be interpreted as a DL technique, though in this regard it is not optimal in general. This robust method to the steering vector uncertainties is considered for comparison purposes because in chapter 5 a method is proposed to deal both with the uncertainties in the steering vector and the finite sample size regime.

1.4.1 Robust Techniques to the finite sample size

DL techniques

The rationale behind DL methods is to obtain a better estimate of the signal of interest, than the sample MVDR or LMMSE, based on a regularization technique. More specifically, recall that in the small sample size regime the SCM is not a well conditioned estimator, or in other words the condition number (which is the ratio between the maximal and minimal eigvenalues of the SCM) can be very large. In addition, the estimation of the signal of interest is a function os the SCM, i.e. $\hat{x}(n) = f(\hat{\mathbf{R}})$ see (1.9), thus by the definition of the condition number, this implies that even small estimation errors in the SCM can lead to large errors in the estimation of the signal of interest $\hat{x}(n)$. A way of improving the condition number of the SCM is by adding a positive real number δ to its diagonal, which clearly must be properly selected to obtain the desired effect. The parameter δ is known in the literature as loading factor. Therefore, the DL approach leads to obtain the next expressions for the practical implementations of the LMMSE and MVDR, respectively,

$$\mathbf{w}_{dl,l} = \gamma(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \tag{1.10}$$

$$\mathbf{w}_{dl,c}=\frac{(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}}{\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}} \tag{1.11}$$

and thus they are so called DL-LMMSE or DL-MVDR filters herein. Note that they reduce to the sample LMMSE and MVDR for $\delta = 0$, see (1.9). On the other hand, for a large enough δ they tend to a matched filter $\mathbf{w} \propto \mathbf{s}$. In fact, the expression of the DL methods arise from adding a regularization or penalty term $\delta \|\mathbf{w}\|^2$ in the objective function of the LMMSE and MVDR problems and thus as it was mentioned above is a regularization technique. Note that the key point in DL methods is how to choose the regularization parameter δ. A lot of research was devoted to this end, though the choice for δ was traditionally controversial. Namely, the pioneering works [42–44] dealing with DL in array processing tried to analyze the ratio between the output SINR of DL methods and the SINR of the optimum beamformers (i.e. the ones with known covariance). When the signal of interest is not present in the training set, they made the next assumptions for their analysis. They assumed that the interefence sources where received with high power, the loading factor (δ) was chosen higher than noise power but much lower than the minimum interference eigenvalue and $N \geqslant K$ (where K is the dimension of the interference subspace). Then, they concluded that the ratio between the output SINR of the DL and

the output SINR of the MVDR is beta distributed with parameters $(N-K+1, K)$, which implies that N must be greater than $2K$ to achieve an average output SINR within 3 dB of the optimal one. This means that under those assumptions DL methods converge much faster to the optimum methods than the sample MVDR methods, which need $N \approx 2M$ to achieve the optimum performance. However, the probability distribution of the ratio between the SINR of the DL and the optimum methods did not depend on δ and as a consequence the optimum value for the loading factor was not clear. When the signal of intereset is present in the training set it was also shown, see e.g. [45], that DL converge faster than the SMI technique to the MVDR, but still the value for δ was not clear. Therefore, rather ad-hoc values for δ have been traditionally used. Based on the analysis mentioned above, assuming that the signal of interest is not present in the training set, a traditional approach [14, p. 748] has been to chose,

$$\delta = 10\hat{\lambda}_{\min} \tag{1.12}$$

where $\hat{\lambda}_{\min}$ is the minimum eigenvalue of the SCM. Unlike those ad hoc methods, herein the analytical expression for the asymptotically optimal shrinkage filters will be given. Moreover, an improved performance will be obtained, as those methods rely on the sample eigenvalues $\hat{\lambda}$, whose distribution suffers a spread in the small sample size regime [34], which degrades the performance. An example is $\delta = 10\hat{\lambda}_{\min}$, due to the eigenvalue spread $\hat{\lambda}_{\min}$ can be smaller than expected, which implies that the benefits of regularization are worse than expected and even may tend to vanish. More insights are given in the simulation results. On the contrary we circumvent this problem using RMT tools. Namely, considering the asymptotic regime where both M and N grow large at a constant rate, we design (M, N)-consistent methods which implicitly use estimations of the real eigenvalues that are robust to the small sample size regime.

Another choice for the DL factor was proposed in [77] and it was based on analyzing the estimation error of the covariance. Namely, they assumed that the SCM is related to the real covariance through the expression $\hat{\mathbf{R}} = \mathbf{R} + \epsilon\mathbf{B}$, where $\mathbf{B}$ was a random matrix with zero mean and unit variance entries and ϵ was a positive constant indicating the estimation error of the estimated covariance. This implies that the diagonally loaded covariance can be expressed as $\check{\mathbf{R}} = \mathbf{R} + \epsilon\mathbf{B} + \delta\mathbf{I}$. By analyzing this expression they concluded that the loading factor is upper and lower bounded, $\epsilon \leqslant \delta < \mathbf{R}(i,i)$. Moreover, they proposed to estimate those bounds yielding the next expression, $\text{std}(\text{diag}(\hat{\mathbf{R}})) \leqslant \delta < \text{Trace}(\hat{\mathbf{R}})/M$. Finally, it was proposed to select δ as the estimation of the lower bound,

$$\delta = \text{std}(\text{diag}(\hat{\mathbf{R}})) \tag{1.13}$$

Therefore, this method is still rather ad hoc and it was based on enhancing the estimation of the SCM. The methods proposed in this thesis deal directly with the estimation of the parameter of interest and this leads to performance improvement compared to [77], as the numerical simulations will highlight.

More recently, the work of Mestre et al. [9] proposed a more analytical methodology to select the value of the loading factor, as they found an estimation of the asymptotically optimal δ. Namely, they considered the asymptotic regime where $M, N \to \infty$ and $M/N \to c \in (0, \infty)$, which implicitly takes into account the small sample size regime. Moreover, they used results from RMT, this permited to found that the SINR of the DL-MVDR filter converges in probability to the next expression,

$$\overline{\text{SINR}} = \left(\frac{1}{\gamma(1 - c\xi)} \frac{\mathbf{s}^H (\mathbf{R} + \rho \mathbf{I})^{-1} \mathbf{R} (\mathbf{R} + \rho \mathbf{I})^{-1} \mathbf{s}}{(\mathbf{s}^H (\mathbf{R} + \rho \mathbf{I})^{-1} \mathbf{s})^2} - \beta \right)^{-1} \tag{1.14}$$

where $\rho = \delta(1 + cb)$, $\beta = 1$ if the signal of interest is present in the observations or zero otherwise, ξ is given by the next expression

$$\xi = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{\lambda_i}{\lambda_i + \rho} \right)^2$$

and b is the unique positive solution to the following equation

$$b = \frac{1}{M} \sum_{i=1}^{M} \left(\frac{\lambda_i (1 + cb)}{\lambda_i + \rho} \right)$$

where $\lambda_{max} = \lambda_1 \geq \ \ldots \geq \lambda_M = \lambda_{min}$ are the eigenvalues of $\mathbf{R}$. Note that the asymptotically optimal loading factor is obtained through the optimization of $\overline{\text{SINR}}$ in (1.14). However, this expression depends on the unknown $\mathbf{R}$. To overcome this problem, in [9] they found an *(M,N)*-consistent estimate of $\overline{\text{SINR}}$. That is, an estimate that tends to (1.14) when $M, N \to \infty$ and $M/N \to c \in (0, \infty)$. To obtain this estimate they relied on results from random matrix theory. Thereby, they finally obtained the next estimation of the asymptotically optimal loading factor,

$$\hat{\delta} = \arg\min_{\delta} \ \frac{\mathbf{s}^H (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \hat{\mathbf{R}} (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \mathbf{s}}{(1 - c\hat{\varphi}(\delta))^2 (\mathbf{s}^H (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \mathbf{s})^2}. \tag{1.15}$$

Where $\hat{\varphi}(\delta) = 1/M \, \mathrm{Tr}(\hat{\mathbf{R}}((\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}))$. Note that $\hat{\delta}$ requires a grid search. Compared to [9], some of the methods proposed herein can obtain almost the same performance in terms of SINR when the vector $\mathbf{s}$ is perfectly known and better performance when there is an uncertainty in $\mathbf{s}$. Moreover, we propose some estimators in this thesis, based on a shrinkage of the sample LMMSE and the regularized sample LMMSE, which in terms of MSE obtain better performance than [9]. Moreover, some of the methods proposed herein avoid the grid search. All these statements will be studied in more detail in the next chapters.

Methods based on a Shrinkage of the SCM

This kind of methods enhance and regularize the SCM estimation by means of a shrinkage, i.e. $\check{\mathbf{R}} = \beta_1\hat{\mathbf{R}}+\beta_2\mathbf{I}$. That is, in some sense they are a generalization of DL. Within this class it is worth mentioning the work of Ledoit and Wolf (LW), see [33]. They propose a shrinkage of the SCM that not only regularizes and improves the estimation error of the SCM, but also minimizes the asymptotic MSE of the data covariance, though they assume real data. For the complex case, according to [78], one can stack the real and imaginary parts of the data, then estimate the associated covariance using the LW method and finally obtain the complex covariance from it, though this may give suboptimal performance as the circular symmetry property of complex data is not used. Substituting $\check{\mathbf{R}}$ in the expressions of the LMMSE and MVDR yields the so called LW-LMMSE, $\hat{x}_{lw,l}(n)$, and LW-MVDR, $\hat{x}_{lw,c}(n)$, estimators herein, respectively,

$$\hat{x}_{lw,l}(n) = \mathbf{w}_{lw,l}^H\mathbf{y}(n); \;\; \mathbf{w}_{lw,l}=\gamma(\beta_1\hat{\mathbf{R}}+\beta_2\mathbf{I})^{-1}\mathbf{s} \tag{1.16}$$

$$\hat{x}_{lw,c}(n) = \mathbf{w}_{lw,c}^H\mathbf{y}(n); \mathbf{w}_{lw,c}=\frac{(\beta_1\hat{\mathbf{R}}+\beta_2\mathbf{I})^{-1}\mathbf{s}}{\mathbf{s}^H(\beta_1\hat{\mathbf{R}}+\beta_2\mathbf{I})^{-1}\mathbf{s}} \tag{1.17}$$

$$\beta_1 = 1-\beta \quad , \quad \beta_2 = \beta\frac{\mathrm{Tr}(\hat{\mathbf{R}}_s)}{2M} \tag{1.18}$$

$$\beta = \frac{\sum_{i=1}^{N}\left\|\mathbf{y}_s(i)\mathbf{y}_s^T(i)-\hat{\mathbf{R}}_s\right\|_F^2}{N^2\left[\mathrm{Tr}(\hat{\mathbf{R}}_s^2)-\frac{\mathrm{Tr}^2(\hat{\mathbf{R}}_s)}{2M}\right]} \tag{1.19}$$

where $\hat{\mathbf{R}}_s = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}_s(i)\mathbf{y}_s^T(i)$, $\|\cdot\|_F$ denotes the Frobenius norm, $\mathbf{y}_s(i) = [\mathrm{Re}(\mathbf{y}(i)), \mathrm{Im}(\mathbf{y}(i))]^T$ and $\mathrm{Tr}(\cdot)$ denotes the trace operator. Other methods that stem or are related to LW

are [41], [59] and references therein. In any case, all these methods aim to improve the estimation of the SCM and they do not deal directly with the estimation of the parameter of interest. Herein this fact is taken into account to obtain estimators with an improved performance.

Another worth remarking method based on the shrinkage of the SCM is the one proposed in [79]. Namely, they propose the next type of shrinkage filter to counteract the degradation of the sample LMMSE in the small sample size regime

$$\mathbf{w} = (\tau_1 \hat{\mathbf{R}} + \tau_2 \mathbf{I})^{-1} \mathbf{s}.$$

This is the same type of filter structure than the LW method. However, in order to design the shrinkage parameters τ_1, τ_2 [79] proposes to minimize the MSE in the estimation of the parameter of interest $x(n)$ in (1.1). To this aim, first they define the parameter τ and the functions $L_1(\tau), L_2(\tau)$ as

$$\tau = \frac{\tau_2}{\tau_1}, \quad L_1(\tau) = \mathbf{s}^H (\hat{\mathbf{R}} + \tau \mathbf{I})^{-1} \mathbf{s}, \quad L_2(\tau) = \mathbf{s}^H (\hat{\mathbf{R}} + \tau \mathbf{I})^{-1} \mathbf{R} (\hat{\mathbf{R}} + \tau \mathbf{I})^{-1} \mathbf{s}.$$

Then, after some easy manipulations they obtain that the optimal shrinkage factors are given by,

$$\tau_1 = \frac{L_2(\tau)}{L_1(\tau)}, \quad \tau = \arg \max_{\tau} \frac{(L_1(\tau))^2}{L_2(\tau)}. \tag{1.20}$$

Note that these optimal shrinkage factors depend on the unknown $\mathbf{R}$. Thereby, they use RMT results to obtain (M, N)-consistent estimates of the optimal shrinkage factors, i.e. which tend to the values in (1.20) within the asymptotic regime where $M, N \to \infty$ and $M/N \to c \in (0, \infty)$. The expression for these (M, N)-consistent estimates is given by

$$\begin{aligned} \tau_1 &= \frac{\hat{L}_2(\tau)}{L_1(\tau)}, \quad \tau = \arg \max_{\tau} \frac{(L_1(\tau))^2}{\hat{L}_2(\tau)} \\ \hat{L}_2 &= \frac{\mathbf{s}^H (\hat{\mathbf{R}} + \tau \mathbf{I})^{-1} \hat{\mathbf{R}} (\hat{\mathbf{R}} + \tau \mathbf{I})^{-1} \mathbf{s}}{\left(1 - \frac{1}{N} \mathrm{Tr}[\hat{\mathbf{R}} (\hat{\mathbf{R}} + \tau \mathbf{I})^{-1}]\right)^2}. \end{aligned} \tag{1.21}$$

In section 3.6, a more general form of shrinkage than the one in [79] is proposed. This leads to a performance gain in terms of MSE when the signature vector $\mathbf{s}$ is known.

Moreover, when there is an uncertainty in $\mathbf{s}$, this gain in performance increases. This is thanks to the structure of the proposed filter, as it combines a regularized LMMSE with a matched filter. The latter is more robust to the signal cancelation effect than the sample LMMSE or the regularized LMMSE, when the regularization does not tackle properly the uncertainty in $\mathbf{s}$. This is explained in detail in sections 3.6 and 5.2.

1.4.2 Robust Techniques to model mismatches

In this section some methods that deal with an uncertainty in the vector $\mathbf{s}$ of the signal model (1.1), i.e. the steering vector in array processing, are reviewed. First of all, it is interesting to comment that DL techniques are in some regard robust to model mismatches in $\mathbf{s}$. This was noticed in the seminal works dealing with DL, see e.g. [48], where it was observed that many sources of error are in practice uncorrelated from sensor to sensor, e.g. errors in the theoretical distance between sensors in the array. Thus, they degrade the system performance in a similar way than adding spatially white noise to each sensor. Therefore, DL can be viewed as a regularization technique, see e.g. [80], as it is based on adding in the MVDR formulation a regularization term $\delta \left\|\mathbf{w}\right\|^2$ that penalizes the increment of spatially white noise,

$$\begin{aligned}\mathbf{w} = \arg\min_{\mathbf{w}} \ & \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} + \delta \left\|\mathbf{w}\right\|^2 \\ s.t. \ & \mathbf{w}^H \mathbf{s} = 1\end{aligned}$$

This yields the DL MVDR filter $\mathbf{w} \propto (\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$. However, the main drawback of the DL technique is how to choose a value for δ that deals properly with the uncertainties in $\mathbf{s}$. For instance, in [48] it was proposed the white noise gain constraint methodology to obtain reasonable values of δ. This relies on adding a constraint for the white noise gain $\left\|\mathbf{w}\right\|^2 \leq \phi$. One problem of this approach is that the relationship between the loading factor δ and the parameters of the white noise gain constraint is not simple. Namely, to adjust δ and fulfill the constraint, a multistep iterative procedure is required to solve the next equation, see e.g. [81],

$$\frac{\mathbf{s}^H(\hat{\mathbf{R}} + \delta\mathbf{I})^{-2}\mathbf{s}}{\left(\mathbf{s}^H(\hat{\mathbf{R}} + \delta\mathbf{I})^{-1}\mathbf{s}\right)^2} = \phi$$

Even more important is the fact that the choice of the DL factor δ depends on the white noise gain constraint through ϕ and as a consequence this methodology does not clearly

depend on the paramteres that model the uncertainty of the steering vector **s.** Therefore, this is still a rather ad-hoc technique to deal with the uncertainties in **s**. To circumvent this drawback, several works [37,46,47] appeared in the literature that incorporate explicitly an uncertainty set for **s** in an optimization problem which arises from the modification of the original MVDR problem explained above in (1.6). These works differ in the way that they model the uncertainty set for **s**. Both [46] and [47] considered ellipsoidal uncertainty sets for the uncertain steering vector. On the other hand, in [37] a spherical uncertainty set was considered for the error vector between the actual and the presumed steering vector. Next, [37] is explained in more detail, as its notion of robustness to the steering vector uncertainties is used to derive a new method, proposed in chapter 5, which is robust to both the finite sample size and to uncertainties in **s**. Unlike [37], the new method deals explicitly with the small sample size regime and it improves the methods proposed in chapters 3 and 4, because it relaxes the assumption of known **s**. Let us focus now on explain [37]. This work, assumes that the true steering vector $\tilde{\mathbf{s}}$ differs from the presumed steering vector **s** in the signal model (1.1), i.e. $\tilde{\mathbf{s}} = \mathbf{s} + \boldsymbol{\Delta}$, where $\boldsymbol{\Delta}$ describes the unknown steering vector distortions. Moreover, they assume the next model for the uncertainty set of the steering vector,

$$\mathcal{A}(\varepsilon) = \{\mathbf{a} | \mathbf{a} = \mathbf{s} + \mathbf{e}, \|\mathbf{e}\| \leq \varepsilon\}, \tag{1.22}$$

where $\varepsilon \in [0, \|\mathbf{s}\|)$ is a user parameter, see [37] [47]. Given this model, the Capon beamformer is made robust by imposing a distortionless constraint for all the steering vectors within $\mathcal{A}(\varepsilon)$,

$$\min_{\mathbf{w}} \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} \text{ subject to } |\mathbf{w}^H \mathbf{a}| \geqslant 1 \text{ for all } \mathbf{a} \in \mathcal{A}(\varepsilon). \tag{1.23}$$

This is a non-convex problem, but [37] showed that (1.23) can be reformulated as the next problem, which is convex as it can be cast as a second order cone program (SOCP),

$$\begin{aligned} \underset{\mathbf{w}}{\text{minimize}} \quad & \mathbf{w}^H \hat{\mathbf{R}} \mathbf{w} \\ \text{subject to} \quad & \mathbf{w}^H \mathbf{s} \geq \varepsilon \|\mathbf{w}\| + 1, \\ & \text{Im}\{\mathbf{w}^H \mathbf{s}\} = 0. \end{aligned} \tag{1.24}$$

Interestingly, it was shown in [37] that this beamformer belongs to the class of DL techniques, as $\mathbf{w} \propto (\hat{\mathbf{R}} + \lambda \varepsilon^2 \mathbf{I})^{-1} \mathbf{s}$. Where λ is a Lagrange multiplier and cannot be obtained in closed form [37]. Moreover, unlike the DL technique based on the white noise gain

constraint, the loading factor depends clearly on the amount of uncertainty considered for the steering vector through ε. Moreover, due to the DL interpretation, (1.24) offers certain protection against the small sample size effect. However, the robustness will depend on the choice of ε, whose value is related to the uncertainty of the steering vector, but it has not a clear relation with the covariance mismatches. Thereby, the RCB will require an ad-hoc tuning of ε to counteract properly the finite sample size effect. As (1.24), the new method proposed in chapter 5 deals explicitly with the uncertainties in $\mathbf{s}$ through the inclusion of the uncertainty set (1.22) in the formulation of the MVDR. However, unlike (1.24), the new method deals directly with the finite sample size impairments and it does not require an additional tuning of ε.

1.5 Contribution and Problem Statement

The aim throughout this thesis is to propose corrections of the sample LMMSE and MVDR that permit to deal with their sources of degradation. That is, the performance degradation in the estimation of the SOI $\hat{x}(n)$ due to the finite sample size effect in the estimation of the covariance and the imprecise knowledge of the steering vector of the SOI, i.e. $\mathbf{s}$ in (1.1). Namely, the bulk of this thesis is devoted to propose methods that tackle the degradation due to the small sample size limitation, when $\mathbf{s}$ is perfectly known. This is the material presented in chapters 3 and 4, the associated contribution and problem statement is presented next in section 1.5.1. Afterwards, in section 1.5.2, the contribution associated to the case where both sources of degradation are present is exposed, this is the material related to chapter 5.

1.5.1 Shrinkage of the sample LMMSE and MVDR to deal with the small sample size

The proposed methods to deal with the finite sample size limitation of the sample LMMSE and MVDR rely on the concept of shrinkage estimation. Namely, this is used to define the structure of the filters $\mathbf{w}$ that carry out the linear estimation of the parameter of interest, i.e. $\hat{x}(n) = \mathbf{w}^H\mathbf{y}(n)$. Shrinkage estimation is considered herein to define the structure of the estimators to be designed, as it leads to methods that 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 [82]. The inception of those methods can be traced back to the works of Stein [83], [84] and later on of Brown [85], see [82] for a thorough discussion on this topic, see

also the references above. 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 [86]. 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. Therefore, the idea of shrinkage estimation is to introduce a correction of this sample methods. This correction can be a linear transformation of the sample estimators or more in general a linear combination of them with an a priori information stemming from the problem at hand. The aim of these corrections is to optimize the bias variance tradeoff of the estimator and to diminish the overall MSE. Namely, in most cases the error of sample methods comes from the estimation variance. 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. Herein, several shrinkage estimators are proposed, whose aim is to propose corrections of the sample LMMSE and MVDR in the small sample size regime.

Thus, following the shrinkage estimation philosophy, first the most basic structure of a shrinkage estimator of the sample LMMSE is considered in this thesis to estimate the parameter of interest. It is based on the next linear transformation of the sample LMMSE and it is constrained to $N > M$,

$$\hat{x}(n) = \mathbf{w}^H \mathbf{y}(n);\ \mathbf{w} = \alpha \hat{\mathbf{R}}^{-1} \mathbf{s}. \tag{1.25}$$

Where, α is the parameter to be designed and it implements the correction of the sample LMMSE. Namely, the aim of α is to optimize the bias variance tradeoff to reduce the overall MSE, compared to the sample LMMSE method. Note that the SINR is insensitive to this scaling, and thus this type of filter it is not considered as a correction of the sample MVDR, which is based on optimizing that measure of performance. However, in terms of MSE, α leads to obtain an improved performance, compared to the sample LMMSE, as it will be shown in chapter 3. Moreover, it is important to stress that this scaling is important in applications where the aim is to obtain an to guarantee a good estimate of the signal amplitude, this is the case of e.g. subband beamforming [17]. Furthermore, within the context of wireless communications, this scaling can be interpreted as an automatic gain control which is necessary in any real MIMO system [76]. To design this scaling α, two

strategies are proposed. The first one seeks to obtain the scaling that optimizes the MSE of the parameter of interest after substituting (1.25) into the expression of the MSE stated above in (1.5). As this leads to an estimate that depends on the unknown $\mathbf{R}$, then RMT is used to obtain a consistent estimate of the optimal though unrealizable scaling factor within the regime where $M, N \to \infty$ and $M/N \to c \in (0, 1)$. Note that this asymptotic regime contemplates implicitly the small sample size regime, i.e. situations where N can be comparable to M, though assuming that $N > M$.

The second strategy designs the scaling correction α in (1.25) by means of multivariate statistical analysis tools. As the filter 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. That is, 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 obtains an optimal solution for the finite regime.

The numerical simulations in chapter 3 show that the shrinkage filter (1.25) obtains similar performance when the scaling factor α is designed using RMT or multivariate analysis tools. Moreover, it outperforms the sample LMMSE, which was the target of this initial basic shrinkage estimator. In addition, compared to other techniques that are robust to the small sample size, its performance is comparable to that of the LW technique based on the shrinkage of the SCM (1.16) and better than the ad-hoc DL technique in (1.13), provided that $M/N \in (0, 1)$. Although the comparison of the proposed method to those robust method is not simple, the rationale can be as follows. The proposed approach is more analytical than the DL technique in (1.13) and unlike the ad-hoc DL, the proposed method faces directly the estimation of the parameter of interest. This last statement can be also the rationale for obtaining similar performance than the LW (1.16) method, whose aim is to minimize the MSE in the estimation of the covariance.

Another contribution proposed in this thesis is a generalization of (1.25) to overcome the limitation $M/N \in (0, 1)$, i.e. to support $M \geqslant N$, and to obtain more insights in the comparison to the LW method. Namely, a scaling or shrinkage of a regularized sample LMMSE filter is proposed. That is to say, a double shrinkage is contemplated. The first one regularizes the SCM, i.e considers a covariance of the type $\check{\mathbf{R}} = \beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I}$ in the LMMSE expression and permits to deal with cases where $M > N$. The second shrinkage is a scaling of the LMMSE implemented with $\check{\mathbf{R}}$ and seeks to further reduce the MSE compared to only considering $\mathbf{w} = \check{\mathbf{R}}^{-1}\mathbf{s}$. Thus, the expression of the proposed estimator is,

$$\hat{x}(n) = \mathbf{w}^H \mathbf{y}(n);\ \mathbf{w} = \alpha \check{\mathbf{R}}^{-1} \mathbf{s}. \tag{1.26}$$

where $\check{\mathbf{R}}$ is a shrinkage of the SCM, i.e. $\check{\mathbf{R}} = \beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I}$. Intuitively (1.26) is better than (1.25) as $\check{\mathbf{R}}$ is a better estimate than the SCM. Moreover, thanks to the additional shrinkage of the filter one can improve the performance of an implementation of the LMMSE based on a shrinkage of the SCM, such as the LW (1.16) and the ad-hoc DL (1.13). The rationale for this improvement is that our method deals directly with the estimation of $x(n)$, whereas the ad-hoc DL and LW try to obtain a better estimation of the covariance. These intuitions are confirmed with the simulations in chapter 3, which show that the proposed approach not only outperforms dramatically the sample LMMSE, but also the first shrinkage of the sample LMMSE proposed above in (1.25) and the implementations of the LMMSE when considering robust techniques to the small sample size regime such as the ad-hoc DL (1.13) or LW estimations of the covariance (1.16). To achieve these results, the scalar α, which controls the shrinkage of the filter, is designed as the one minimizing the asympotic MSE of the parameter of interest, given a shrinkage of the SCM. That is, first the optimal shrinkage of the filter is obtained for a given shrinkage of the covariance. Then, as the optimal shrinkage depends on $\mathbf{R}$, a RMT approach is proposed to obtain a consistent estimate of the optimal shrinkage of the filter within the regime where $M, N \to \infty$ at a constant rate $M/N \to c \in (0, \infty)$. Note that this asymptotic regime considers intrinsically the small sample size situations. Regarding the scalars governing the shrinkage of the SCM, the optimal approach would be to substitute the optimal shrinkage of the filter, obtained previously, in the MSE expression and optimize again respect to the shrinkage factors of the SCM. Nonetheless, they could not be isolated due to their presence within the inverse of the covariance in the expression of the LMMSE. Indeed, neither one can use a numerical search method to propose a realizable estimator which obtain the optimal shrinkage factors due to the unknown $\mathbf{R}$ in the expression of the MSE. To circumvent this problem and to obtain a realizable filter, one could proceed as in [9], find the consistent estimate for the asymptotic MSE and find the α_1,α_2 minimizing it. Although a realizable filter is obtained, one still must carry out a numerical search to find α_1 and α_2. Instead of that an alternative approach is proposed. Namely, the shrinkage factors α_1,α_2 of the SCM are selected as the ones proposed by LW in [33], i.e. the ones minimizing the asymptotic MSE of the data covariance, given by the expressions (1.18) and (1.19) described above. This approach permits to obtain a more clear insight in the comparison of the proposed shrinkage method to the LW method.

Next, assuming that M may be comparable to N, but $M < N$, a more general form of shrinkage estimator is studied. Unlike the previous proposed methods, this shrinkage applies not only for the sample LMMSE but also for the sample MVDR. It is based on a

linear combination or weighted average between the sample filters and an a priori guess of the filter. This a priori guess is obtained by substituting the unknown noise covariance in the LMMSE and MVDR filters by the identity matrix. In array processing this guess filter corresponds to the case where only the SOI and additive noise are presumed to be in the scenario and thus corresponds to the conventional or Bartlett beamformer, i.e. it is a kind of matched filter. From a shrinkage point of view, with this strategy we are blending by means of a weighted average the sample estimators of the LMMSE or MVDR filters, which take into account the available information of the samples, with the conventional beamformer, which does not take into account the information of the samples and only relies on the a priori information. The combination of these two type of informations seeks to optimize the bias variance tradeoff and to diminish the overall MSE compared to just considering the sample LMMSE or MVDR. Thus, the expression of the proposed shrinkage filter reads,

$$\hat{x}_s(n) = \mathbf{w}^H \mathbf{y}(n); \;\; \mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s} \tag{1.27}$$

The filter (1.27) can be rewritten as $\mathbf{w} = (\alpha_1 \hat{\mathbf{R}}^{-1} + \alpha_2 \mathbf{I})\mathbf{s}$. This highlights that (1.27) is carrying out a shrinkage of the inverse of the SCM. Next, the shrinkage factors or scalings that perform the linear combination are designed by optimizing the MSE and the SINR, in the case of the LMMSE and the MVDR respectively. Direct optimization of these metrics leads to optimal scalings dependent on the true but unknown correlation $\mathbf{R}$, i.e. to unrealizable methods. To overcome this problem RMT results are used to obtain consistent estimates of the optimal weights within the regime where $M, N \to \infty$ at a constant rate $M/N \to c \in (0, 1)$. Note that this approach naturally deals with the small sample size regime. Interesting enough, the numerical simulations in chapters 3 and 4 show that the shrinkage of the sample LMMSE or MVDR towards the conventional beamformer outperforms the sample counterparts. Moreover, they also improve robust techniques to the small sample size regime such as LW (1.16) and the ad-hoc DL techniques (1.13) and (1.12) provided that $M/N \to c \in (0, 1)$. This is thanks to face directly the estimation of the parameter of interest, instead of trying to enhance the covariance estimate. Also compared to most of the DL techniques, that are rather ad hoc, here an analytical expression is given for the optimal and asymptotically optimal filters. Compared to the DL in [9], which obtains an asymptotically optimal filter in terms of SINR, the next conclusions are obtained from the simulations of chapters 3 and 4. The proposed shrinkage of the sample LMMSE obtains better performance in terms of MSE than the DL in [9], thanks to focusing on the optimization of the MSE, recall that as stated above the MSE is important in applications where the aim is to obtain a good estimate of the complex amplitude of the SOI, e.g. in subband beamforming [17]. Moreover, the shrinkage of the LMMSE and

MVDR obtain similar performance than the DL in [9] in terms of SINR. Moreover, in the numerical simulations of chapter 5 it is observed that the proposed shrinkage filters are more robust to uncertainties in $\mathbf{s}$ in the small sample size regime than the DL in [9], in terms of SINR. The reason is that the DL factor of [9] is designed to counteract the small sample size regime though assumes a known $\mathbf{s}$, thus it suffers a degradation when the knowledge of $\mathbf{s}$ is imprecise as an additional tuning of the loading factor is required. On the other hand, the proposed shrinkage filters tend to a scaling of the conventional beamformer in the small sample size. As the main lobe is wider than the one of the MVDR or LMMSE it is less sensitive to uncertainties in $\mathbf{s}$, though it has less resolution. Or in other words, the DL filters may undergo the signal cancelation effect due to the additional tuning required in the loading factors, i.e. they may tend to cancel the SOI as they confuse it as an interference. On the other hand, the conventional beamformer can attenuate the SOI when the knowledge of $\mathbf{s}$ is imprecise, though it does not undergo the signal cancelation effect.

The next contribution is the generalization of the type of shrinkage filters proposed in (1.27) to deal with cases where $M \geqslant N$. To achieve this aim, a regularization of the SCM is considered in the type of shrinkage filters proposed in (1.27). Thereby, the expression of the shrinkage filters reads as follows,

$$\hat{x}(n) = \mathbf{w}^H \mathbf{y}(n); \quad \mathbf{w} = \alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \mathbf{s} + \alpha_2 \mathbf{s} \tag{1.28}$$

In fact, note that this filter leads to improve (1.27) even when $M < N$. This is because it relies on a regularization of the SCM, which is a better estimate of $\mathbf{R}$ than the SCM. Moreover, it is important to stress that the filter proposed in (1.28) is the most complete or general shrinkage filter proposed in this thesis, as all the other filters proposed in (1.25)-(1.27) can be viewed as a particular case of (1.28). For any given δ, the shrinkage factors $\alpha_1,\ \alpha_2$ are obtained by optimizing the MSE or the SINR, in the case of the shrinkage of the regularized LMMSE or the regularized MVDR, respectively. Moreover, the regularization parameter δ is obtained by first substituting the optimal $\alpha_1,\ \alpha_2$ in the performance metrics, i.e. in the MSE or the SINR, and second by optimizing the resulting expression of the metric. As the optimal expressions of $\alpha_1,\ \alpha_2$ and δ depend on the unknown $\mathbf{R}$, then the next RMT approach is followed to obtain realizable filters. (M, N)-consistent estimates of the optimal $\alpha_1,\ \alpha_2$ and δ are obtained within the asymptotic regime where $M, N \to \infty$ and $M/N \to c \in (0, \infty)$. That is the estimates tend to the optimal $\alpha_1,\ \alpha_2$ and δ within that asymptotic regime. The numerical simulations of chapter 3 show that the proposed shrinkage of the regularized sample LMMSE outperforms the type of DL filter proposed in [9] and the regularized LMMSE in [79] in terms of MSE. These

filters are important because, bearing in mind the asymptotic regime $M, N \rightarrow \infty$ and $M/N \rightarrow c \in (0, \infty)$, [9] obtains the loading factor which optimizes asymptotically the SINR and [79] optimizes asymptotically the regularization parameters in terms of the MSE, see section 1.4.1. Furthermore in terms of SINR the proposed shrinkage of the regularized LMMSE obtains similar performance than [9] and [79]. Moreover, in chapter 5 it is shown that the proposed shrinkage of the regularized LMMSE outperforms, in terms of SINR, the DL in [9] and the regularized LMMSE in [79] when there is an uncertainty in the signature vector of the SOI. With regard to the shrinkage of the regularized MVDR it is shown in chapter 4 that it obtains the same performance in terms of SINR than the DL [9] and better performance than the ad-hoc DL and LW techniques mentioned above in (1.13) and (1.16), respectively. Moreover, in chapter 5 it is shown that the proposed shrinkage of the regularized MVDR outperforms the DL [9] when there is an uncertainty in the signature vector of the SOI. The rationale for the performance improvement of the proposed methods compared to the related work estimators is detailed in the numerical simulations of chapters 3, 4 and 5.

The design of the proposed shrinkage filters in (1.25)-(1.28) involve to solve certain problems that are tackled in detail in chapters 3 and 4. These problems can be summarized in compact form, by means of the next problem statement.

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 minimizes a functional* $f(\text{MSE}(\mathbf{w}))$ *of the* MSE *in (1.5) when the estimation is based on the type of shrinkage filters in (1.25), (1.26), (1.27) or (1.28) and subject to a set of constraints* $\mathcal{C}$ *on* $\mathbf{w}$*. This problem is mathematically formulated as follows,*

$$\hat{x}(n) = \mathbf{w}^H \mathbf{y}(n); \qquad \mathbf{w} = \underset{\substack{\mathbf{w} \\ s.t.\ \mathbf{w} \in \mathcal{C}}}{\arg\min}\ f(\text{MSE}(\mathbf{w})) \tag{1.29}$$

where, according to (1.5), $\text{MSE}(\mathbf{w}) = \mathbf{w}^H \mathbf{R} \mathbf{w} + \gamma(1 - \mathbf{w}^H \mathbf{s} - \mathbf{s}^H \mathbf{w})$.

Remark 1: The different shrinkage methods of the sample LMMSE proposed in chapter 3, i.e. the ones arising from (1.25), (1.26), (1.27) or (1.28), impose the constraint $\mathbf{w} \in \mathcal{C} \Leftrightarrow \mathbf{w} = \alpha \hat{\mathbf{R}}^{-1} \mathbf{s}$, $\mathbf{w} = \alpha \breve{\mathbf{R}}^{-1} \mathbf{s}$, $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1} \mathbf{s} + \alpha_2 \mathbf{s}$ or $\mathbf{w} = \alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \mathbf{s} + \alpha_2 \mathbf{s}$ respectively.

Remark 2: In the case of the filters based on a shrinkage of the sample MVDR proposed in chapter 4 and arising from (1.27) or (1.28), the next constraint is imposed $\mathbf{w} \in \mathcal{C} \Leftrightarrow \mathbf{w}^H \mathbf{s} = 1$ and $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1} \mathbf{s} + \alpha_2 \mathbf{s}$ in the case of (1.27) or $\mathbf{w}^H \mathbf{s} = 1$ and $\mathbf{w} = \alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \mathbf{s} + \alpha_2 \mathbf{s}$ in the case of (1.28). Moreover, note that this problem is equivalent to optimize the SINR at the output of the shrinkage filters.

Remark 3: Moreover, $f(\text{MSE}(\mathbf{w})) = \text{MSE}(\mathbf{w})$, for all the methods proposed in chapters 3 and 4, except for the shrinkage filter proposed in section 3.3. As the optimization of $\text{MSE}(\mathbf{w})$ yields unrealizable methods, subsequently random matrix theory tools are used to obtain $(M, N)-$consistent estimates of the optimal methods. That is, estimates that tend to the optimal methods within the asymptotic regime where both M, N tend to infinity at a constant rate. Unlike classical asymptotics where M remains fixed and N grows large this general asymptotic regime deals naturally with the small sample size regime. Another advantage of this RMT approach is that it does not rely on any assumption about the distribution of the observations. On the other hand, the method proposed in section 3.3 is an alternative to the RMT approach used by the rest of the methods. Namely, for the particular case where $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$, we consider the average MSE to design the proposed method, provided that there are not other constraints in $\mathcal{C}$ and that the observations are gaussian. Namely, note that the MSE in (1.5) is obtained for a generic filter $\mathbf{w}$ and a generic data sample $\mathbf{y}(n)$, whose optimization leads to the LMMSE depending on the unknown $\mathbf{R}$. To obtain a realizable filter, one considers N available snapshots or training samples $\mathbf{Y} = [\mathbf{y}(1), \ldots, \mathbf{y}(N)]$, which permit to build the SCM and the proposed filter $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$, which is substituted in the MSE to evaluate its performance. Thereby, the MSE performance associated to the filter $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$ is a random quantity as it depends on the available snapshots $\mathbf{Y}$, which are samples of a random process. Thereby, one can obtain the average MSE performance of $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$ by means of the next expectation $\mathbb{E}_{\mathbf{Y}\mathbf{Y}^H}\left[\text{MSE}\left(\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}\right)\right]$, thereby $f(\text{MSE}(\mathbf{w})) = \mathbb{E}_{\mathbf{Y}\mathbf{Y}^H}\left[\text{MSE}\left(\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}\right)\right]$ in (1.29).

1.5.2 Shrinkage of the sample MVDR to cope with the small sample size and steering vector uncertainties

The contribution summarized in this section, and exposed in detail in chapter 5, deals with the two sources of degradation of the MVDR, the small sample size impairment and uncertainties in the signature vector of the SOI $\mathbf{s}$ in (1.1). This permits to generalize the contribution of the last section which assumes a known $\mathbf{s}$.

Namely, in order to deal with the small sample size effect, the shrinkage MVDR filter (1.27) is taken into account and random matrix theory is used to obtain consistent estimates of the expressions depending on $\mathbf{R}$. Moreover, the shrinkage MVDR filter is extended to support a mismatch in the assumed $\mathbf{s}$. To this end, following the approach of [37], the true but unknown steering vector of the SOI, $\tilde{\mathbf{s}} = \mathbf{s} + \boldsymbol{\Delta}$, is assumed to lie within an uncertainty region, where $\boldsymbol{\Delta}$ is an unknown distortion vector. Namely, the type of uncertainty set $\mathcal{A}(\varepsilon)$ described above in (1.22) is considered,

$$\mathcal{A}(\varepsilon) = \{\mathbf{a} | \mathbf{a} = \mathbf{s} + \mathbf{e}, \|\mathbf{e}\| \leq \varepsilon\}.$$

Thereby, the shrinkage MVDR method is made robust to uncertainties in $\mathbf{s}$, by adding a new constraint in the formulation of the MVDR problem. Namely, the new optimization problem requires a no distortion constraint for all the steering vectors within the uncertainty set $\mathcal{A}(\varepsilon)$. As it will be shown in chapter 5 this approach permits to outperform the shrinkage MVDR method that assumes a known $\mathbf{s}$, i.e. the one stemming from (1.29) and dealt with in chapter 4. Moreover, it leads to improve the DL method [9], which obtains the asymptotically optimal loading factor (1.15) to combat the small sample size regime, though assuming a known $\mathbf{s}$. Compared to the robust MVDR method [37], which was the first proposing to deal with the uncertainties in $\mathbf{s}$ by incorporating the no distortion constraint for all the steering vectors in $\mathcal{A}(\varepsilon)$ in the MVDR, our method offers the next benefit. Thanks to the shrinkage structure of the filter and to an approach based on RMT, the finite sample size effect is dealt with directly, which avoids the parameter tuning required in [37] to treat properly this effect. Namely, [37] can be interpreted as a DL, but its loading factor only depends on the parameter ε which models the uncertainty in $\mathbf{s}$. Therefore, [37] may offer an insufficient protection against the small sample size regime in some situations and it will require an ad-hoc parameter tuning of the loading factor.

According to the above description, the next optimization problem needs to be solved to obtain the proposed robust shrinkage MVDR filter, which permits to deal both with the finite sample size effect and uncertainties in $\mathbf{s}$.

Problem statement:

Consider a set of observations $\{\mathbf{y}(n)\}_{n=1}^{N}$ fulfilling the model in (1.1). Moreover, assume that the true signature vector of the SOI $\tilde{\mathbf{s}}$ differs from the presumed signature vector $\mathbf{s}$ in (1.1), namely $\tilde{\mathbf{s}} = \mathbf{s} + \boldsymbol{\Delta}$ where $\boldsymbol{\Delta}$ is an unknown distortion vector. In addition by assumption $\tilde{\mathbf{s}}$ lies within an uncertainty set $\mathcal{A}(\varepsilon) = \{\mathbf{a} | \mathbf{a} = \mathbf{s} + \mathbf{e}, \|\mathbf{e}\| \leq \varepsilon\}$, where ε is a design parameter. Then, obtain an estimation of the unknown parameter $x(n)$ in (1.1) that maximizes the SINR subject to a no distortion constraint for all the signature vectors within $\mathcal{A}(\varepsilon)$ and subject to the constraint that the filter $\mathbf{w}$ fulfills the shrinkage structure in (1.27). This problem is mathematically formulated as follows,

$$\begin{aligned} \hat{x}(n) = \mathbf{w}^H \mathbf{y}(n); \quad \mathbf{w} = \arg\min_{\mathbf{w}} \quad & \mathbf{w}^H \mathbf{R} \mathbf{w} \\ \text{subject to} \quad & |\mathbf{w}^H \mathbf{a}| \geqslant 1 \text{ for all } \mathbf{a} \in \mathcal{A}(\varepsilon), \\ & \mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1} \mathbf{s} + \alpha_2 \mathbf{s}. \end{aligned} \tag{1.30}$$

1.6 Organization of the Thesis and list of publications

The organization of the rest of this thesis is as follows. In chapter 2, an introduction to the main tools, used herein to obtain the proposed corrections of the sample LMMSE and MVDR methods to the small sample size regime, is presented. These tools are random matrix theory and shrinkage estimation. Moreover, chapter 2 presents some RMT results that are the cornerstone to obtain the proposed shrinkage filters in the subsequent chapters of this thesis. Chapter 3 obtains the shrinkage filters that counteract the small sample size degradation of the sample LMMSE. That is, it deals with the problem stated in (1.29) for the particular cases exposed in the remarks 1 and 3. 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 F. Rubio, "Asymptotically optimal linear bias corrections in minimum mean square error estimation," *Presentation at the International Conference on Trends and Perspectives in Linear Statistical Inference (LinStat'2010),* 27-31 July 2010, Tomar, (Portugal).
- 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).
- J. Serra and M. Nájar,"Double Shrinkage correction in sample LMMSE estimation". *in Proceedings of the European Signal Processing Conference (EUSIPCO 2013)*, 9-13 September 2013, Marrakech (Morocco).
- J. Serra and M. Nájar,"Asymptotically Optimal Linear Shrinkage of sample LMMSE and MVDR filters," IEEE Transactions on Signal Processing, vol. 62, no. 14, pp. 3552-3564, July 2014.
- J. Serra and M. Nájar,"On the Shrinkage of regularized sample LMMSE and MVDR filters", in preparation to be submitted to the IEEE Transactions on Signal Processing.

Chapter 4 proposes optimal shrinkage corrections for large sample MVDR filters, which are based on RMT. That is it deals with the problem stated in (1.29) for the particular

case exposed in the remark 2. The material of this chapter has been included in the next journal papers,

- J. Serra and M. Nájar, "Asymptotically Optimal Linear Shrinkage of sample LMMSE and MVDR filters," IEEE Transactions on Signal Processing, vol. 62, no. 14, pp. 3552-3564, July 2014.
- J. Serra and M. Nájar, "On the Shrinkage of regularized sample LMMSE and MVDR filters", in preparation to be submitted to the IEEE Transactions on Signal Processing.

The aim of chapter 5 is twofold. First, it evaluates the shrinkage filters proposed in chapters 3 and 4, as well as their related work, when there is both a finite sample size situation and a mismatch in the presumed signature vector of the SOI. This is important as the filters proposed in chapters 3, and 4 and their related work, assumed a known signature vector. The second aim of chapter 5 deals with the problem stated in (1.30), i.e. it proposes a shrinkage MVDR filter which deals directly with both the finite sample size impairment and the imprecise knowledge of the signature vector of the SOI. The work of this chapter can be found in,

- J. Serra and M. Nájar, "Robust Shrinkage MVDR beamforming," submitted to the IEEE Signal Processing Letters.
- J. Serra and M. Nájar, "On the Shrinkage of regularized sample LMMSE and MVDR filters", in preparation to be submitted to the IEEE Transactions on Signal Processing.

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

Chapter 2

Technical Background and Random Matrix Theory results

2.1 Introduction

The aim of this chapter is twofold, first the tools used to design the proposed estimators in the upcoming chapters are introduced. 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 introducing the most important features of these techniques in the context of this thesis. Namely, certain definitions and propositions motivating the use of these tools for the design of the proposed estimators will be dealt with. The second aim of this chapter is to present several RMT results that are the cornerstone for the design of the proposed estimators. The organization of this chapter is as follows. Section 2.2 introduces the RMT tool as well as the important RMT results used to derive the proposed estimators in the next chapters. Section 2.3 exposes the theory of shrinkage estimation.

2.2 Random Matrix Theory

2.2.1 Introduction

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 [87], [88] and [89]; the study of small and large random matrices with invariance properties (e.g. free probability theory [90] [91], combinatronics [92] [93] and Gaussian methods [94] [95]); And finally the study of large random matrices with independent entries [73], [96] and [97]. This latter topic is the one considered in this thesis.

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 [87]. 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 function of the eigenvalues, also known as Empirical Spectral Distribution (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 [98], showed that the expected value of the empirical eigenvalue distribution of a large random matrix, called Wigner matrix nowadays, tends to a deterministic distribution function with an associated density function known as the semi-circle law. With this work he founded the field of random matrix theory, which deals with the asymptotic study of the eigenvalue and eigenvector distribution 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, [96], which is commented below in the next section. Since then, a plethora of research have been conducted by researchers such as Bai or Silverstein, see [99] and references therein. For this 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. [100] or [101].

Random matrix theory has found applications in fields as diverse as nuclear physics [102], mathematical finance [103] or computational biology [104]. In wireless communica-

tions, information theory and signal processing, random matrix theory has found several applications, see [105], [106] and references therein for a thorough description of them. In this regard, the first contributions were probably the ones dealing with the multiplexing gain of multiple antenna communications [107] and the performance analysis of multi-user linear receivers in spread spectrum systems [108]. More in general, a possible, non-exhaustive, classification of the applications of RMT is described next. The first group deals with the performance study of wireless communications technologies and systems. The second and third group treat the signal detection and estimation problems in sensor networks and wireless communications systems. Examples of the performance analyis category are the studies of the capacity in multiple antenna systems for different types of channel models, such as time varying Rayleigh channels [107], Rician channels [109], or frequency selective channels [110]. An example of signal detection is [111], which deals with the performance study of hypothesis testing methods in sensor arrays, such as the GLRT, within the framework where the number of statistical samples may be comparable to the number of sensors. Finally, examples of the application of RMT to parameter estimation are the energy estimation of multiple sources in cognitive wireless networks [11], the performance study of subspace based methods to propose new subspace algorithms such as the G-MUSIC, see [12]. Also in this regard, it is worth mentioning the performance analysis of the sample estimates of eigenvalues and eigenvectors of covariance matrices [34], and the proposition of new estimators of them that cope with the performance degradation in the small sample size regime, see [73].

2.2.2 Large Dimensional Random Matrix Theory

In order to begin this section it is important to clarify some points. First, it is interesting to justify why it is attractive to study the distribution of eigenvalues associated to large dimensional random matrices instead of just focusing on the case of fixed size random matrices. In fact, probably the first result in RMT, i.e. [87], gave the expression of the pdf of a random matrix consisting of a central Wishart distribution. However, in general it is difficult to obtain the expression of the pdf of the eigenvalues associated to a fixed size random matrix and it requires extremely involved mathematical results [106]. This is exemplified in e.g. [106, Theorem 2.7] where the joint pdf of the unordered eigenvalues associated to a central Wishart matrix $\mathbf{X}\mathbf{X}^H$ is given, where the columns of $\mathbf{X}$ are iid Gaussian with zero mean and non-negative definite covariance $\mathbf{R}$. In contrast, in general large dimensional random matrix theory leads to much simple results, in the study of the distribution of the eigenvalues related to a random matrix. Moreover, as it is stated in [106, p. 29] large dimensional RMT leads to stunningly precise approximations of finite

case scenarios stemming from random matrices.

The Stieltjes Transform

In this thesis, results from large dimensional random matrix theory will be used to derive most of the proposed estimators. This discipline is devoted to study the asymptotic distribution of eigenvalues associated to random matrices, whose dimensions grow without bound. In order to proceed, let $\mathbf{X} \in \mathbb{C}^{M \times M}$ be a Hermitian random matrix[1] with associated eigenvalues $\{\lambda_i\}_{i=1}^{M}$. Then, the empirical distribution function of the eigenvalues, also known as empirical spectral distribution (ESD) in the RMT literature, is defined as follows[2]

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

Where # denotes the cardinality of a set and $\mathcal{I}$ denotes the indicator function. Then, one of the tasks of random matrix theory is to study the convergence, if it exists, of the random ESD $F^{\mathbf{X}}(\lambda)$ towards a limiting non-random probability distribution function, called limiting spectral distribution (LSD), when the dimensions of $\mathbf{X}$ increase without bound. In this regard, a tool of paramount importance is the Stieltjes transform, which is defined as follows,

Definition 2.1 *Let F be a real-valued distribution function and $z \in \mathbb{C}$ be taken outside the support of F. Then the Stieltjes transform of F at point z, denoted by $m_F(z)$, is defined as*[3]*,*

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

■

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

[1] The Hermitian property is needed to ensure that all the eigenvalues of $\mathbf{X}$ are real, though the extension to non-Hermitian matrices has been considered sometimes e.g. in [99, (1.2.2)].

[2] Although here $\mathbf{X}$ is defined as random, the ESD is defined for non random matrices in the same way [106].

[3] In general , the Stieltjes transform can be applied when F is a real valued bounded measurable function over $\mathbb{R}$, see [106].

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 the empirical eigenvalue distribution function of* $\mathbf{X}$ *as defined in (2.1). Then, the Stieltjes transform of* $F^{\mathbf{X}}$, *denoted by* $m_{\mathbf{X}}$ *is given by,*

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

■

The Stieltjes transform allows to simplify the asymptotic analysis of the eigenvalue distribution. That is, 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 function, say $F^{\mathbf{B}}$, towards a limiting eigenvalue distribution function, say $F^{\mathbf{L}}$, a possible procedure is as follows. First, the Stieltjes transform of $F^{\mathbf{B}}$, say $m_{\mathbf{B}}$, is found. Then, one founds that $m_{\mathbf{B}}$ converges to the Stieltjes 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 Stieltjes transform, defined below, and the next property of the Stieltjes transform [106, Theorem 3.10], where $\Rightarrow$ denotes weakly convergence and $\xrightarrow{a.s.}$ denotes almost sure convergence,

$$F^{\mathbf{B}} \Rightarrow F^{\mathbf{L}} \Leftrightarrow m_{\mathbf{B}} \xrightarrow{a.s.} m_{\mathbf{L}}. \tag{2.4}$$

Definition 2.3 *For all distribution functions* $F(x)$ *which admit a Stieltjes Transform, then the inverse Stieltjes Transform is defined for all* x *where* $F(x)$ *is continuous as follows,*

$$F(x) = \frac{1}{\pi}\lim_{y\to 0^+}\int_{-\infty}^{x}\operatorname{Im}\{m_F(x+iy)\}dx \tag{2.5}$$

■

The procedure just described above to find the LSD $F^{\mathbf{L}}$, based on finding the convergence of the Stieltjes transform associated to the ESD, is applied for instance to obtain the Marčenko-Pastur law, which is exposed next. The Marčenko-Pastur law is one of the most remarkable and seminal works in RMT and it illustrates the spreading phenomenon of the eigenvalues associated to a Gram matrix built from a random matrix with iid entries with zero mean and unit variance. That is, a basic sample covariance matrix associated to a

population covariance given by the identity matrix. This spreading phenomenon is more pronounced as the sample size becomes smaller as it is shown below in figure 2.1 and it can be viewed as a motivation to propose corrections of estimators relying on the SCM in the small sample size regime, which is the topic treated in this thesis.

Proposition 2.1 *(Marčenko Pastur Law) [96] [35, 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* $\mathbf{X}$ *be a random matrix with i.i.d entries such that the real and imaginary parts of the entries are independent with zero mean and variance 1/2. Then, as* $M, N \to \infty$ *with* $M/N \to c < +\infty$*, the Stieltjes transform of the empirical eigenvalue distribution function* $F^{\mathbf{B}}$ *associated to* $\mathbf{B}$*, denoted by* $m_{\mathbf{B}}$*, converges almost surely to the Stieltjes transform* $m_{\mathbf{L}}$*, associated to the LSD of* $\mathbf{B}$*. Moreover, the ESD of* $\mathbf{B}$ *converges weakly and almost surely to a non-random distribution function* $F_{\mathbf{L}}$ *with an associated pdf* $f_{\mathbf{L}}$ *described next.*

$$m_{\mathbf{L}}(z) = \frac{1-c-z+\sqrt{(z-1-c)^2-4c}}{2cz}, \forall z \in \mathbb{C}^+ \tag{2.6}$$

$$f_{\mathbf{L}}(x) = \begin{cases} \max(0, 1-c^{-1})\delta(x) + \frac{1}{2\pi cx}\sqrt{(b-x)(x-a)} & \text{if } a \le x \le b \\ 0, & \text{otherwise} \end{cases} \tag{2.7}$$

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

■

Indeed, the Marčenko-Pastur law is a particular case of [35, Theorem 1.1], which provides the Stieltjes transform associated to the limiting eigenvalue distribution function of random matrices of the form $\mathbf{B} = \frac{1}{N}\mathbf{R}^{1/2}\mathbf{X}\mathbf{X}^H\mathbf{R}^{1/2}$. Where $\mathbf{R}$ is assumed to be a hermitian square positive-definite matrix, whose eigenvalues are uniformly bounded for all M, and the entries of $\mathbf{X}$ are assumed to 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 method.

In figure 2.1 the Marčenko-Pastur law is exemplified by displaying the eigenvalue pdf in (2.7) for different values of c, namely c =0.1, 0.2 and 0.5. One can see that when $c \to 0$ 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 $\mathbf{B} \to \mathbf{I}_M$,

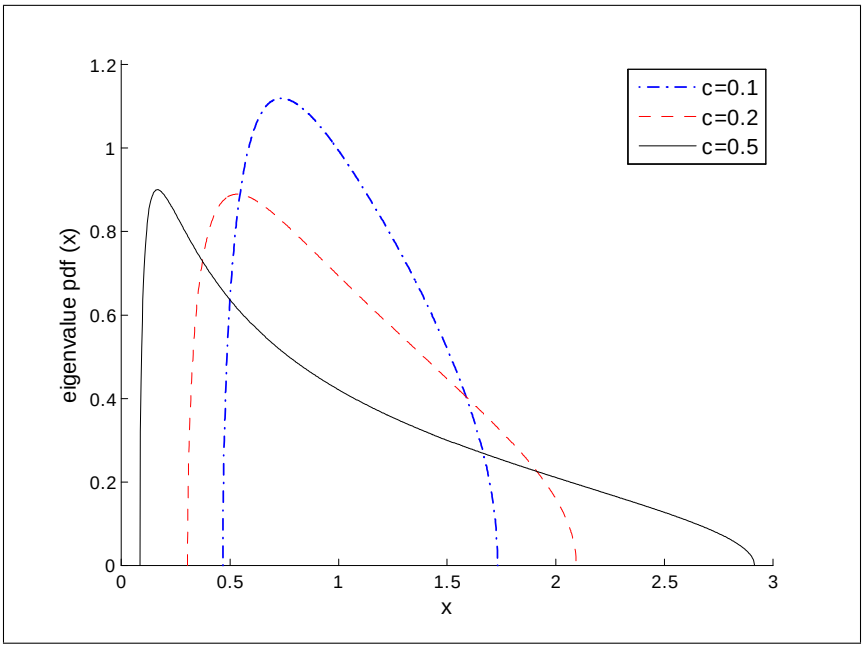

Figure 2.1: Marčenko-Pastur Law: Limiting eigenvalue pdf in (2.7) for different values of $c = M/N$.

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.

Following with the illustration of the Marčenko-Pastur law, in figure 2.2 we represent the histogram of the eigenvalues of $\mathbf{B}$, defined in proposition 2.1, and the associated limiting pdf predicted by the Marčenko-Pastur law (2.7), 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.

The Stieltjes transform presented in (2.3) is appropiate to study the asymptotic behavior of eigenvalues. Nonetheless, for the purposes of this 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 square hermitian random matrix $\mathbf{X} \in \mathbb{C}^{M \times M}$, which was introduced by Girko in e.g. [112] and it is a generalization of the empirical eigenvalue distribution function in (2.1),

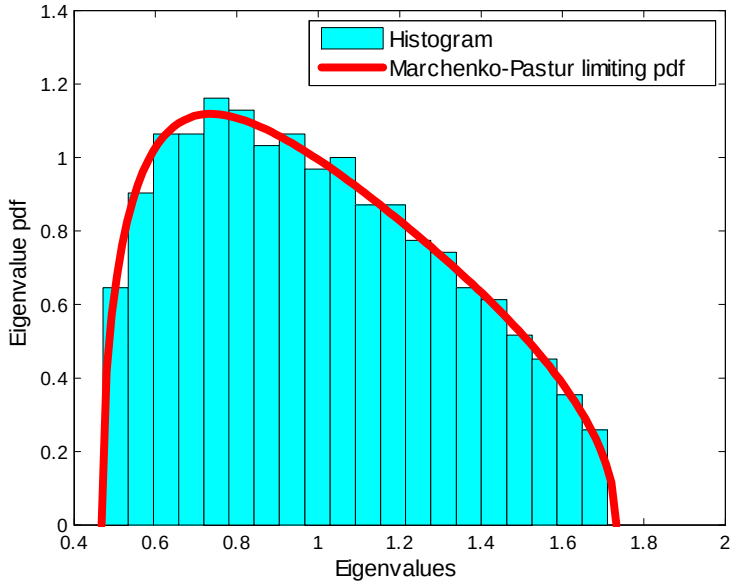

Figure 2.2: Comparison between eigenvalue pdf of $\mathbf{B}$ and its limiting pdf given by Marčenko Pastur law in (2.7), when $M = 500$, $N = 5000$ and $c = 0.1$.

$$G^{\mathbf{X}}(\lambda) = \sum_{m=1}^{M} \mathbf{a}^H \mathbf{e}_m \mathbf{e}_m^H \mathbf{b} \mathcal{I}_{\lambda_m \leq \lambda}(\lambda) \tag{2.8}$$

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 Stieltjes transform given by the next expression, see [112],

$$\mathbf{a}^H (\mathbf{X} - z\mathbf{I}_M)^{-1} \mathbf{b} = \sum_{m=1}^{M} \frac{\mathbf{a}^H \mathbf{e}_m \mathbf{e}_m^H \mathbf{b}}{\lambda_m - z} \tag{2.9}$$

Now we can see the importance of (2.9). 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, (2.9) 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. [100]. 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 $M/N \to c \in (0, \infty)$. For instance, 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. [101]. 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 Stieltjes transform of a certain spectral function, analogous to its complex counterpart in (2.9),

$$\mathcal{T}_{\mathbf{R}}(x) = \mathbf{a}^H (\mathbf{I}_M + x\mathbf{R})^{-1} \mathbf{b} = \sum_{m=1}^{M} \frac{\mathbf{a}^H \mathbf{e}_m \mathbf{e}_m^H \mathbf{b}}{1 + x\lambda_m}, \quad x \in \mathbb{R}^+ \tag{2.10}$$

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 $\mathbf{R} \in \mathbb{C}^{M \times M}$, respectively. Then, an (M, N)-consistent estimator of (2.10) when $M, N \to \infty$ and $M/N \to c \in (0, \infty)$ reads as follows [101],

$$\hat{\mathcal{T}}_{\mathbf{R}}(x) = \mathbf{a}^H (\mathbf{I}_M + \theta(x)\hat{\mathbf{R}})^{-1} \mathbf{b} \tag{2.11}$$

where $\theta(x)$ is the positive solution to 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 (2.10). As a consequence, the

G_{25}-estimator paves the way to obtain the (M, N)-consistent estimators of those quantities. For instance, the inverse correlation matrix $\mathbf{R}^{-1}$ or quadratic forms of $\mathbf{R}^k, k > 0$, see [9],

$$\left[\mathbf{R}^{-1}\right]_{i,j} = \lim_{x\to\infty} x\mathbf{u}_i^H(\mathbf{I}_M + x\mathbf{R})^{-1}\mathbf{u}_j \tag{2.12}$$

$$\mathbf{a}^H\mathbf{R}^k\mathbf{b} = \frac{(-1)^k}{k!}\left[\frac{d^k}{dx^k}\mathbf{a}^H(\mathbf{I}_M + x\mathbf{R})^{-1}\mathbf{b}\right]_{|_{x=0}} \tag{2.13}$$

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 quantity is as follows. First, one expresses the parameter to be estimated, say ϕ, as a function of the Stieltjes transform associated to 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 Stieltjes transform associated to the available matrix of observations $\mathbf{Y}$, i.e. $m_{\mathbf{T}} \asymp g(m_{\mathbf{Y}})$, where $\asymp$ denotes usually almost surely convergence. Finally, one estimates ϕ as $\hat{\phi} = f((g(m_{\mathbf{Y}}))$.

2.2.3 Random Matrix Theory Results

In this section asymptotic equivalences that pave the way to obtain (M, N)-consistent estimators in the subsequent chapters of this thesis are presented. Namely, they build upon large dimensional random matrix theory and they 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_n denote a sequence of random variables and b denote a deterministic quantity. Let $a_n \asymp b$ mean that a_n and b are asymptotic equivalents, i.e. $|a_n - b| \to 0$ in probability. Let $\mathbf{s} \in \mathbb{C}^M$ be a generic deterministic vector with uniformly bounded norm. Moreover, assume that $\mathbf{R} \in \mathbb{C}^{M\times M}$ denotes a generic positive definite correlation matrix whose eigenvalues are uniformly bounded for all M and they have a limiting spectral distribution. Let $\hat{\mathbf{R}} \in \mathbb{C}^{M\times M}$ denote the sample estimate of $\mathbf{R}$ 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,*

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

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

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp (1-c)^{-3}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} \tag{2.16}$$

$$\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s} \asymp \mathbf{s}^H\mathbf{R}\mathbf{s} \tag{2.17}$$

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

■

Lemma 2.2 *Let $a \asymp b$ mean that a and b are asymptotic equivalents, i.e. $|a-b| \to 0$ in probability. Let $\mathbf{s} \in \mathbb{C}^M$ be a generic deterministic vector with uniformly bounded norm. Moreover, assume that $\mathbf{R} \in \mathbb{C}^{M\times M}$ denotes a generic positive definite correlation matrix whose eigenvalues are uniformly bounded for all M and they have a limiting spectral distribution. Let $\hat{\mathbf{R}} \in \mathbb{C}^{M\times M}$ denote the sample estimate of $\mathbf{R}$ as defined in (1.8). Finally, let denote $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta\mathbf{I}$, with δ a deterministic real positive value. Then, considering the asymptotic regime where $M, N \to \infty$ at a constant rate $M/N \to c \in (0,\infty)$, the next equivalences hold,*

$$\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} \asymp \frac{1}{\left(1-\frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}]\right)^2}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s} \tag{2.18}$$

$$\mathbf{s}^H\mathbf{R}\mathbf{s} \asymp \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}. \tag{2.19}$$

$$\mathbf{s}^H\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} \asymp \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\mathbf{s}^H\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s} \tag{2.20}$$

$$\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \asymp \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\mathbf{s} \tag{2.21}$$

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 [83], 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 [84], [85] and [113–117]. See also [82] 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 [82], 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,*

$$\hat{\boldsymbol{\mu}}^s = (1-\alpha)\hat{\boldsymbol{\mu}} + \alpha \mathbf{b} \tag{2.22}$$

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 [118, Appendix 4.5].

■

Indeed, there are other possible choices for $\mathbf{b}$ other than a constant vector. E.g. Jorion in [119] 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 [120].

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 (2.22), 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 [121] 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 [122] 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 [123] and references therein. Moreover, James-Stein estimation has been applied to other fields of science such as quantitative finance [119].

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 [84]. More recently, Ledoit and Wolf proposed in [33] a shrinkage estimator of the SCM $\hat{\mathbf{R}}$, consisting of shrinking $\hat{\mathbf{R}}$ towards a scaled 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. For instance, Stoica, Guerci et al. [41], 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 [59], 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,

$$\begin{aligned} &\min_{\rho} \mathbb{E}\left[\left\|\check{\mathbf{\Sigma}}-\mathbf{\Sigma}\right\|_F^2\right] \\ &s.t. \check{\mathbf{\Sigma}} = (1-\rho)\hat{\mathbf{\Sigma}} + \rho\mathbf{\Sigma}_0 \end{aligned} \tag{2.23}$$

Where $\|\cdot\|_F$ *denotes the Frobenius norm. Then, an (M,N)-consistent estimator of the optimal, though unrealizable, solution to (2.23) 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,*

$$\begin{aligned} \check{\mathbf{\Sigma}}^{LW} &= (1-\hat{\rho}^{LW})\hat{\mathbf{\Sigma}} + \hat{\rho}^{LW}\mathbf{\Sigma}_0 \\ \hat{\rho}^{LW} &= \frac{\sum_{n=1}^{N}\left\|\mathbf{x}_n\mathbf{x}_n^T - \hat{\mathbf{\Sigma}}\right\|_F^2}{N^2\,\mathrm{Tr}\left[\hat{\mathbf{\Sigma}}^2\right] - \frac{\mathrm{Tr}^2[\hat{\mathbf{\Sigma}}]}{M}} \end{aligned} \tag{2.24}$$

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

Proof: For a proof of this proposition the reader is referred to [33].

■

Appendix: Proof of the useful equivalences

Proof of (2.14) and (2.15).

The proofs of these asymptotic equivalences are based on [34, Theorem 1]. First, note that $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}$, $\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}$ can be expressed in terms of a Stieltjes transform $\hat{m}(z)$ as follows,

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

$$\begin{aligned} \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} &= \lim_{z\to 0}\hat{m}(z) \\ \hat{m}(z) \text{ s.t. } \mathbf{a} &= \mathbf{s},\ \mathbf{b} = \mathbf{R}\mathbf{s} \end{aligned} \tag{2.25}$$

$$\begin{aligned} \mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} &= \lim_{z\to 0}\hat{m}(z) \\ \hat{m}(z) \text{ s.t. } \mathbf{a} &= \mathbf{R}\mathbf{s},\ \mathbf{b} = \mathbf{s} \end{aligned} \tag{2.26}$$

$$\begin{array}{c} \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} = \lim\limits_{z\to 0} \hat{m}(z) \\ \hat{m}(z) \text{ s.t. } \mathbf{a}=\mathbf{s},\ \mathbf{b}=\mathbf{s} \end{array} \tag{2.27}$$

Therefore, in order to proof (2.14) and (2.15) the asymptotics of $\hat{m}(z)$ must be studied. This is provided by means of [34, 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 $0 < c < \infty$,

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

$$\bar{m}(z) = \sum_{m=1}^{M} \frac{\mathbf{a}^H \mathbf{e}_m \mathbf{e}_m^H \mathbf{b}}{\lambda_m(1-c-cz\bar{b}(z))-z} \tag{2.28}$$

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

$$\bar{b}(z) = \frac{1}{M}\sum_{m=1}^{M} \frac{1}{\lambda_m(1-c-cz\bar{b}(z))-z} \tag{2.29}$$

Now, recalling the relations in (2.25), (2.26), (2.27) and 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 $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}$, $\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}$ which concludes the proof,

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

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

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

■

Proof of (2.18)

The proof of the equivalence in (2.18) is based on the results obtained in [9, Appendix I] and [124]. Namely, the proof is divided in two parts. In the first part, the convergence

of $\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$ towards a deterministic function of $\mathbf{R}$ is obtained. This result was found in [9, Appendix I] and it relies on expressing $\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$ in terms of a Stieltjes transform associated to the type of random matrices $\mathbf{B} = \alpha\mathbf{R}^{-1}+\mathbf{\Phi}\mathbf{\Phi}^H$, where $\mathbf{\Phi}$ is an $M\times N$ random matrix with iid complex entries with zero mean and variance $1/2N$. The second part of the proof obtains the (M,N)-consistent estimate of the function of $\mathbf{R}$ obtained in the first part of the proof. Thereby it paves the way to obtain the (M,N)-consistent estimate of $\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$. This result was obtained in [124]. Next, we give the details of the overall proof. First the next definition is used,

$$\eta_n(\delta) \ = \mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} \tag{2.30}$$

where recall that $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta\mathbf{I}$. Now, assume that the SCM can be related to the theoretical covariance $\mathbf{R}$ through the expression,

$$\hat{\mathbf{R}} = \mathbf{R}^{1/2}\mathbf{\Phi}\mathbf{\Phi}^H\mathbf{R}^{1/2} \tag{2.31}$$

where $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$ is a random matrix with iid entries and whose real and imaginary parts are independent, have zero mean and variance $1/2N$. The assumption that the equality (2.31) holds is usually accepted in the random matrix theory literature, see e.g. [9, 34, 73]. Namely, as it is pointed out in [9, Appendix I], which has the same data model than the one assumed herein in (1.1), it can be easily shown that (2.31) holds if the next two assumptions are fulfilled. First, the signal of interest $x(n)$ in (1.1) is a sequence of iid complex random variables with independent real and imaginary parts, which have zero mean, variance $\gamma/2$ and bounded moments. Second, the interference plus noise $\mathbf{n}(n)$ in (1.1) is a sequence of iid complex random vectors, which are statistically independent of $x(n)$, whose real and imaginary parts are independent, have zero mean, posititve definite covariance matrix $\mathbf{R}_n/2$ and bounded moments. Taking into account (2.31) in the expression (2.30), it is easily shown that (2.30) can be expressed in terms of the next function,

$$m(z) = \mathbf{s}^H\mathbf{R}^{-1/2}(\mathbf{\Phi}\mathbf{\Phi}^H+\delta\mathbf{R}^{-1} - z\mathbf{I}_M)^{-1}\mathbf{R}^{-1/2}\mathbf{s}. \tag{2.32}$$

Namely, the next relation holds between $\eta_n(\delta)$ and $m(z)$,

$$\eta_n(\delta) = \frac{dm(z)}{dz}|_{z=0}. \tag{2.33}$$

It turns out that (2.32) has the form of a Stieltjes transform of the type $\mathbf{a}^H(\mathbf{M}-z\mathbf{I})^{-1}\mathbf{a}$ introduced above in (2.9). Thereby, due to the relation in (2.33), the convergence of the expression of interest $\eta_n(\delta) = \mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$ implies finding the convergence of the Stieltjes transform $m(z)$ associated to the type of random matrices $\mathbf{M} = \mathbf{\Phi}\mathbf{\Phi}^H + \delta\mathbf{R}^{-1}$. This result was found in [9, Lemma 1 in Appendix I], which is stated in the next lemma.

Lemma 2.3 *Let* $\mathbf{M} = \mathbf{\Phi}\mathbf{\Phi}^H + \delta\mathbf{R}^{-1}$ *with* $\delta\mathbf{R}^{-1}$ *a positive definite Hermitian matrix of dimensions* $M \times M$*, whose eigenvalues are strictly positive and uniformly bounded (i.e. there exists a constant that bounds all the eigenvalues) and they have a limiting distribution function as* M *grows large. Moreover, assume that* $\mathbf{\Phi} \in \mathbb{C}^{M\times N}$ *is a complex random matrix with iid entries and whose real and imaginary parts are independent, have zero mean, variance* $1/2N$ *and bounded moments. Let denote* $\mathbf{A} = \delta\mathbf{R}^{-1}$ *and* $\lambda_M(\mathbf{A})$ *the maximum eigenvalue of* $\mathbf{A}$*. Moreover, consider* D_ϵ *an open disk centered at* $z = 0$ *with radius* $\epsilon < \lambda_M(\mathbf{A})/3$*. Finally, suppose that the vector* $\mathbf{a} = \mathbf{R}^{-1/2}\mathbf{s}$ *has a uniformly bounded norm and it fulfills the next assumption,* $\sup_M \sup_{z\in D_{2\epsilon}} |\mathbf{a}^H(\delta\mathbf{R}^{-1} - z\mathbf{I})^{-1}\mathbf{a}| < \infty$*. Then, assuming the asymptotic regime where* $M, N \to \infty$ *and* $M/N \to c \in (0, \infty)$*,*

$$\lim_{M,N\to\infty} \sup_{z\in D_\epsilon} |m(z) - \bar{m}(z)| = 0 \tag{2.34}$$

almost surely, where

$$m(z) = \mathbf{s}^H\mathbf{R}^{-1/2}(\mathbf{\Phi}\mathbf{\Phi}^H + \delta\mathbf{R}^{-1} - z\mathbf{I}_M)^{-1}\mathbf{R}^{-1/2}\mathbf{s}. \tag{2.35}$$

$$\bar{m}(z) = \sum_{m=1}^{M} \frac{\mathbf{s}^H\mathbf{R}^{-1/2}\mathbf{e}_m\mathbf{e}_m^H\mathbf{R}^{-1/2}\mathbf{s}(1+cb(z))}{1+(\delta\lambda_m^{-1} - z)(1+cb(z))}. \tag{2.36}$$

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 equation,*

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

Proof: The proof is shown in [9, Appendix I] and references therein.

■

At this point, recall that the quantity of interest $\eta_n(\delta) = \mathbf{s}^H \check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$ is related to $m(z)$ through a derivative operation, see equation (2.33). Moreover, observe that $m(z)$ is an holomorphic function and convergence of holomorphic functions implies convergence of all its derivatives (by the Weierstrass theorem). Thereby, from Lemma 2.3 it is easily deduced that $\eta_n(\delta)$ converges in probability towards a function $\bar{\eta}_n(\delta)$. More specifically

$$\lim_{M,N\to\infty} |\eta_n(\delta) - \bar{\eta}_n(\delta)| = 0 \text{ in probability.} \tag{2.37}$$

Where

$$\bar{\eta}_n(\delta) = \frac{d\bar{m}(z)}{dz}|_{z=0} = \sum_{m=1}^{M} \frac{\left|\mathbf{s}^H \mathbf{e}_m\right|^2 \lambda_m((1+cb)^2 + cb')}{(\lambda_m + \delta(1+cb))^2}. \tag{2.38}$$

Note that $b = b(z)|_{z=0}$, $b(z)$ was defined in lemma 2.3 and $b' = \frac{db(z)}{dz}|_{z=0}$ has the next expression,

$$b' = \frac{db(z)}{dz}|_{z=0} = \left[1 - \frac{1}{M}\sum_{m=1}^{M} \frac{c\lambda_m^2}{(\lambda_m + \delta(1+cb))^2}\right]^{-1} \left[\frac{1}{M}\sum_{m=1}^{M} \frac{\lambda_m^2(1+cb)^2}{(\lambda_m + \delta(1+cb))^2}\right]. \tag{2.39}$$

Therefore at this point the first part of the proof, which aimed to find the convergence of $\eta_n(\delta)$ towards a deterministic function of $\mathbf{R}$ is finished. More specifically, equations (2.37) and (2.38) yield the next asymptotic equivalence, where the convergence is in probability,

$$\mathbf{s}^H \check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} = \eta_n(\delta) \asymp \bar{\eta}_n(\delta) = \sum_{m=1}^{M} \frac{\left|\mathbf{s}^H \mathbf{e}_m\right|^2 \lambda_m((1+cb)^2 + cb')}{(\lambda_m + \delta(1+cb))^2}. \tag{2.40}$$

In order to proceed, note that the right hand side of equation (2.40) can be rewritten in terms of $\mathbf{R}$, just taking into account the eigendecomposition of $\mathbf{R}$ and its properties. This leads to rewrite (2.40) as follows,

$$\mathbf{s}^H \check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} \asymp [(1+cb)^2 + cb']\mathbf{s}^H(\mathbf{R} + \delta(1+cb)\mathbf{I})^{-1}\mathbf{R}(\mathbf{R} + \delta(1+cb)\mathbf{I})^{-1}\mathbf{s} \tag{2.41}$$

Now note that (2.41) can be rewritten in a more compact form. To this end first, observe that the term b' in (2.39) can be expressed in terms of a new parameter ξ,

$$
\begin{aligned}
b' &= \frac{(1+cb)^2\xi}{1-c\xi} \\
\xi &= \frac{1}{M}\sum_{m=1}^{M}\left(\frac{\lambda_m}{\lambda_m+\delta(1+cb)}\right)^2 .
\end{aligned}
\tag{2.42}
$$

This permits to rewrite (2.41) as follows,

$$
\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} \asymp \left[(1+cb)^2\left(\frac{1}{1-c\xi}\right)\right]\mathbf{s}^H(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{s} \tag{2.43}
$$

This equation establishes the beginning of the second part of the proof, which consists of obtaining an *(M,N)*-consistent estimate of the right hand side of (2.43) and thereby of the expression of interest herein which is the left hand side of (2.43). More specifically, to achieve this aim one can see that one must estimate the next three expressions,

$$
\begin{aligned}
b &= \frac{1}{M}\sum_{m=1}^{M}\frac{\lambda_m(1+cb)}{\lambda_m+\delta(1+cb)} \\
\xi &= \frac{1}{M}\sum_{m=1}^{M}\left(\frac{\lambda_m}{\lambda_m+\delta(1+cb)}\right)^2 \\
\bar{\eta}_n &= \mathbf{s}^H(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{s}.
\end{aligned}
$$

These estimates are given in [124, Appendix A] and the rationale relies on expressing b, ξ, $\bar{\eta}_n$ in terms of the next functions $t(x)$ and $s(x)$, which are particular cases of a real Stieltjes transform $m_{\varphi_k}(x)$, see e.g. [100, 101]

$$
\begin{aligned}
m_{\varphi_k}(x) &= \sum_{k=1}^{M}\frac{\varphi_k}{1+x\lambda_k},\ x\in\mathbb{R}^+ \\
t(x) &= m_{\varphi_k}(x)\big|_{\varphi_k=|\mathbf{s}^H\mathbf{e}_k|^2} = \sum_{k=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_k\right|^2}{1+x\lambda_k},\ x\in\mathbb{R}^+ \\
s(x) &= m_{\varphi_k}(x)\big|_{\varphi_k=1/M} = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{1+x\lambda_k}\ ,\ x\in\mathbb{R}^+.
\end{aligned}
\tag{2.44}
$$

Thereby it can be shown that b, ξ and $\bar{\eta}_n$ can be expressed as follows, where the relation between b and $s(x)$ is readily obtained by using the matrix inversion lemma, the eigendecomposition of $\mathbf{R}$ and the properties of the trace operator.

$$
\begin{aligned}
\xi &= 1 - s(x)|_{x=(\delta(1+cb))^{-1}} + (\delta(1+cb))^{-1} \frac{ds(x)}{dx}|_{x=(\delta(1+cb))^{-1}} \\
\frac{b}{(1+cb)} &= 1 - s(x)|_{x=(\delta(1+cb))^{-1}} \\
\bar{\eta}_n &= -\left[x^2 \frac{dt(x)}{dx} \right] |_{x=(\delta(1+cb))^{-1}}
\end{aligned}
\tag{2.45}
$$

Therefore, one can see that the estimation of the quantities of interest b, ξ and $\bar{\eta}_n$ requires to obtain the estimation of the Stieltjes transforms $t(x)$ and $s(x)$. The *(M,N)*-consistent estimator of $m_{\varphi_k}(x)$ and thereby of $t(x)$ and $s(x)$ was obtained in e.g. [100, 101]. These are summarized in the next equation,

$$
\begin{aligned}
m_{\varphi_k}(x) \asymp \hat{m}_{\varphi_k}(x) &= \sum_{k=1}^{M} \frac{\varphi_k}{1+\theta(x)\hat{\lambda}_k} \\
t(x) \asymp \hat{t}(x) &= \sum_{k=1}^{M} \frac{\left|\mathbf{s}^H \hat{\mathbf{e}}_k\right|^2}{1+\theta(x)\hat{\lambda}_k} \\
s(x) \asymp \hat{s}(x) &= \frac{1}{M} \sum_{k=1}^{M} \frac{1}{1+\theta(x)\hat{\lambda}_k}.
\end{aligned}
\tag{2.46}
$$

where $\theta(x)$ is the positive solution to the next equation,

$$
\theta(x) \left[1 - c + c \frac{1}{M} \operatorname{Tr}\left[\left(\mathbf{I} + \theta(x) \hat{\mathbf{R}} \right)^{-1} \right] \right] = x, x > 0. \tag{2.47}
$$

As a consequence, considering (2.46) into (2.45) one obtains the desired estimates, i.e. b, ξ and $\bar{\eta}_n$, which are expressed in the next equations. [4]

$$
b \asymp \hat{b} = \frac{1 - \frac{\delta}{M} \operatorname{Tr}[(\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}]}{1 - c\left(1 - \frac{\delta}{M} \operatorname{Tr}[(\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}] \right)} \tag{2.48}
$$

[4] The key to obtain these expressions is to note that in (2.47) $1/M \operatorname{Tr}[(\mathbf{I} + \theta(x)\hat{\mathbf{R}})^{-1}] = \hat{s}(x)$ and to use that $\dfrac{\hat{b}}{(1+c\hat{b})} = 1 - \hat{s}(x)|_{x=(\delta(1+c\hat{b}))^{-1}}$, which leads to obtain that $\theta(x)|_{x=(\delta(1+c\hat{b}))^{-1}} = 1/\delta$.

$$\xi \asymp \hat{\xi} = \frac{\frac{1}{M}\mathrm{Tr}[\hat{\mathbf{R}}^2(\hat{\mathbf{R}}+\delta\mathbf{I})^{-2}] - \frac{c}{M^2}(\mathrm{Tr}[\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}])^2}{1-c+c\delta^2\frac{1}{M}\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-2}]} \tag{2.49}$$

$$\bar{\eta}_n \asymp \hat{\eta}_n = \frac{\left(1-c\left(1-\frac{\delta}{M}\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]\right)\right)^2}{1-c+\frac{c}{M}\mathrm{Tr}[(\delta^{-1}\hat{\mathbf{R}}+\mathbf{I})^{-2}]}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \tag{2.50}$$

In order to finish the proof, it only remains to substitute the quantities b, ξ and $\bar{\eta}_n$ of (2.43) by their estimates obtained in (2.48), (2.49) and (2.50), respectively. After some manipulations, which mainly require the matrix inversion lemma and the identity displayed in the footnote [5], one obtains the *(M,N)*-consistent estimation of the quantity of interest,

$$\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} \asymp \frac{1}{\left(1-\frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}]\right)^2}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s}. \tag{2.51}$$

Where recall that $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta\mathbf{I}$. Equation (2.51) coincides with the identity (2.18) thereby the proof is now concluded.

■

Proof of (2.16)

This proof is a particular case of the proof for (2.18), i.e. it is mainly based on the proof provided in [9, Appendix I], (cf. [10, Chapter4], [32] and [125]). First, let us define the next random quantity,

$$\eta_n(\alpha) \ = \mathbf{s}^H(\hat{\mathbf{R}}+\alpha\mathbf{I}_M)^{-1}\mathbf{R}(\hat{\mathbf{R}}+\alpha\mathbf{I}_M)^{-1}\mathbf{s} \tag{2.52}$$

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

$$\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} = \lim_{\alpha\to 0}\eta_n(\alpha) \tag{2.53}$$

Therefore, in order to proof (2.16), the convergence of η_n must be obtained. But, this was obtained above in (2.37)-(2.38)

[5] $\frac{1}{M}\mathrm{Tr}[\hat{\mathbf{R}}^2(\hat{\mathbf{R}}+\delta\mathbf{I})^{-2}] = 1-\frac{2}{M}\mathrm{Tr}[(\delta^{-1}\hat{\mathbf{R}}+\mathbf{I})^{-1}] + \frac{1}{M}\mathrm{Tr}[(\delta^{-1}\hat{\mathbf{R}}+\mathbf{I})^{-2}]$. [10, p. 248]

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

$$\bar{\eta}_n(\alpha) = \sum_{m=1}^{M} \frac{\left|\mathbf{s}^H \mathbf{e}_m\right|^2 \lambda_m((1+cb)^2 + cb')}{\left(\lambda_m + \alpha(1+cb)\right)^2} \tag{2.54}$$

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

$$b = \frac{1}{M} \sum_{m=1}^{M} \frac{\lambda_m(1+cb)}{\lambda_m + \alpha(1+cb)} \tag{2.55}$$

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

$$b' = \frac{db(z)}{dz}|_{z=0} = \left[1 - \frac{1}{M} \sum_{m=1}^{M} \frac{c\lambda_m^2}{\left(\lambda_m + \alpha(1+cb)\right)^2}\right]^{-1} \left[\frac{1}{M} \sum_{m=1}^{M} \frac{\lambda_m^2(1+cb)^2}{\left(\lambda_m + \alpha(1+cb)\right)^2}\right] \tag{2.56}$$

Now, 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 (2.53) and considering the asymptotic equivalent of $\eta_n(\alpha)$ in (2.54),

$$\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} \asymp \lim_{\alpha \to 0} \bar{\eta}_n(\alpha) \tag{2.57}$$

According to (2.55) and (2.56) $b \xrightarrow{\alpha \to 0} (1-c)^{-1}$ and $b' \xrightarrow{\alpha \to 0} (1-c)^{-3}$, respectively. Therefore, after operating the limit $\alpha \to 0$, we can rewrite (2.57) 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, recalling the eigendecomposition properties of $\mathbf{R}$, we obtain the next asymptotic equivalence, which concludes the proof for (2.16),

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

■

Proof of (2.17) and (2.19)

We will begin by giving the proof of (2.19), as (2.17) is a particular case of it. The proof is based on [9], cf. also [126, Proposition 1]. First, note that according to (2.13), $\mathbf{s}^H\mathbf{R}\mathbf{s}$ can be expressed as,

$$\mathbf{s}^H\mathbf{R}\mathbf{s} = -\left[\frac{d}{dx}\mathbf{s}^H(\mathbf{I}_M + x\mathbf{R})^{-1}\mathbf{s}\right]|_{x=0} \tag{2.58}$$

Now, the key in the proof is to recall the G-25 estimator of the real Stieltjes transform of $\mathbf{R}$ stated in (2.10) and (2.11), which states the next almost sure convergence within the asymptotic regime where $M, N \to \infty$ and $M/N \to c \in (0, \infty)$

$$\mathbf{s}^H(\mathbf{I}_M + x\mathbf{R})^{-1}\mathbf{s} \asymp \mathbf{s}^H(\mathbf{I}_M + \theta(x)\hat{\mathbf{R}})^{-1}\mathbf{s} \tag{2.59}$$

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

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

Now considering the asymptotic equivalence (2.59) in (2.58) we obtain the next relation,

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

Where the second equality follows from the chain rule and the next property. Given a matrix $\mathbf{X}$ which depends on a parameter p, then $d[\text{Tr}(\mathbf{X}^{-1}\mathbf{A})]/dp = -\text{Tr}[(\mathbf{X}^{-1}\mathbf{A}\mathbf{X}^{-1})d\mathbf{X}/dp]$, see [14, p. 1401]. Finally, after easy manipulations 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,

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

The proof of (2.17) is a particular case of the proof of (2.19) shown above. Namely, in the proof of (2.19) the next almost surely convergence within the asymptotic regime $M, N \to \infty$ and $M/N \to c \in (0, \infty)$ was shown, $\mathbf{s}^H\mathbf{R}\mathbf{s} \asymp \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}$. Thereby, (2.17) is shown directly from (2.19), as recall that (2.17) proposed that $\mathbf{s}^H\mathbf{R}\mathbf{s} \asymp \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}$ within the regime $M, N \to \infty$ and $M/N \to c \in (0, 1)$.

■

Proof of (2.20) and (2.21)

The proof of (2.21) can be separated in two parts, as in the proof of (2.18). The first part obtains the convergence of (2.20) and (2.21) towards a deterministic function that only depends on $\mathbf{R}$ and the regularization parameter δ. To do so, it interprets (2.20) and (2.21) as a type of Stieltjes transforms associated to the random matrix $\mathbf{B} = \alpha\mathbf{R}^{-1} + \mathbf{\Phi}\mathbf{\Phi}^H$, where $\mathbf{\Phi}$ is an $M \times N$ random matrix with iid complex entries with zero mean and variance $1/2N$. In the second part, all the unknown quantities of the deterministic function obtained in the first part are estimated. To this end, it is observed that all the quantities can be expressed as a function of the real Stieltjes transform,

$$m_{\varphi_k}(x) = \sum_{k=1}^{M} \frac{\varphi_k}{1 + x\lambda_k}, x \in \mathbb{R}^+$$

whose (M,N)-consistent estimate is known.

In order to begin the first part of the proof, the quantities of interest are denoted as follows for the sake of the clarity.

$$\begin{aligned} \eta_1 &= \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s} \\ \eta_2 &= \mathbf{s}^H \mathbf{R} \check{\mathbf{R}}^{-1} \mathbf{s} \end{aligned} \tag{2.60}$$

with $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta\mathbf{I}$. Next, bear in mind the relation between $\hat{\mathbf{R}}$ and $\mathbf{R}$, already discussed above in (2.31), see also [9, 34, 73].

$$\hat{\mathbf{R}} = \mathbf{R}^{1/2} \mathbf{\Phi}\mathbf{\Phi}^H \mathbf{R}^{1/2}. \tag{2.61}$$

where $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$ is a random matrix with iid entries and whose real and imaginary parts are independent, have zero mean and variance $1/2N$. Taking into account (2.61) into (2.60) one obtains that both η_1 and η_2 can be expressed in terms of the next function

$$h(z) = \mathbf{c}_1^H (\mathbf{\Phi}\mathbf{\Phi}^H + \delta\mathbf{R}^{-1} - z\mathbf{I})^{-1} \mathbf{c}_2. \tag{2.62}$$

Where $\mathbf{c}_1$ and $\mathbf{c}_2$ are two deterministic vectors. More specifically,

$$\eta_1 = \mathbf{s}^H \mathbf{R}^{-1/2} (\mathbf{\Phi}\mathbf{\Phi}^H + \delta\mathbf{R}^{-1} - z\mathbf{I})^{-1} \mathbf{R}^{-1/2} \mathbf{R}\mathbf{s} = h(z)|_{z=0, \mathbf{c}_1^H = \mathbf{s}^H \mathbf{R}^{-1/2}, \mathbf{c}_2 = \mathbf{R}^{-1/2}\mathbf{R}\mathbf{s}}. \tag{2.63}$$

$$\eta_2 = \mathbf{s}^H \mathbf{R}\mathbf{R}^{-1/2}(\mathbf{\Phi}\mathbf{\Phi}^H + \delta\mathbf{R}^{-1} - z\mathbf{I})^{-1}\mathbf{R}^{-1/2}\mathbf{s} = h(z)|_{z=0,\mathbf{c}_1^H=\mathbf{s}^H\mathbf{R}\mathbf{R}^{-1/2},\mathbf{c}_2=\mathbf{R}^{-1/2}\mathbf{s}}. \quad (2.64)$$

Therefore, in order to obtain the convergence of the quantities of interest η_1 and η_2, one must find the convergence of $h(z)$. Note that $h(z)$ has the form of the Stieltjes Transform introduced above in (2.9). Thereby, the convergence of the Stieltjes transform associated to the type of matrices $M = \mathbf{\Phi}\mathbf{\Phi}^H + \delta\mathbf{R}^{-1}$ must be found. This convergence is actually found in [127], [9, Appendix I] and references therein for the case where $\mathbf{c}_1 = \mathbf{c}_2$. Though analyzing the proofs in [127], see also [9, Appendix I], it is easily shown that the results hold for two generic deterministic complex vectors $\mathbf{c}_1, \mathbf{c}_2$ with uniformly bounded norm for all M. Thereby, from [9, Lemma I,Appendix I], one obtains that considering the asymptotic regime where $M, N \to \infty, M/N \to c \in (0, \infty)$ the next almost surely convergence for $h(z)$ holds,

$$h(z) = \mathbf{c}_1^H(\mathbf{\Phi}\mathbf{\Phi}^H + \delta\mathbf{R}^{-1} - z\mathbf{I})^{-1}\mathbf{c}_2 \asymp \bar{h}(z) = \sum_{m=1}^{M} \frac{\mathbf{c}_1^H\mathbf{e}_m\mathbf{e}_m^H\mathbf{c}_2(1+cb(z))}{1+(\delta\lambda_m^{-1}-z)(1+cb(z))}. \quad (2.65)$$

Where, recall that λ_m is the m-th eigenvalue of $\mathbf{R}$ and $\mathbf{e}_m$ the associated eigenvector. Furthermore, $b(z)$ is the positive solution to the next equation

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

Thus, for the particular case of $z = 0$, and bearing in mind the eigendecomposition of $\mathbf{R}$ and its properties, one obtains the next expression

$$\bar{h}(z)|_{z=0} = (1+cb)\mathbf{c}_1^H(\mathbf{I} + \delta(1+cb)\mathbf{R}^{-1})^{-1}\mathbf{c}_2. \quad (2.66)$$

Where $b = b(z)|_{z=0}$. As a consequence, taking into account (2.63)-(2.66) into (2.60) one obtains the next convergence in probability for the quantities of interest η_1, η_2 within the asymptotic regime where $M, N \to \infty, M/N \to c \in (0, \infty)$

$$\begin{aligned} \eta_1 &= \mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \asymp (1+cb)\mathbf{s}^H(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{R}\mathbf{s}, \\ \eta_2 &= \mathbf{s}^H\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} \asymp (1+cb)\mathbf{s}^H\mathbf{R}(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{s}. \end{aligned} \quad (2.67)$$

Where recall that $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta\mathbf{I}$. Therefore, the first part of the proof is concluded. The second part of the proof aims to find the (M,N)-consistent estimate of the unknown terms

in (2.67). To this end, first bear in mind the property $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ of two generic invertible complex matrices $\mathbf{A}$ and $\mathbf{B}$. Then, it is easy to show that

$$\begin{array}{l} \mathbf{s}^H(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{R}\mathbf{s} = \mathbf{s}^H(\mathbf{I}+\delta(1+cb)\mathbf{R}^{-1})^{-1}\mathbf{s} \\ \mathbf{s}^H\mathbf{R}(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{s} = \mathbf{s}^H(\mathbf{I}+\delta(1+cb)\mathbf{R}^{-1})^{-1}\mathbf{s}. \end{array} \tag{2.68}$$

Thereby, it is enough to estimate the next unknown terms in the right hand side of (2.67)

$$\begin{array}{c} b \\ \mathbf{s}^H\mathbf{R}(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{s}. \end{array} \tag{2.69}$$

The rationale to obtain the (M,N)-consistent estimators of these quantities relies on a twofold approach. First, the quantities in (2.69) are expressed in terms of the real Stieltjes transforms $t(x)$ and $s(x)$, defined above in (2.44). Then, the (M,N)-consistent estimators of $t(x)$ and $s(x)$ obtained by Girko in [100, 101], and summarized above in (2.46), are used. This approach lead to obtain the estimate for b in [124, Appendix A], recall that this procedure was exposed above in equations (2.45)-(2.48). Namely, recall that according to (2.48),

$$b \asymp \hat{b} = \frac{1-\dfrac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]}{1-c\left(1-\dfrac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]\right)} \tag{2.70}$$

It remains to obtain the (M,N)-consistent estimator of $\mathbf{s}^H\mathbf{R}(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{s}$. To this end, first note that applying the matrix inversion lemma and the property $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ of two generic invertible complex matrices $\mathbf{A}$ and $\mathbf{B}$, one can write $\mathbf{s}^H\mathbf{R}(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{s}$ as follows,

$$\begin{aligned} \mathbf{s}^H\mathbf{R}(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{s} &= \mathbf{s}^H\mathbf{R}\left(\mathbf{R}^{-1}-\mathbf{R}^{-1}\delta(1+cb)\left(\mathbf{I}+\mathbf{R}^{-1}\delta(1+cb)\right)^{-1}\mathbf{R}^{-1}\right)\mathbf{s} \\ &= \mathbf{s}^H\mathbf{s}-\mathbf{s}^H((\delta(1+cb))^{-1}\mathbf{R}+\mathbf{I})^{-1}\mathbf{s} \end{aligned} \tag{2.71}$$

At this point, recall the expression of the real Stieltjes transform $t(x)$ in e.g. (2.46).

$$t(x) = \mathbf{s}^H(\mathbf{I} + x\mathbf{R})^{-1}\mathbf{s}, x \in \mathbb{R}^+.$$

Thereby, it is clear that the expression of interest (2.71) can be rewritten in terms of the real Stieltjes transform $t(x)$ as follows,

$$\mathbf{s}^H\mathbf{R}(\mathbf{R} + \delta(1+cb)\mathbf{I})^{-1}\mathbf{s} = \mathbf{s}^H\mathbf{s} - t(x)|_{x=(\delta(1+cb))^{-1}}. \tag{2.72}$$

As a consequence, in order to obtain the (M,N)-consistent estimator of $\mathbf{s}^H\mathbf{R}(\mathbf{R}+\delta(1+cb)\mathbf{I})^{-1}\mathbf{s}$, one must obtain the (M,N)-consistent estimator of $t(x)$, which is denoted as $\hat{t}(x)$. This estimator was obtained by Girko in e.g. [100, 101], its expression is given above in (2.46). Considering $\hat{t}(x)$ in (2.72) the next almost surely convergence within the asymptotic regime where $M, N \to \infty$ and $M/N \to c \in (0, \infty)$ is obtained

$$\begin{aligned}\mathbf{s}^H\mathbf{R}(\mathbf{R} + \delta(1+cb)\mathbf{I})^{-1}\mathbf{s} &\asymp \mathbf{s}^H\mathbf{s} - \hat{t}(x)|_{x=(\delta(1+cb))^{-1}} \\ &= \mathbf{s}^H\mathbf{s} - \mathbf{s}^H(\mathbf{I} + \theta(x)\hat{\mathbf{R}})^{-1}\mathbf{s}|_{x=(\delta(1+c\hat{b}))^{-1}}.\end{aligned} \tag{2.73}$$

Where $\theta(x)$ is the unique positive solution to

$$\theta(x)\left[1 - c + c\frac{1}{M}\mathrm{Tr}\left[(\mathbf{I} + \theta(x)\hat{\mathbf{R}})^{-1}\right]\right] = x. \tag{2.74}$$

Next, the expression for $\theta(x)|_{x=(\delta(1+c\hat{b}))^{-1}}$ must be found. To this end, observe that according to (2.45) and (2.46)

$$\begin{aligned}\hat{s}(x) &= \tfrac{1}{M}\mathrm{Tr}\left[(\mathbf{I} + \theta(x)\hat{\mathbf{R}})^{-1}\right] \\ \frac{\hat{b}}{(1+c\hat{b})} &= 1 - \hat{s}(x)|_{x=(\delta(1+c\hat{b}))^{-1}}.\end{aligned} \tag{2.75}$$

Therefore, taking into account (2.75) into (2.74) one arrives at the conclusion that $\theta(x)|_{x=(\delta(1+c\hat{b}))^{-1}} = 1/\delta$. Considering this result, the expression (2.73) can be rewritten as follows,

$$\begin{aligned}\mathbf{s}^H\mathbf{R}(\mathbf{R} + \delta(1+cb)\mathbf{I})^{-1}\mathbf{s} &\asymp \mathbf{s}^H\mathbf{s} - \mathbf{s}^H(\mathbf{I} + \frac{1}{\delta}\hat{\mathbf{R}})^{-1}\mathbf{s} \\ &= \mathbf{s}^H\hat{\mathbf{R}}(\hat{\mathbf{R}} + \delta\mathbf{I})^{-1}\mathbf{s}.\end{aligned} \tag{2.76}$$

Where the second equality follows after bearing in mind the matrix inversion lemma and after easy manipulations. Therefore, taking into account (2.76), (2.70) and (2.68) in (2.67) one obtains the next (M,N)-consistent estimators, which concludes the proof.

$$\begin{aligned}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{R}\mathbf{s} &\asymp \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}\mathbf{s}\\ \mathbf{s}^H\mathbf{R}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} &\asymp \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]}\mathbf{s}^H\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}.\end{aligned} \tag{2.77}$$

■

Chapter 3

Shrinkage of the sample LMMSE to tackle the finite sample size effect.

3.1 Introduction

This chapter deals with shrinkage corrections of the sample LMMSE filter. They permit to obtain robust methods to the small sample size regime and maintain the optimal properties of the sample LMMSE for a large sample size situation. First, in section 3.2 the most basic shrinkage correction of the sample LMMSE is considered, i.e. a linear transformation of the sample LMMSE of the type $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$ is considered. Where the aim of α is to reduce the MSE induced by the sample LMMSE filter in the estimation of the parameter of interest $x(n)$ by means of optimizing the bias variance tradeoff. That is, a bias in the estimation is allowed with the aim of reducing the overall MSE. Thereby, α is designed to optimize the MSE, which leads to an optimal expression depending on the unknown $\mathbf{R}$. To circumvent this problem random matrix theory tools are used to obtain an asymptotically optimal estimator, i.e. which tends to the optimal estimator in the asymptotic regime where $M, N \to \infty$ and $M/N \in (0, 1)$. This regime permits to deal naturally with the small sample size situation. Next, in section 3.3 the type of shrinkage estimator $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$ is considered as well. However, an alternative approach to the method described in section 3.2 is proposed to obtain the shrinkage factor α. Instead of a direct minimization of the MSE, it is suggested 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. In section 3.4 the previous filters are extended to support the cases where $M \geqslant N$ and to improve in general their performance for any sample size. To this end a regularization of the SCM is considered, i.e. the filter has the expression $\mathbf{w} = \alpha \check{\mathbf{R}}^{-1}\mathbf{s}$, with $\check{\mathbf{R}} = \beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I}$. The scalar α controlling the shrinkage of the filter is designed as the one which minimizes the MSE of the signal of interest for any regularization of the SCM, i.e. for any β_1, β_2. And as the optimal α depends on the unknown $\mathbf{R}$, then random matrix theory results are used to obtain an estimation that tends to the optimal α in the asymptotic regime where $M, N \to \infty$ and $M/N \in (0, \infty)$. Moreover, the scalars β_1, β_2 governing the shrinkage of the SCM are the ones proposed by Ledoit and Wolf in [33], i.e. the ones minimizing the asymptotic MSE in the estimation of the data covariance. Next, in section 3.5, a more general version of the previous shrinkage filter is considered, namely the next shrinkage correction of the sample LMMSE, $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, is considered. This is a linear combination of the sample LMMSE and a matched filter or conventional beamformer. That is, it combines the available information of the observations in the sample LMMSE with a kind of prior information represented by the matched filter, as this conventional beamformer is obtained by estimating the unknown noise covariance by just the identity matrix. Moreover, on the one hand the sample LMMSE has better rejection against the interference than the matched filter and on the other hand, the conventional beamformer is not affected by a small sample size situation, i.e. the signal cancelation effect in the sample LMMSE. Thereby, the shrinkage coefficients α_1, α_2 try to obtain the benefits of both filters and they are designed to optimize a bias variance tradeoff by means of the optimization of the MSE in the estimation of the signal of interest. As in the previous shrinkage filters this leads to expressions of α_1, α_2 which depend on the unknown data covariance $\mathbf{R}$. To circumvent this problem a random matrix theory approach is proposed. This paves the way to obtain an $(M, N)-$consistent estimator of the optimal shrinkage filter for $M/N \in (0, 1)$, or in other words a method that is asymptotically optimal in the regime where $M, N \to \infty$ and $M/N \in (0, 1)$. In section 3.6, the previous shrinkage filter of section 3.5 is extended to support the situations where $M \geqslant N$ and to improve in general its performance thanks to a regularization of the SCM, thereby the proposed filter has the structure $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, with $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I}$. This is the more complete form of shrinkage proposed in this chapter, in the sense that all the shrinkage filters presented in the previous sections are a particular case of it. The shrinkage factors α_1, α_2 are designed to minimize the MSE of the signal of interest for a given δ. Then, after inserting the optimal α_1, α_2 in the MSE, the optimal δ is obtained as the argument which minimizes the MSE. Those optimal α_1, α_2 and δ depend on the unknown covariance $\mathbf{R}$. Thereby, random matrix theory results are used to obtain estimators that tend to the optimal α_1, α_2 and δ in the regime where $M, N \to \infty$ and

$M/N \in (0, \infty)$, which deals explicitly with the small sample size regime. Section 3.7 deals with the numerical simulations. They highlight that the most simple forms of shrinkage proposed in sections 3.2 and 3.3 outperform the sample LMMSE in terms of MSE provided that $M/N \in (0, 1)$. Moreover, the shrinkage of a regularized LMMSE in 3.4 outperforms other robust methods to the small sample size regime such as the LW and the ad-hoc DL methods explained in section 1.4.1 for any sample size regime in terms of MSE. The more general form of shrinkage in 3.5 outperforms the LW and the ad-hoc DL methods both in terms of SINR and MSE provided that $M/N \in (0, 1)$. Moreover, it outperforms the asymptotically SINR optimal DL method proposed in [9] in terms of MSE and it gives almost the same performance in SINR, provided that $M/N \in (0, 1)$. Finally the most complete form of shrinkage proposed in 3.6 outperforms the asymptotically SINR optimal DL method proposed in [9] in terms of MSE and it gives similar performance in terms of SINR for any sample size regime. The method in 3.6 also gives some performance gains in terms of MSE compared to the asymptotically MSE optimal DL method in [79]. Even more important in chapter 5 it will be shown that when there is an uncertainty in the signature vector of the SOI, the proposed shrinkage filters in 3.5 and 3.6 outperform [9] and [79] in terms of SINR. Finally, in the appendix the proofs of the lemmas and theorems of this chapter are presented.

3.2 Shrinkage of the sample LMMSE

Next we study the shrinkage of the sample LMMSE relying on the class of filters expressed as $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$. This type of shrinkage estimators introduce a shrinkage coefficient α. This is a correction or calibration term whose aim is to counteract the degradation of the sample LMMSE in the small sample size regime and keep its optimal properties in large sample size situations. Indeed note that these type of methods have the structure of a basic linear shrinkage estimator which are based on a scaling of a sample estimator. In our case the sample estimator is the sample LMMSE due to the presence of the SCM, i.e. $\hat{x}(n) = \mathbf{w}^H\mathbf{y}(n) = (\hat{\mathbf{R}}^{-1}\mathbf{s})^H\mathbf{y}(n)$. Thereby, the role of α is to modify the bias and variance in such a way that the MSE in the estimation of the SOI is reduced, compared to the sample LMMSE.

Next, the shrinkage factor α is designed as the one which minimizes the MSE in the estimation of the signal of interest. This is formally stated in the next lemma.

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, 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. $\mathbf{w}_{l,s}{=}\alpha_l \hat{\mathbf{R}}^{-1}\mathbf{s}$. This problem is mathematically formulated as follows,

$$\hat{x}_{l,s}(n) = \mathbf{w}_{l,s}^H \mathbf{y}(n); \qquad \mathbf{w}_{l,s} = \arg\min_{\mathbf{w}} \mathbb{E}\left[\left|x(n) - \mathbf{w}^H \mathbf{y}(n)\right|^2 \mid \hat{\mathbf{R}}\right] \quad s.t. \mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s} \tag{3.1}$$

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

$$\alpha_l = \frac{\gamma \mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}} \tag{3.2}$$

Proof: See section 3.8.

■

In fact, α has different interpretations that are explained next. First, α can be interpreted as an automatic gain control, which is needed in any MIMO communication system according to [76], note that the system model in (1.1) corresponds to the particular case of a SIMO system. Next, another interpretation is given for α. Note that $\hat{x}$ is a function of $\hat{\mathbf{R}}$, which on its turn is an estimation of $\mathbf{R}$. Thereby, the estimation errors of $\hat{\mathbf{R}}$ are translated in $\hat{x}$. Therefore, the role of α is to control how sensitive is $\hat{x}$ due to the errors provoked by $\hat{\mathbf{R}}$ or in other words how much errors in the output result from errors in the input.

Moreover, the type of shrinkage filters $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$ in (3.1), which use the MSE as a design criterion, are important in applications where the complex amplitude of the signal of interest is important. This is the case of subband beamforming, see e.g. [17]. In this kind of applications a wideband signal is decomposed, at each antenna of the array, into a number of frequency bands and decimated by means of a filter bank. Thereby, the n-th snapshot of the signal at the k-th subband associated to the m-th antenna is a narrowband signal denoted by y_m^k. Then, stacking the signals associated to each antenna at the k-th subband, one would have the vector $\mathbf{y}^k(n) = [y_1^k(n), \ldots, y_M^k(n)]^T$. After that one can apply a narrowband beamformer, such as the one in lemma 3.1, to the signal $\mathbf{y}^k(n) = [y_1^k(n), \ldots, y_M^k(n)]^T$ associated to the k-th subband. The output of this beamforming is upsampled and interpolated, which yields the estimation of the signal of interest at the k-th subband. Finally, the estimation of the signal of interest is obtained as the summation of the estimations at each subband. This highlights that a proper estimation of the magnitude of the signal of interest at each subband is fundamental to obtain a good estimation of

the signal. This is accomplished thanks to the scaling α of the proposed shrinkage filter $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$ and the design based on minimizing the MSE.

Interestingly enough (3.2) highlights that the sample LMMSE is not in general an optimal estimator in the MSE sense. Indeed, it is obtained when substituting the unknown $\mathbf{R}$ for the SCM in the optimal shrinkage factor (3.2) and as a consequence is only optimal in the large sample size regime. This is because the SCM tends to the theoretical covariance when $N \to \infty$ for a fixed M, but obviously this is not the case in a small sample size situation. In fact, this highlights the necessity of the proposed shrinkage correction herein. In order to proceed, consider the asymptotic regime $M, N \to \infty$, $M/N \to c \in (0, 1)$, which has appealing properties. It permits to study the convergence of the optimal shrinkage factor in (3.2) within an asymptotic regime which deals naturally with the small sample size regime. Moreover, it paves the way to obtain an asymptotically optimal estimation of α_l in (3.2) by means of RMT tools. By means of the next theorem we present the consistent estimate for the optimal shrinkage of the sample LMMSE method in (3.1). That is, for the class of filters $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$.

Theorem 3.1 *A realizable and consistent estimate of the optimal shrinkage of the sample LMMSE filter (3.1), within the general asymptotics where $M, N \to \infty$, $M/N \to c \in (0, 1)$, reads as follows,*

$$\begin{array}{c} \check{x}_{l,s}(n) = \check{\mathbf{w}}_{l,s}^H \mathbf{y}(n); \check{\mathbf{w}}_{l,s} = \check{\alpha}_l \hat{\mathbf{R}}^{-1}\mathbf{s} \\ \check{\alpha}_l = \gamma (1-c)^2 \end{array} \tag{3.3}$$

Proof: See section 3.8.

■

The proposed estimator in Theorem 3.1 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 it 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 mentioning that in the large sample size regime the performance of the proposed method tends to the one of the sample LMMSE, which in turn tends to the one of the LMMSE.

3.3 Shrinkage of the sample LMMSE for Gaussian distributed data.

Next, as in the previous section, the type of shrinkage filter $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$ is considered to counteract the degradation of the sample LMMSE method in the small sample size regime. However, instead of a design of α based on RMT a different approach is presented in this section. First of all, let us begin this section by recalling that the MSE in the estimation of $x(m)$, when considering the linear model in (1.1) for the observed signal $\mathbf{y}(m)$ and a linear estimator $\hat{x}(m) = \mathbf{w}^H\mathbf{y}(m)$ of $x(m)$, reads as follows for the generic filter $\mathbf{w}$,

$$\text{MSE}\left(\mathbf{w}\right) \triangleq \mathbb{E}\left[\left|x(m) - \mathbf{w}^H\mathbf{y}(m)\right|^2\right] = \mathbf{w}^H\mathbf{R}\mathbf{w} + \gamma(1 - \mathbf{w}^H\mathbf{s} - \mathbf{s}^H\mathbf{w}) \tag{3.4}$$

Now, consider the shrinkage filter $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$. Note that $\hat{\mathbf{R}}$ relies implicitly on the availability of N samples of the random process $\mathbf{y}(n)$. Or in other words, note that by definition $\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$**,** where $\mathbf{Y}$ stacks in its columns the vectors $\mathbf{y}(n)$, $n = 1, \ldots, N$. Thereby, $\mathbf{Y}$ and $\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ are both random matrices. Observe that the index of the random process $\mathbf{y}(n)$ used to build $\hat{\mathbf{R}}$ is denoted n, whereas in the estimation of $\hat{x}(m)$ in (3.4) it is denoted by m. This is for the sake of the clarity and can be intrepreted as follows. In a training period of N samples $\hat{\mathbf{R}}$ is built from $\mathbf{y}(n)$, $n = 1, \ldots, N$, and in the evaluation period the signal of interest $\hat{x}(m)$ is estimated from $\mathbf{y}(m)$, by means of $\hat{x}(m) = \mathbf{w}^H\mathbf{y}(m)$. Thereby, if one considers a realization of the random matrix $\mathbf{Y}$ and constructs the filter $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$, then one can subsitute this filter in the expression of the MSE to obtain the performance associated to the filter $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}$ for a given realization of the underlying random matrix $\mathbf{Y}$. Thereby, actually the MSE is in this case a conditional expectation, i.e. the expectation of the squared error conditioned to a given realization of the random matrix $\mathbf{Y}$ or analogously to a given realization of $\hat{\mathbf{R}}$

$$\text{MSE}\left(\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1}\mathbf{s}\right) \triangleq \mathbb{E}\left[\left|x(m) - \mathbf{w}^H\mathbf{y}(m)\right|^2 \mid \hat{\mathbf{R}}\right] \tag{3.5}$$

Thus, by taking the expectation of (3.5) over $\mathbf{Y}\mathbf{Y}^H$, or in other words over all the possible realizations of the underlying random matrix, we obtain a statistical average of the outcomes of the MSE, which arise from the values of the support of the random matrix $\hat{\mathbf{R}}$. That is, an average MSE is obtained. Therefore, next statement proposes 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 samples of the random process $\mathbf{y}(n)$, i.e. $\{\mathbf{y}(n)\}_{n=1}^{N}$, is available to build $\hat{\mathbf{R}}$. Moreover, assume that the samples are modeled according to (1.1) when assumptions (a)-(f) hold, i.e. they are i.i.d gaussian distributed. Then, obtain an estimate of $x(m)$ in (1.1) by solving the next optimization problem,

$$\begin{aligned}
&\hat{x}_{ls,s}(m) = \hat{\mathbf{w}}_{ls,s}^{H}\mathbf{y}(m) \\
&\hat{\mathbf{w}}_{ls,s} = \arg\min_{\mathbf{w}} \mathbb{E}_{\mathbf{Y}\mathbf{Y}^{H}}\left[\mathbb{E}_{x,\mathbf{n}}\left[\left|x(m) - \mathbf{w}^{H}\mathbf{y}(m)\right|^{2} \mid \hat{\mathbf{R}}\right]\right] \\
&s.t.\ \mathbf{w} = \hat{\alpha}_{ls,s}\hat{\mathbf{R}}^{-1}\mathbf{s}
\end{aligned} \tag{3.6}$$

Observe that this problem formulation is a particular case of the general problem posed in (1.29) which aims to summarize all the problems discussed in this thesis. Namely, the functional $f(\cdot)$ corresponds to the expectation operator $\mathbb{E}[\cdot]$, which is taken over $\hat{\mathbf{R}}$.

The solution to the problem stated in (3.6) is exposed by means of the next theorem. Then the proof of this theorem and the corresponding comments are exposed.

Theorem 3.2 *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 $x(m)$ 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 (3.6), reads as follows,*

$$\begin{aligned}
\hat{x}_{ls,s}(m) = \hat{\mathbf{w}}_{ls,s}^{H}\mathbf{y}(m); \hat{\mathbf{w}}_{ls,s} = \hat{\alpha}_{ls,s}\hat{\mathbf{R}}^{-1}\mathbf{s} \\
\hat{\alpha}_{ls,s} = \gamma((1-c_f)^{2} - \tfrac{1}{N^{2}})
\end{aligned} \tag{3.7}$$

Proof: The proof is provided below in this section.

■

Remark: As the estimator proposed in section 3.2 in (3.3), the method proposed herein in (3.7) 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 (3.3) 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 (3.2). On the other hand, the method proposed in this section in (3.7),

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 (3.3), based on RMT, does not require any assumption about the type of probability distribution of the observations.

Proof of Theorem 3.2

Next, the proof that the proposed shrinkage LMMSE method in (3.7) solves the problem stated in (3.6) is provided. To this end, observe that (3.6) may be rewritten as follows, after introducing the signal model for $\mathbf{y}(m)$ in (1.1) into (3.6), after solving the inner conditional expectation, which is operated upon the joint pdf of $x(m)$ and $\mathbf{n}(m)$ and bearing in mind the model of $\mathbf{R}$ in assumption *(b)* of (1.1). Moreover, obviously the type of filter $\mathbf{w} = \alpha_s \hat{\mathbf{R}}^{-1}\mathbf{s}$ must be considered.

$$\hat{\alpha}_{ls,s} = \arg\min_{\alpha_s} |\alpha_s|^2 \, \mathbb{E}\left[\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right] + \gamma - \gamma(\alpha_s^* + \alpha_s)\mathbb{E}\left[\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}\right] \tag{3.8}$$

Where with some abuse of notation, the subindex in the expectation, indicating that it is operated upon $\mathbf{Y}\mathbf{Y}^H$ has been dropped. The solution to (3.8) is found after setting the first derivative of the cost function equal to zero and after straigthforward manipulations,

$$\hat{\alpha}_{ls,s} = \frac{\gamma \mathbb{E}\left[\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}\right]}{\mathbb{E}\left[\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right]} \tag{3.9}$$

Therefore, in order to obtain the proposed shrinkage LMMSE estimator (3.6), 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 $\hat{\mathbf{R}} = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ 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,

$$\hat{\mathbf{R}}^{-1} \overset{d}{=} N\mathbf{R}^{-1/2}(\mathbf{X}\mathbf{X}^H)^{-1}\mathbf{R}^{-1/2} \tag{3.10}$$

Substituting (3.10) into (3.9) the next equalities can be readily verified,

$$\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right] = N\mathbf{s}^H\mathbf{R}^{-1/2}\mathbb{E}\left[(\mathbf{X}\mathbf{X}^H)^{-1}\right]\mathbf{R}^{-1/2}\mathbf{s} \tag{3.11}$$

$$\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right] = N^2\mathbf{s}^H\mathbf{R}^{-1/2}\mathbb{E}\left[(\mathbf{X}\mathbf{X}^H)^{-1}(\mathbf{X}\mathbf{X}^H)^{-1}\right]\mathbf{R}^{-1/2}\mathbf{s} \tag{3.12}$$

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 [128], 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 (3.9) the first and second moments of a complex inverse Wishart distribution must be found. Namely, in (3.11) the first moment is needed, which according to [128, 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. That is, $\mathbb{E}[\mathbf{\Omega}^{-1}] = 1/(N-M)\mathbf{I}_M$, which substituted in (3.11) yields,

$$\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\right] = \frac{N}{N-M}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} \tag{3.13}$$

In order to obtain the proposed estimator it remains to obtain an expression for (3.12), 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 ([128, eq. 41]) reads component-wise as follows,

$$\mathbb{E}[[\mathbf{\Omega}^{-1}]_{i,j}[\mathbf{\Omega}^{-1}]_{l,k}] = \frac{[\mathbf{I}_M]_{i,j}\,[\mathbf{I}_M]_{l,k} + \frac{1}{N-M}\,[\mathbf{I}_M]_{l,j}\,[\mathbf{I}_M]_{i,k}}{(N-M)^2-1} \tag{3.14}$$

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 (3.14),

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

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

$$\sum_{i=1}^{M}\mathbb{E}[[\mathbf{\Omega}^{-1}]_{p,i}[\mathbf{\Omega}^{-1}]_{i,p}] = \frac{N}{(N-M)((N-M)^2-1)} \tag{3.15}$$

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 (3.14) we can conclude that $\forall i,k \in \{1,\dots,M\}$ with $i \neq k$ the next equality holds,

$$\sum_{j=1}^{M} \mathbb{E}[[\mathbf{\Omega}^{-1}]_{i,j}[\mathbf{\Omega}^{-1}]_{j,k}] = 0 \tag{3.16}$$

As a consequence, considering (3.15) and (3.16), $\mathbb{E}[\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}]$ is given by,

$$\mathbb{E}[\mathbf{\Omega}^{-1}\mathbf{\Omega}^{-1}] = \frac{N}{(N-M)((N-M)^2-1)}\mathbf{I}_M \tag{3.17}$$

Recalling that $\mathbf{\Omega} \triangleq \mathbf{X}\mathbf{X}^H$ and inserting (3.17) in (3.12) we obtain the desired expression for the denominator of the optimal shrinkage factor (3.9),

$$\mathbb{E}\left[\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\right] = \frac{N^3}{(N-M)((N-M)^2-1)}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s} \tag{3.18}$$

Finally, substituting (3.13) and (3.18) in the expression of the optimal shrinkage factor in (3.9) and after straightforward manipulations we obtain the optimal shrinkage LMMSE estimator for our problem statement in (3.6).

$$\begin{aligned} \hat{x}_{l,s}(m) &= \mathbf{w}_{l,s}^H\mathbf{y}(m) \\ \mathbf{w}_{l,s} &= \gamma\left((1-\tfrac{M}{N})^2 - \tfrac{1}{N^2}\right)\hat{\mathbf{R}}^{-1}\mathbf{s} \end{aligned} \tag{3.19}$$

This concludes the proof as (3.19) coincides with the proposed shrinkage estimator in (3.7).

■

3.4 Shrinkage of the regularized sample LMMSE

The shrinkage methods proposed in last sections assume that the observation dimension M is smaller than the sample size N, i.e $M/N \in (0,1)$. The aim of this section is to propose a shrinkage method which may deal with any sample size regime, i.e. $M/N \in (0,\infty)$. Thus we extend the proposed shrinkage framework to cases where $M \geq N$. As an example, such situations may happen for instance in array processing in applications with large arrays,

where there may be up to hundreds or thousands of sensors, e.g. in radioastronomy or over-the-horizon radar [9]. Also it may happen due to a short sample size N produced by a short stationarity of the signal of interest. Regarding the methods that involve inverting the covariance such as the LMMSE, or the MVDR in the next chapter, the conventional approach to deal with $M \geq N$ has been to regularize or shrink the sample covariance matrix, see section 1.4. That is, one substitutes the unknown covariance $\mathbf{R}$ for the next linear combination or shrinkage of the SCM, $\check{\mathbf{R}} = \beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I}$, which permits to invert $\check{\mathbf{R}}$ even when $M \geq N$. This leads to the next type of regularized sample LMMSE,

$$\mathbf{w} = (\beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I})^{-1} \mathbf{s}.$$

For instance, DL techniques consider $\beta_1 = 1$ and focus on the design of β_2 under different criteria, whereas shrinkage techniques are more general as they consider the design of both β_1 and β_2, see section 1.4. On the other hand, recall that the shrinkage filters proposed in the last sections have the form $\mathbf{w} = \alpha \hat{\mathbf{R}}^{-1} \mathbf{s}$ and their rationale is as follows. A correction factor α is introduced to diminish the MSE, achieved by the sample methods such as the sample LMMSE, in the estimation of the parameter of interest. In other words α faces directly the estimation of the parameter of interest and controls the bias variance tradeoff to diminish the MSE respect to the one achieved by the sample LMMSE.

Therefore, one may think of the next double shrinkage or shrinkage of the regularized sample LMMSE. The shrinkage of the SCM $\check{\mathbf{R}} = \beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I}$ is considered in the LMMSE to deal with the cases where $M \geq N$ and because it is a better estimate than the SCM, thanks to its shrinkage structure. Moreover, we explore the possibility to improve further the estimators of the type $\mathbf{w} = (\beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I})^{-1} \mathbf{s}$ by shrinking the filter, i.e. by introducing a correction on this type of regularized methods which reduces further the MSE by controlling the bias variace tradeoff. Therefore, next we deal with the type of filters which are based on a shrinkage of a regularized sample LMMSE

$$\mathbf{w} = \alpha (\beta_1 \hat{\mathbf{R}} + \beta_2 \mathbf{I})^{-1} \mathbf{s}. \tag{3.20}$$

Next, the design of the shrinkage coefficients α, β_1 and β_2 of the proposed filter in (3.20) is dealt with. Namely, first the optimal α is obtained for a given value of β_1 and β_2 by means of the minimization of the MSE. The result is presented in the next lemma.

Lemma 3.2 *Consider a set of observations $\{\mathbf{y}(n)\}_{n=1}^{N}$ fulfilling the model in (1.1) with assumptions (a)-(e). Given $\{\mathbf{y}(n)\}_{n=1}^{N}$, consider the problem of estimating the parameter of interest $x(n)$ in (1.1), based on minimizing the MSE, when the estimator $\hat{x}_{ds}(n) = \mathbf{w}_{ds}^H \mathbf{y}(n)$*

is a shrinkage of the regularized sample LMMSE, i.e. $\mathbf{w}_{ds}{=}\alpha\check{\mathbf{R}}^{-1}\mathbf{s}$ *with* $\check{\mathbf{R}}{=}\beta_1\hat{\mathbf{R}}{+}\beta_2\mathbf{I}$. *Then the shrinkage* α *which minimizes the MSE* $\hat{x}_{ds}(n)$ *for given* β_1, β_2 *has the next expression,*

$$\alpha_o = \arg\min_{\alpha} \mathbb{E}\left[\left|x(n) - \mathbf{w}^H\mathbf{y}(n)\right|^2 \mid \hat{\mathbf{R}}, \beta_1, \beta_2\right] = \gamma\frac{\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}} \tag{3.21}$$

Proof: The proof follows easily from the proof of lemma 3.1 by considering $\check{\mathbf{R}}$ *instead of* $\hat{\mathbf{R}}$.

This highlights that the shrinkage of the filter is not superfluous, i.e. $\alpha_o \neq 1$ and thereby does not lead to the case of just considering $\mathbf{w}_{ds}{=}(\beta_1\hat{\mathbf{R}}{+}\beta_2\mathbf{I})^{-1}\mathbf{s}$. In other words, one may further reduce the MSE of an estimate relying on a regularized sample LMMSE filter $\mathbf{w}_{ds}{=}(\beta_1\hat{\mathbf{R}}{+}\beta_2\mathbf{I})^{-1}\mathbf{s}$. Nonetheless, the optimal α_o in (3.21) leads to an unrealizable filter, as it depends on the unknown $\mathbf{R}$. To circumvent this problem, an (M,N)-consistent estimate of (3.21), denoted by $\hat{\alpha}_o$ will be obtained by means of RMT and G-estimation tools, i.e. $\hat{\alpha}_o$ will converge in probability towards α_o when $M, N \to \infty$ and $M/N \to c \in (0,\infty)$, in compact notation $\hat{\alpha}_o \asymp \alpha_o$. Thereby $\hat{\alpha}_o$ will minimize the asymptotic MSE for any given β_1, β_2. Note that this general asymptotic regime naturally deals with small sample size situations. Moreover, the asymptotic regime $M/N \to c \in (0,\infty)$ highlights the claim stated above where it was said that the proposed method of this section generalizes the shrinkage methods of last section, which are restricted to $M/N \to c \in (0,1)$. To achieve our aim, i.e. the (M,N)-consistent estimate of (3.21), the procedure is as follows,

1. Find the asymptotic deterministic expressions of $\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$, denoted by $f(\mathbf{R},\beta_1,\beta_2)$ and $g(\mathbf{R},\beta_1,\beta_2)$, respectively.

2. Obtain (M,N)-consistent estimates of $f(\mathbf{R},\beta_1,\beta_2)$ and $g(\mathbf{R},\beta_1,\beta_2)$, denoted by $\hat{f}(\hat{\mathbf{R}},\beta_1,\beta_2)$ and $\hat{g}(\hat{\mathbf{R}},\beta_1,\beta_2)$.

3. Estimate $\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$ using $\hat{f}(\hat{\mathbf{R}},\beta_1,\beta_2)$ and $\hat{g}(\hat{\mathbf{R}},\beta_1,\beta_2)$, respectively, i.e. $\hat{\alpha}_o = \gamma\frac{\hat{f}(\hat{\mathbf{R}},\beta_1,\beta_2)}{\hat{g}(\hat{\mathbf{R}},\beta_1,\beta_2)} \asymp \alpha_o$.

Therefore, in order to achieve our aim we first found the asymptotic deterministic expressions of $\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}$, which are given in the next lemma.

Lemma 3.3 *Let consider the general asympotic regime where $M, N \to \infty$ and $M/N \to c \in (0, \infty)$. Then α_o in (3.21) converges in probability towards the next deterministic expression,*

$$\alpha_o \asymp \gamma\beta_1 \frac{1 - c\xi}{1 + cb} \frac{\mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}}{\mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}} \tag{3.22}$$

where $\xi = \frac{1}{M}\sum_{i=1}^{M} \frac{\lambda_i^2}{(\lambda_i+\rho)^2}$, λ_i are the eigenvalues of $\mathbf{R}$, $\rho = \delta(1+cb)$, $\delta \triangleq \beta_2/\beta_1$ and b is the positive solution to the next equation $b = \frac{1}{M}\sum_{i=1}^{M} \frac{\lambda_i(1+cb)}{\lambda_i+\delta(1+cb)}$.

Proof: See section 3.8.

Next step towards obtaining the $(M, N)-$consistent estimate of the optimal shrinkage coefficient α_o in (3.21) is obtaining the $(M, N)-$consistent estimate of b, ξ, $\eta_d = \mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$ and $\eta_n = \mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$ in (3.22). These are obtained in [124, Appendix A], and summarized in the next lemma.

Lemma 3.4 *Let consider the general asympotic regime where $M, N \to \infty$ and $M/N \to c \in (0, \infty)$. Then, an (M, N)-consistent estimate of b, $\eta_d \triangleq \mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$, ξ and $\eta_n \triangleq \mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$ in (3.22) is given, respectively, by the next expressions,*

$$\hat{b} = \frac{1 - \frac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]}{1 - c(1 - \frac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}])}$$

$$\hat{\eta}_d = (1 - c + c\tfrac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}])\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$$

$$\hat{\xi} = \frac{\frac{1}{M}\operatorname{Tr}[\hat{\mathbf{R}}^2(\hat{\mathbf{R}}+\delta\mathbf{I})^{-2}] - \frac{c}{M^2}\operatorname{Tr}^2[\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]}{1 - c + c\delta^2\frac{1}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-2}]}$$

$$\hat{\eta}_n = \frac{(1 - c(1 - \frac{\delta}{M}\operatorname{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]))^2}{1 - c + \frac{c}{M}\operatorname{Tr}[(\delta^{-1}\hat{\mathbf{R}} + \mathbf{I})^{-2}]}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$$

Proof: The proof is provided in [124, Appendix A], a sketch of the proof is available in section 3.8.

Now, in order to obtain the (M,N)-consistent estimate of the optimal shrinkage factor α_o, one needs to substitute the estimates of b, ξ, η_n and η_d into (3.22). After this step, one needs some manipulations based on the matrix inversion lemma and the next identity [10, p. 248],

$$\frac{1}{M}\,\mathrm{Tr}[\mathbf{R}^2(\mathbf{R}+\delta\mathbf{I})^{-2}] = 1 - \frac{2}{M}\,\mathrm{Tr}[(\delta^{-1}\mathbf{R}+\mathbf{I})^{-1}] + \frac{1}{M}\,\mathrm{Tr}[(\delta^{-1}\mathbf{R}+\mathbf{I})^{-2}]. \tag{3.23}$$

Thus, in the next theorem the expression of the (M,N)-consistent estimate of α_o is given.

Theorem 3.3 *The (M,N)-consistent estimate of the MSE optimal shrinkage factor α_o in (3.21), within the doubly asymptotic regime where $M,N\to\infty$ and $M/N\to c\in(0,\infty)$, is given by the next expression,*

$$\alpha_o \asymp \hat{\alpha}_o = \gamma\beta_1\Big(1-\frac{c}{M}\,\mathrm{Tr}[\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]\Big)^2\cdot\frac{\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}}{\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}} \tag{3.24}$$

Proof: The proof is obtained by substituting the asymptotic equivalences of lemma 3.4 into (3.22) and after some manipulations by taking into account the equality in (3.23).

Note that $\hat{\alpha}_o$ is asymptotically optimal in an MSE sense for any value of β_1 and β_2. At this point, the design of β_1 and β_2 must be tackled. The optimal approach would be to substitute (3.21) into (3.20) and to minimize the MSE (1.5) with respect to β_1 and β_2. Nonetheless, β_1,β_2 could not be isolated due to their presence within the inverse. Indeed, neither one can use a numerical search method to propose a realizable estimator which obtain the optimal β_1,β_2 due to the unknown $\mathbf{R}$ in the expression of the MSE (1.5). To circumvent that and to obtain a realizable filter, one could proceed as in [9], find the (M,N)-consistent estimate for the asymptotic MSE($\alpha=\alpha_o$,β_1,β_2) and find the β_1,β_2 minimizing it. Although a realizable filter is obtained, one still must carry out a numerical search to find β_1 and β_2. Instead of this, an alternative approach is to substitute β_1 and β_2 by the estimates proposed by Ledoit and Wolf in [33], i.e. β_1^{lw} and β_2^{lw} in (1.18). Although this is a suboptimal approach, in this way the numerical search is avoided. Moreover this fits perfectly in our framework and some optimality properties are still kept. That is, β_1^{lw} and β_2^{lw} are asymptotically optimal, as they are (M,N)-consistent estimates of the β_1 and

β_2 that optimize the MSE in the estimation of the covariance, when using the shrinkage estimator $\check{\mathbf{R}} = \beta_1\hat{\mathbf{R}}+\beta_2\mathbf{I}$, $\beta_2 = (1-\beta_1)\,\mathrm{Tr}(\hat{\mathbf{R}}_s)/2M$, and $\hat{\mathbf{R}}_s$ defined above in (1.19). The numerical results in section 3.7 compare this approach to the theoretical lower MSE bound when considering the optimal α_o in (3.21) and β_1, β_2 obtained by a 2D search in (1.5) for a given $\mathbf{R}$, which in practice is unknown, but for simulation purposes is available. Recalling that $\check{\mathbf{R}} = \beta_1\hat{\mathbf{R}}+\beta_2\mathbf{I}$ and that $\delta = \beta_2/\beta_1$, we can manipulate (3.24) to obtain the final expression of the proposed algorithm, which is summarized next.

Proposed estimator

Theorem 3.4 *Consider a set of observations $\{\mathbf{y}(n)\}_{n=1}^N$ fulfilling the model in (1.1) with assumptions (a)-(e). Given $\{\mathbf{y}(n)\}_{n=1}^N$, consider the problem of estimating the parameter of interest $x(n)$ in (1.1) based on a shrinkage of the regularized sample LMMSE. That is, $\hat{x}_{ds}(n) = \mathbf{w}_{ds}^H\mathbf{y}(n)$ with $\mathbf{w}_{ds}=\hat{\alpha}_o\check{\mathbf{R}}^{-1}\mathbf{s}$ and $\check{\mathbf{R}} =\beta_1\hat{\mathbf{R}}+\beta_2\mathbf{I}$. Then, the next estimator asymptotically minimizes the MSE of $\hat{x}_{ds}(n)$ for given $\beta_1^{lw}, \beta_2^{lw}$. And $\beta_1^{lw}, \beta_2^{lw}$ asymptotically optimize the MSE of the data covariance $\check{\mathbf{R}} = \beta_1\hat{\mathbf{R}}+\beta_2\mathbf{I}$, $\beta_2 = (1-\beta_1)\,\mathrm{Tr}(\hat{\mathbf{R}}_s)/2M$.*

$$\hat{x}_{ds}(n) = \mathbf{w}_{ds}^H\mathbf{y}(n); \mathbf{w}_{ds}=\hat{\alpha}_o\check{\mathbf{R}}^{-1}\mathbf{s}$$

$$\hat{\alpha}_o = \gamma(1-\frac{c\beta_1^{lw}}{M}\,\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}])^2\frac{\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s}} \tag{3.25}$$

$$\check{\mathbf{R}} =\beta_1^{lw}\hat{\mathbf{R}}+\beta_2^{lw}\mathbf{I}$$

where the expression of β_1^{lw}, β_2^{lw} and $\hat{\mathbf{R}}_s$ are given in (1.18).

Proof: The proof follows from the asymptotic optimality of $\hat{\alpha}_o$, which is proven in Theorem 3.3 and the asymptotic optimality of $\beta_1^{lw}\beta_2^{lw}$, which is proven in [33].

The simulation results in section 3.7 show that the proposed shrinkage method in (3.25) outperforms other robust methods to the small sample size such as the ad-hoc DL techniques and the LW-LMMSE implementation exposed in section 1.4.1. This is thanks to the additional shrinkage α of the regularized LMMSE filter and thanks to the approach

based on dealing directly with the estimation of the parameter of interest. Moreover, in the numerical results section it is shown that the shrinkage of the regularized LMMSE in (3.25) outperforms the shrinkage of the sample LMMSE proposed in the previous sections. The reason is that in order to estimate $\mathbf{R}$, the former uses a shrinkage of the SCM, which is a better estimator than the SCM, which is the one used in the shrinkage of the sample LMMSE.

3.5 Shrinkage of the sample LMMSE towards a matched filter

Next, a more general shrinkage of the sample LMMSE is introduced. Namely, a common fact in shrinkage estimation is to combine the sample method to be corrected with some a priori information about the parameter to be estimated, which is called shrinkage target. In the problem at hand, we know that the signal of interest is observed through a known signature vector $\mathbf{s}$, as recall that the signal model is $\mathbf{y}(n) = x(n)\mathbf{s} + \mathbf{n}(n)$. For instance in the context of beamforming in array processing $\mathbf{s}$ is the steering vector associated to the direction of arrival of the signal of interest [14]. A conventional filter to recover $x(n)$ is $\mathbf{w} \propto \mathbf{s}$. This is called conventional or Bartlett beamformer in array processing [129] and corresponds to a kind of matched filter. Indeed it maximizes the SINR provided that the term $\mathbf{n}(n)$ only contains additive white noise. Note that in fact it corresponds to an estimate of the LMMSE where one substitutes the unknown covariance $\mathbf{R_n}$ for $\mathbf{I}$[1]. That is, the conventional beamformer can be interpreted as an initial guess of the LMMSE where one assumes that only additive white noise and the signal of interest are present in the scenario. Therefore, one can think of a shrinkage filter where the sample LMMSE is combined with an initial guess or a priori information of the LMMSE filter consisting of the conventional beamformer, i.e. $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$. In fact, note thas this filter is a linear combination of an estimator which takes into account the available observations, with some a priori information which does not take the observations into account. This is more clear if one rewrites the shrinkage filter as $\mathbf{w} = (\alpha_1 \hat{\mathbf{R}}^{-1} + \alpha_2 \mathbf{I})\mathbf{s}$. This expression highlights that the proposed filter is estimating $\mathbf{R}^{-1}$ by means of a shrinkage of the inverse of the SCM, i.e. $\alpha_1 \hat{\mathbf{R}}^{-1} + \alpha_2 \mathbf{I}$, which is a better estimate of $\mathbf{R}^{-1}$ than $\hat{\mathbf{R}}^{-1}$. Or even more precisely, consider the eigendecomposition of the SCM $\hat{\mathbf{R}}^{-1} = \hat{\mathbf{E}}\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{E}}^H$, where $\hat{\mathbf{E}}$ is a matrix which stacks in its columns the eigenvectors of the SCM and $\hat{\mathbf{\Lambda}}$ is a diagonal matrix which contains the

[1] Using the Woodbury's identity and noting that $\mathbf{R} = \gamma \mathbf{s}\mathbf{s}^H + \mathbf{R_n}$ it is easy to show that $\mathbf{w} = \gamma \mathbf{R}^{-1}\mathbf{s}|_{\mathbf{R_n}=\mathbf{I}} = \frac{\gamma}{1+\gamma \mathbf{s}^H \mathbf{s}}\mathbf{s} \propto \mathbf{s}$

eigenvalues of the SCM in its main diagonal. Then, it is clear that the proposed filter can be expressed as $\mathbf{w} = \hat{\mathbf{E}}(\alpha_1 \hat{\mathbf{\Lambda}}^{-1} + \alpha_2 \mathbf{I})\hat{\mathbf{E}}^H \mathbf{s}$. Therefore the proposed filter is implementing a correction of the sample LMMSE which consists of a shrinkage of the eigenvalues of $\hat{\mathbf{R}}^{-1}$. Another interpretation is that the linear combination takes into account the benefits of the sample LMMSE and the matched filter. On the one hand, the sample LMMSE is optimal for a large sample size regime and it has better rejection against the interference than the matched filter. On the other hand, the conventional beamformer may give in general better performance than the sample LMMSE in small sample size situations, as it is not affected by the signal cancelation effect arising from the scarcity of available observations.

Note that the shrinkage method proposed in Lemma 3.1 is indeed a particular case of $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$ with $\alpha_2 = 0$. Thus, the filter that we are now introducing incorporates more a priori information or in other words is a more general form of shrinkage filter than the one of Lemma 3.1. Thereby, the filter proposed in this section is expected to give better performance than the filter introduced in lemma 3.1. Next, we study the estimation of the parameter of interest $x(n)$ when using the type of shrinkage filter $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$. Namely, first the optimal shrinkage factors α_1 and α_2 are designed to optimize the MSE in the estimation of the parameter of interest. This is presented in the next lemma.

Lemma 3.5 *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, 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 matched filter. This problem is mathematically formulated as follows,*

$$\hat{x}_{lb,s}(n) = \mathbf{w}_{lb,s}^H \mathbf{y}(n); \qquad \mathbf{w}_{lb,s} = \arg\min_{\mathbf{w}} \mathbb{E}\left[\left|x(n) - \mathbf{w}^H \mathbf{y}(n)\right|^2 \mid \hat{\mathbf{R}}\right] \quad s.t. \mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s} \tag{3.26}$$

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

$$\boldsymbol{\alpha}_{lb} = \frac{\gamma \begin{pmatrix} \mathbf{s}^H \mathbf{R}\mathbf{s}\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \\ \mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H \mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \end{pmatrix}}{\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H \mathbf{R}\mathbf{s} - \mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}\mathbf{s}^H \mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}} \tag{3.27}$$

Proof: See section 3.8.

■

The expression of the optimal shrinkage LMMSE estimator in (3.27) highlights the dependance on the unknown $\mathbf{R}$. As a consequence it is not a realizable estimator. 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. [130] 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 the proposed shrinkage estimator of the sample LMMSE in (3.27) leads to the conventional sample LMMSE method (1.9), as $\boldsymbol{\alpha}_{lb|\mathbf{R}=\hat{\mathbf{R}}} = (\gamma, 0)^T$, and as a consequence the potencial benefits of the shrinkage approach are lost.

Next, in order to tackle this problem, i.e. to obtain a realizable estimator, a strategy based on RMT is proposed. This approach is based on obtaining a consistent estimator of the shrinkage coefficients in (3.27). That is, it leads to obtain an asymptotically optimal shrinkage estimator. More specifically, in order to obtain the consistent estimate, the asymptotic regime where $M, N \to \infty$ with $M/N \to c \in (0, 1)$ is considered. Note that this general asymptotics considers implicitly the small sample size regime when $c \to 1$ and as a consequence the proposed method is robust to this regime. Indeed, this generalizes classical consistent methods based on classical asymptotics where M remains fixed and N tends to infinity. Thereby, in the next theorem the consistent estimate of the optimal shrinkage coefficients α_1, α_2 in (3.27) is presented.

Theorem 3.5 *Let define* $\check{\boldsymbol{\alpha}}_{lb} \triangleq (\check{\alpha}_{lb,1}\ ,\ \check{\alpha}_{lb,2})^T$*, then a realizable and consistent estimate of the optimal shrinkage of the sample LMMSE towards a matched filter (3.26), within the general asymptotics where* $M, N \to \infty$*,* $M/N \to c \in (0, 1)$*, reads as follows,*

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

$$\check{\boldsymbol{\alpha}}_{lb} = \frac{\gamma \begin{pmatrix} (1-c)^2 \mathbf{s}^H \hat{\mathbf{R}} \mathbf{s}\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s} - (1-c) \\ c\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s} \end{pmatrix}}{\mathbf{s}^H \hat{\mathbf{R}} \mathbf{s}\mathbf{s}^H \hat{\mathbf{R}}^{-1}\mathbf{s} - 1} \tag{3.28}$$

Proof: See section 3.8.

■

The proposed estimator in (3.28) is robust to the small sample size regime. Effectively, on the one hand it incorporates a shrinkage correction through the coefficients α_1, α_2. On

the other hand, the RMT approach implicitly considers this scenario through $c \to 1$. 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 (3.28) outperforms the traditional sample LMMSE estimator (1.9) in any sample size regime, 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. $N \gg M$ 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. Interestingly enough, the numerical results section shows that provided that $M/N \in (0,1)$ the proposed shrinkage method in Theorem 3.5 outperforms other methods that are robust to the small sample size regime, such as ad-hoc DL techniques or the LW implementation of the LMMSE, cf. chapter 1 and section 1.4.1. This is because on the one hand the DL methods in (1.12) and (1.13) are rather ad-hoc. On the other hand, the LW technique in (1.16) relies on a shrinkage estimator of the sample covariance which minimizes asympotitcally the MSE in the estimation of the data covariance. Though this is a better estimate than the SCM it is not the final target, as here we are interested in estimating $x(n)$. On the other hand, the proposed method faces directly the estimation of the parameter of interest $x(n)$ by obtaining the shrinkage coefficients that minimize asymptotically its MSE. The numerical simulations will also highlight that compared to [9], which proposes a DL which optimizes asymptotically the SINR, the proposed shrinkage in Theorem 3.5 obtains almost the same SINR performance and better MSE performance.

In order to gain more insights about how the shrinkage estimation framework affects the method proposed in Theorem 3.5 it is interesting to study the asymptotic values of the shrinkage factors 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. Thereby, when the sample dimension is much larger than the observation dimension, i.e. $c \to 0$, the next expression of $\check{\boldsymbol{\alpha}}_{lb}$ reveals that the performance of the shrinkage filter tends to the one of the traditional LMMSE implementation.

$$c \to 0 \Rightarrow \check{\boldsymbol{\alpha}}_{lb} \to (\gamma, 0)^T. \tag{3.29}$$

This behavior makes sense as in this situation, $\hat{\mathbf{R}}$ is the optimal estimator of $\mathbf{R}$ and as a consequence the performance of the traditional sample implementation of the LMMSE

tends to the one of the optimal LMMSE filter in (1.4). With regard to the case where $c \rightarrow 1$, i.e. in the small sample size regime, it is easy to check that the following relation for the shrinkage factors holds,

$$c \rightarrow 1 \Rightarrow \check{\boldsymbol{\alpha}}_{lb} \rightarrow \left(0, \gamma \frac{\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}}{\mathbf{s}^H \hat{\mathbf{R}} \mathbf{s} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s} - 1} \right)^T . \tag{3.30}$$

That is, in the small sample size regime, the performance of the shrinkage filter tends to the one of a type of matched filter, or conventional beamformer, and it disregards the contribution of the sample LMMSE, as it has in general worse performance than a matched type filter. The expressions (3.29) and (3.30) 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 type of matched filter may behave better than the sample LMMSE, the proposed shrinkage filter tends to a scaling of a matched filter.

It is also worth observing that the method proposed in Theorem 3.1 is likely to behave worse than the one in Theorem 3.5, 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 3.5 incorporates more a priori information in the structure of the estimator, through the presence of the matched filter $\mathbf{w} \propto \mathbf{s}$, which may in general behave better than the sample LMMSE in the small sample size regime.

3.6 Shrinkage and regularization of the sample LMMSE towards a matched filter

The filter presented in this section is the most general among the methods that are proposed in this chapter. It arises from the same shrinkage philosophy than the method proposed in section 3.5, though it is more general as it avoids the assumption that $N > M$. That is, an estimator of the LMMSE relying on the information available from the samples is combined with an estimator which does not take this information into account and only relies on the a priori information. More precisely, on the one hand the unknown inverse covariance in the LMMSE is estimated from the available samples and on the other hand it is linearly combined with the a priori information, i.e. with a matched filter, which

arises after substituting the unknown noise covariance within the inverse data covariance by just the identity matrix. This is mathematically expressed as $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, where $\check{\mathbf{R}}^{-1}$ is an estimator of $\mathbf{R}^{-1}$ based on the available samples. In the previous section $\mathbf{R}^{-1}$ was estimated by considering the inverse of the SCM matrix. Herein, in order to cope with the cases where $M \geq N$, a regularization of the SCM is considered thereby $\check{\mathbf{R}}^{-1} = (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}$. In fact, this may lead to performance improvements for the rest of the cases where $N > M$, as $\check{\mathbf{R}}^{-1}$ is a better estimator of $\mathbf{R}^{-1}$ than just considering the inverse of the SCM. Thereby, the shrinkage filter proposed in this section has the next structure,

$$\mathbf{w} = \alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}\mathbf{s} + \alpha_2 \mathbf{s} \tag{3.31}$$

The expression (3.31) may be interpreted as follows if one rewrites this expression as $\mathbf{w} = (\alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} + \alpha_2 \mathbf{I})\mathbf{s}$. On the one hand, a shrinkage estimation of the inverse covariance is carried out by means of the shrinkage factors α_1, α_2. On the other hand, the covariance is estimated from the available samples by means of the sample covariance matrix, which on its turn is regularized through δ to allow its inversion when $N > M$. Interestingly enough, chapter 5 highlights another appealing feature of this filter. Namely, on the one hand the regularized sample LMMSE has better rejection capabilities against the interference than the matched filter. On the other hand, when there is an uncertainty in $\mathbf{s}$, the regularized sample LMMSE designed to cope with the finite sample size may still undergo a notable performance degradation due to the signal cancellation effect. That is it may tend to cancel the signal of interest, whereas the matched filter will not have this behavior, i.e. may experiment a degradation due to the uncertainty in $\mathbf{s}$, but it will not lead to cancel the signal of interest. Thereby, the type of filter (3.31) will obtain better performance than just considering a type of DL filter $\mathbf{w} = \alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}\mathbf{s}$ designed to cope in an asymptotically optimal manner with the finite sample size, e.g. [9] and [79].

Next, the design of the proposed filter in (3.31) is tackled. Namely, this implies how to obtain a value for the shrinkage factors α_1, α_2 and the regularization parameter δ. In order to achieve this aim, the values of α_1, α_2 and δ are obtained as the ones which minimize the MSE in the estimation of the parameter of interest $x(n)$ in (1.1). Recall that given the signal model (1.1), the expression of the MSE for a linear estimation of $x(n)$ based on a generic filter $\mathbf{w}$ is given by,

$$\text{MSE} = \mathbf{w}^H \mathbf{R} \mathbf{w} + \gamma (1 - \mathbf{w}^H \mathbf{s} - \mathbf{s}^H \mathbf{w}) \tag{3.32}$$

With $\gamma = \mathbb{E}[|x|^2]$. Thereby, the design of the shrinkage factors α_1, α_2 and the regularization parameter δ are obtained by means of the next optimization problem,

$$\begin{array}{ll}\min\limits_{\mathbf{w}} & \mathbf{w}^H\mathbf{R}\mathbf{w}+\gamma(1-\mathbf{w}^H\mathbf{s}-\mathbf{s}^H\mathbf{w}) \\ \text{s.t.} & \mathbf{w}=\alpha_1(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}+\alpha_2\mathbf{s}.\end{array}$$

Namely, this problem is solved following a two step approach. First, the optimal value for $\boldsymbol{\alpha}=(\alpha_1,\alpha_2)^T$ is obtained for any given δ. Then, the optimal $\boldsymbol{\alpha}$ is substituted in the expression of the MSE (3.32) to obtain the optimal δ. This result is formally presented in the next lemma.

Lemma 3.6 *Assume that a set of observations $\{\mathbf{y}(n)\}_{n=1}^N$ fulfilling the model in (1.1), with assumptions (a),(b),(c) and (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, when the estimator $\hat{x}(n)=\mathbf{w}^H\mathbf{y}(n)$ is a shrinkage of a regularized sample LMMSE towards a matched filter, i.e. $\mathbf{w}=\alpha_1(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}+\alpha_2\mathbf{s}$. This problem can be mathematically expressed as*

$$\begin{array}{rl}\hat{x}(n)=\mathbf{w}^H\mathbf{y}(n);\quad \mathbf{w}=\arg\min\limits_{\mathbf{w}} & \mathbf{w}^H\mathbf{R}\mathbf{w}+\gamma(1-\mathbf{w}^H\mathbf{s}-\mathbf{s}^H\mathbf{w}) \\ s.t. & \mathbf{w}=\alpha_1(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}+\alpha_2\mathbf{s}\end{array} \tag{3.33}$$

Then, defining $\check{\mathbf{R}}=\hat{\mathbf{R}}+\delta\mathbf{I}$, the optimal shrinkage factor $\boldsymbol{\alpha}_o\triangleq(\alpha_1\ ,\ \alpha_2)^T$ for any given regularization δ is given by,

$$\boldsymbol{\alpha}_o=\gamma\frac{\begin{pmatrix}\mathbf{s}^H\mathbf{R}\mathbf{s}\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}-\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}\\ \mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}-\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}\end{pmatrix}}{\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H\mathbf{R}\mathbf{s}-\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}\mathbf{s}^H\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s}}. \tag{3.34}$$

Moreover, the optimal regularization δ_o is obtained by means of the next optimization,

$$\begin{aligned}\delta_o=\arg\min_{\delta}\ \boldsymbol{\alpha}_o^H\begin{pmatrix}\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} & \mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}\\ \mathbf{s}^H\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} & \mathbf{s}^H\mathbf{R}\mathbf{s}\end{pmatrix}\boldsymbol{\alpha}_o+\\ \gamma\left(1-\boldsymbol{\alpha}_o^H\begin{pmatrix}\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}\\ \mathbf{s}^H\mathbf{s}\end{pmatrix}-\left(\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s},\mathbf{s}^H\mathbf{s}\right)\boldsymbol{\alpha}_o\right)\end{aligned} \tag{3.35}$$

Proof: The proof for $\boldsymbol{\alpha}_o$ follows easily from the proof for Lemma 3.5, just considering $\check{\mathbf{R}}$ instead of $\hat{\mathbf{R}}$. The proof for Lemma 3.5 is detailed in the appendix of this chapter, section 3.8. On the other hand, δ_o is obtained after easy manipulations by considering the

type of filter $\mathbf{w} = \alpha_1(\hat{\mathbf{R}} + \delta\mathbf{I})^{-1}\mathbf{s} + \alpha_2\mathbf{s}$ and by substituting the optimal shrinkage $\boldsymbol{\alpha}_o$ in the expression of the MSE, i.e. in (3.32).

■

At this point, note that the optimal values of the shrinkage and regularization parameters, i.e. $\boldsymbol{\alpha}_o$ and δ_o in Lemma 3.6 depend on the unknown data covariance $\mathbf{R}$. Namely, $\boldsymbol{\alpha}_o$ and δ_o depend on the next quantities,

$$\begin{aligned} &\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} \\ &\mathbf{s}^H\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} \\ &\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \\ &\mathbf{s}^H\mathbf{R}\mathbf{s}. \end{aligned} \tag{3.36}$$

Thereby, the next task is to estimate these unknown quantities to obtain a realizable estimator of $\boldsymbol{\alpha}_o$ and δ_o in Lemma 3.6. To this end, a random matrix theory approach is considered. This permits to obtain estimators that tend to the optimal $\boldsymbol{\alpha}_o$ and δ_o within the asymptotic regime where $(M, N) \to \infty$ and $M/N \to c \in (0, \infty)$. This framework permits to obtain consistent estimators that take into account the finite sample size regime. The consistent estimation for the terms in (3.36) within the asymptotic regime where $(M, N) \to \infty$ and $M/N \to c \in (0, \infty)$ is provided in Lemma 2.2 and they are stated next for the sake of the clarity in the exposition,

$$\begin{aligned} \mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} &\asymp \frac{1}{\left(1 - \frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}]\right)^2}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s} \\ \mathbf{s}^H\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} &\asymp \frac{1}{1 - c + c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\mathbf{s}^H\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s} \\ \mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} &\asymp \frac{1}{1 - c + c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\mathbf{s} \\ \mathbf{s}^H\mathbf{R}\mathbf{s} &\asymp \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}. \end{aligned} \tag{3.37}$$

Where $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta\mathbf{I}$. The results in (3.37) pave the way to obtain an estimator that tends to the optimal shrinkage of the regularized LMMSE in Lemma 3.6. This is summarized in the next theorem.

Theorem 3.6 *Let denote* $\check{\boldsymbol{\alpha}}_o = (\check{\alpha}_{o,1}, \check{\alpha}_{o,2})^T = \hat{\boldsymbol{\alpha}}_{o_{|\delta=\hat{\delta}_o}}$ *and* $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta\mathbf{I}$. *Then, a realizable and (M,N)-consistent estimate of the optimal shrinkage LMMSE estimator in lemma 3.6, within the general asymptotic framework where* $M, N \to \infty$ *and* $M/N \to c \in (0, \infty)$, *reads as follows,*

$$\check{x}(n) = \mathbf{w}^H \mathbf{y}(n); \mathbf{w} = \check{\alpha}_{o,1}(\hat{\mathbf{R}} + \hat{\delta}_o \mathbf{I})^{-1}\mathbf{s} + \check{\alpha}_{o,2}\mathbf{s} \tag{3.38}$$

$$\hat{\boldsymbol{\alpha}}_o = \frac{\gamma \begin{pmatrix} \mathbf{s}^H \hat{\mathbf{R}} \mathbf{s} \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{s} - \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]} \mathbf{s}^H \check{\mathbf{R}}^{-1} \hat{\mathbf{R}} \mathbf{s} \\ \frac{1}{\left(1-\frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}]\right)^2} \mathbf{s}^H \check{\mathbf{R}}^{-1} \hat{\mathbf{R}} \check{\mathbf{R}}^{-1} \mathbf{s} - \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{s} \left(\frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\right) \mathbf{s}^H \hat{\mathbf{R}} \check{\mathbf{R}}^{-1} \mathbf{s} \end{pmatrix}}{\frac{1}{\left(1-\frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}]\right)^2} \mathbf{s}^H \check{\mathbf{R}}^{-1} \hat{\mathbf{R}} \check{\mathbf{R}}^{-1} \mathbf{s} \mathbf{s}^H \hat{\mathbf{R}} \mathbf{s} - \left(\left(\frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\right) \mathbf{s}^H \hat{\mathbf{R}} \check{\mathbf{R}}^{-1} \mathbf{s}\right)^2} \tag{3.39}$$

$$\hat{\delta}_o = \arg\min_{\delta} \hat{\boldsymbol{\alpha}}_o^H \begin{pmatrix} \frac{1}{\left(1-\frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}]\right)^2} \mathbf{s}^H \check{\mathbf{R}}^{-1} \hat{\mathbf{R}} \check{\mathbf{R}}^{-1} \mathbf{s} & \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]} \mathbf{s}^H \check{\mathbf{R}}^{-1} \hat{\mathbf{R}} \mathbf{s} \\ \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]} \mathbf{s}^H \hat{\mathbf{R}} \check{\mathbf{R}}^{-1} \mathbf{s} & \mathbf{s}^H \hat{\mathbf{R}} \mathbf{s} \end{pmatrix} \hat{\boldsymbol{\alpha}}_o + \gamma \left(1 - \hat{\boldsymbol{\alpha}}_o^H \begin{pmatrix} \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{s} \\ \mathbf{s}^H \mathbf{s} \end{pmatrix} - \left(\mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{s}, \mathbf{s}^H \mathbf{s}\right) \hat{\boldsymbol{\alpha}}_o \right) \tag{3.40}$$

Proof: The proof is based on Lemma 2.2, as it provides the (M, N)-consistent estimates of the unknown quantities in Lemma 3.6.

■

Remark 1: In (3.39) and (3.40) it has been used $\mathbf{s}^H \hat{\mathbf{R}} \mathbf{s}$ as an estimate of the unknown $\mathbf{s}^H \mathbf{R} \mathbf{s}$ in Lemma 3.6. An alternative estimate for $\mathbf{s}^H \mathbf{R} \mathbf{s}$ is $\mathbf{s}^H (\hat{\mathbf{R}} + \hat{\delta}_o \mathbf{I}) \mathbf{s}$. In numerical simulations it has been observed that this latter approach leads to slightly better performance in the estimation of the parameter of interest $x(n)$.

Remark 2: In order to find the optimal value for $\hat{\delta}_o$, a one dimensional search is needed, as $\hat{\delta}_o$ is the argument optimizing (3.40). This requires matrix inversions for each iteration of the search due to the expressions involved in (3.39) and (3.40). Fortunately, these matrix inversions can be avoided, which leads to reduce the computational cost of the numerical search. To achieve this aim, first the next identities can be considered, where $\hat{\lambda}_m$ and $\hat{\mathbf{e}}_m$ denote the m-th sample eigenvalue of $\hat{\mathbf{R}}$ and its associated eigenvector, respectively,

$$
\begin{aligned}
\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} &= \sum_{m=1}^{M}\frac{|\mathbf{s}^H\hat{\mathbf{e}}_m|^2}{\delta+\hat{\lambda}_m} \\
\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} &= \sum_{m=1}^{M}\frac{|\mathbf{s}^H\hat{\mathbf{e}}_m|^2\hat{\lambda}_m}{(\delta+\hat{\lambda}_m)^2} \\
\mathrm{Tr}[\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}] &= \sum_{m=1}^{M}\frac{\hat{\lambda}_m}{\delta+\hat{\lambda}_m} \\
\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}] &= \sum_{m=1}^{M}\frac{1}{\delta+\hat{\lambda}_m} \\
\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}\mathbf{s} &= \sum_{m=1}^{M}\frac{|\mathbf{s}^H\hat{\mathbf{e}}_m|^2\hat{\lambda}_m}{(\delta+\hat{\lambda}_m)} \\
\mathbf{s}^H\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} &= \sum_{m=1}^{M}\frac{|\mathbf{s}^H\hat{\mathbf{e}}_m|^2\hat{\lambda}_m}{(\delta+\hat{\lambda}_m)}.
\end{aligned}
\tag{3.41}
$$

Moreover, note that $\{\hat{\lambda}_m\}_{m=1}^{M}$ and the expression $|\mathbf{s}^H\hat{\mathbf{e}}_m|^2$ need to be computed and stored only once, i.e. there is no need to recompute these values for each iteration of the numerical search.

3.7 Numerical simulations

This section is devoted to study the performance of the proposed estimators by means of numerical simulations. Namely, these methods are the shrinkage of the sample LMMSE proposed in (3.3), the shrinkage of the sample LMMSE for gaussian data in (3.7), the shrinkage of the regularized sample LMMSE in (3.25), the shrinkage of the sample LMMSE towards a matched filter in (3.28) and the shrinkage of the regularized sample LMMSE towards a matched filter in (3.38). The performance of these methods is compared to that of the optimal LMMSE and its traditional sample based implementation in (1.4) and (1.9), respectively. Moreover, the proposed methods are also compared to other methods which are robust to the small sample size regime. These are the ad-hoc choices of the DL in section 1.4.1. Also the LW method, which obtains the asymptotically optimal shrinkage of the sample covariance in terms of the MSE in the estimation of the covariance. Another robust method considered in the comparisons is the DL method in [9], which obtains the asymptotically optimal loading factor in terms of the SINR at the ouput of the filter. And

finally the regularized sample LMMSE in [79], which obtains the shrinkage factors of the regularization of the SCM which optimize asymptotically the MSE in the estimation of the parameter of interest x in (1.1). Recall that all these related work methods are summarized in section 1.4.1.

In order to define the simulations to be carried out observe that the general expression of the MSE in (1.5) depends on $\mathbf{R_n}$, γ and $\mathbf{s}$, whereas the estimators to be compared depend on c, $\frac{M}{N}$, $\hat{\mathbf{R}}$, $\mathbf{R}$, γ and $\mathbf{s}$. Beamforming in array signal processing is considered as an application to specify the models for these parameters. Nonetheless, they are flexible enough to be applied to other fields of signal processing, e.g. in spectrum analysis. According to the assumptions in (1.1),

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

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

$$[\mathbf{s}]_m = \frac{1}{\sqrt{M}} e^{j\pi \sin\theta_0 m}$$

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 [14]. Moreover, for the simulation purposes θ_0 is set to 0, unless stated otherwise. With regard to $\mathbf{R_n}$, a standard model, see [14], is

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

where $\mathbf{S}$ is the matrix of steering vectors of the interferers. $[\mathbf{S}]_{m,k} = \frac{e^{j\pi \sin\theta_k m}}{\sqrt{M}}$, $m = 0, \ldots, M-1$ is the antenna index, $k = 1, \ldots, K$ defines a set of interferers and θ_k is the DOA of the *k-th* interferer. For the simulations, from figure 3.1 to figure 3.4 $\theta_k = (2 + 10(k-1))\frac{\pi}{180}$, from figure 3.5 to 3.6 $\left\{\theta_k \frac{180^\circ}{\pi}\right\}_{k=1}^{4} = \{45^\circ, -45^\circ, 85^\circ, -85^\circ\}$ and from figure 3.7 to 3.10 $\left\{\theta_k \frac{180^\circ}{\pi}\right\}_{k=1}^{8} = \{8^\circ, -15^\circ, 23^\circ, -21^\circ, 46^\circ, -44^\circ, -85^\circ, 74^\circ\}$. $\mathbf{P}$ is the covariance matrix of the interferers and σ^2 is the power of an AWGN. $\mathbf{P}$ 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. 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. The value for SIR_k and SNR is set to 0 dB and to 5 dB, respectively, unless stated otherwise. With regard to c it is set to $c = \frac{M}{N}$. M is fixed in the simulations and N varies fulfilling different sample size regimes. Finally, the sample

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

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

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,

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

Next, with the simulation conditions at hand, the performance of the proposed estimators is assessed. The first set of simulations in figures 3.1 to 3.4 assess the performance of the proposed methods based on the most basic form of shrinkage presented in this chapter, i.e. $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$. These methods are on the one hand, the shrinkage of the sample LMMSE based on optimizing the MSE and using RMT tools, i.e. with (3.3). On the other hand, the estimator in (3.7), based on a shrinkage of the sample LMMSE, which optimizes the average MSE and which uses summary statistics of the complex inverse Wishart distribution. As these methods intend to overcome the performance degradation of the sample LMMSE, in the small sample size regime, we compare their performance with that of the sample and the theoretical LMMSE methods in (1.9) and (1.4), respectively. Given a fixed M, the MSE of the estimators is displayed as a function of N, which varies to simulate different sample size regimes. Moreover, the MSE is computed by substituting the expression of the filter of each estimator in (1.5).

Figure 3.1, compares the sample LMMSE method, the theoretical LMMSE and the proposed Shrinkage LMMSE method in (3.7) when $M = 5$. It can be observed that the proposed method dramatically outperforms the sample LMMSE in the small sample size regime. This behavior is due to the robustness of the shrinkage methods to the small sample size regimes. Moreover, figure 3.1 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 it remains close to the theoretical LMMSE estimator. Another comment worth mentioning is the evolution as the sample size tends to be large. 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. Thereby, the performance of the sample LMMSE tends to the one of the theoretical LMMSE. With regard to the proposed shrinkage of the sample LMMSE, it is easy to check that when N grows large compared to M, then the shrinkage factor in (3.7) fulfills

$$\hat{\alpha}_{ls,s} \rightarrow \gamma.$$

Thereby, the performance of the proopsed shrinkage tends to the one of the sample LMMSE, which on its turn tends to the performance of the LMMSE, as it was just mentioned.

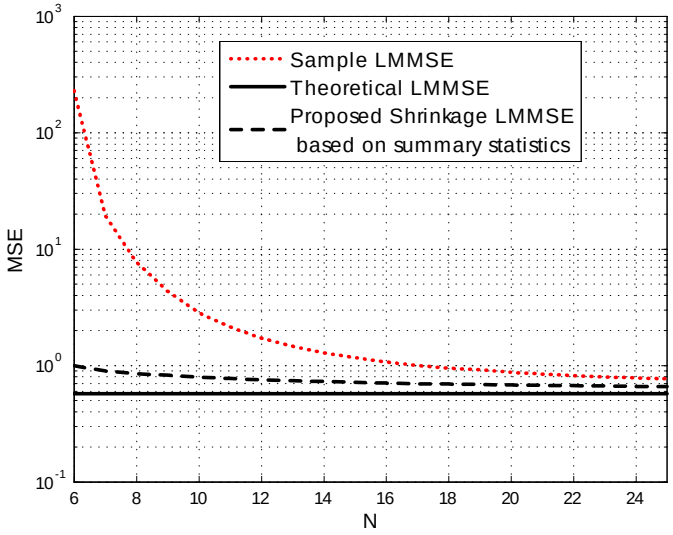

Figure 3.1: Performance comparison between proposed Shrinkage LMMSE based on summary statistics (3.7), theoretical LMMSE estimator (1.4) and sample LMMSE (1.9) when $M = 5$, $SNR = 5$ dB and $SIR_i = 10$ dB.

The next set of simulations is presented in figures 3.2 to 3.4. 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 based on RMT, proposed in (3.3), and the one based on optimizing the average MSE proposed in (3.7). For the simulation purposes, the theoretical LMMSE is also considered and all the simulation parameters, except M, are the same than in figure 3.1. This set of simulations highlight that both shrinkage methods

are robust to the small sample size regime, though the method proposed in (3.7) is slightly better. This behavior is due to the fact that the shrinkage method based on RMT is optimum when $M, N \to \infty$ and $\frac{M}{N} \in (0, 1)$, see Theorem 3.1. That is to say, that method is obtained by finding an estimator which tends asympotically to the optimal though unrealizable method (3.2) 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 (3.7), based on averaging the MSE and using the summary statistics of the complex inverse Wishart distribution, 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 3.4. 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 (3.7) that does not assume any asymptotic approximation.

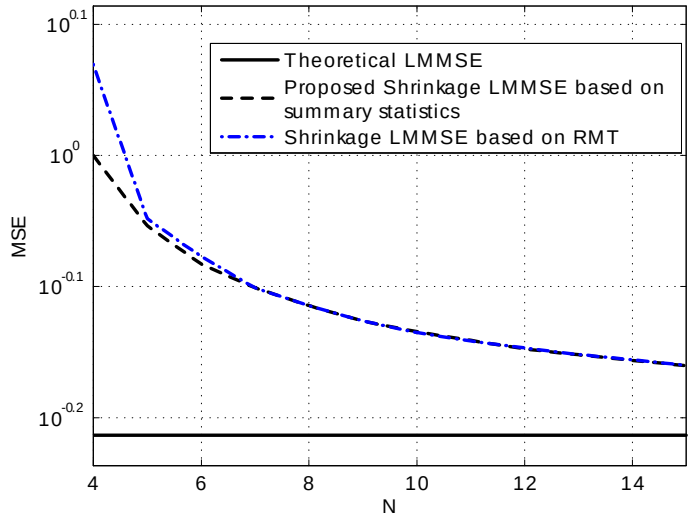

Figure 3.2: Performance comparison between proposed Shrinkage LMMSE based on summary statistics (3.7), theoretical LMMSE estimator (1.4) and Shrinkage LMMSE based on RMT (3.3) when M=3 , $SNR = 5$ dB and $SIR_i = 10$ dB.

Next, in figures 3.5 and 3.6 the performance of the shrinkage of the regularized LMMSE (3.25) is studied. Recall that this method generalizes the shrinkage of the sample LMMSE in (3.3) to support any sample size regime, i.e. $M/N \in (0, \infty)$. The proposed shrinkage

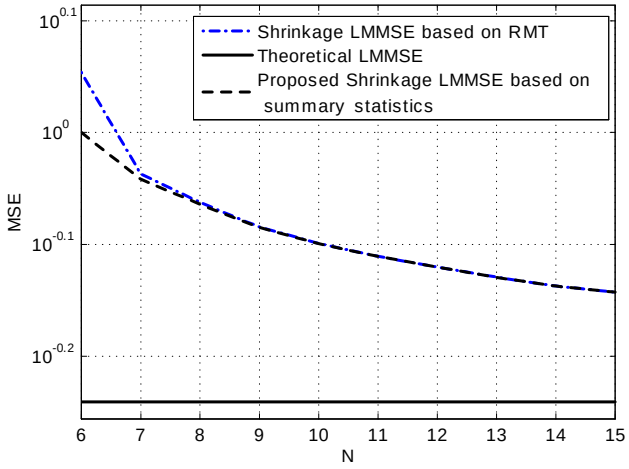

Figure 3.3: Performance comparison between proposed Shrinkage LMMSE based on summary statistics (3.7), theoretical LMMSE estimator (1.4) and Shrinkage LMMSE based on RMT (3.3) when M=5, $SNR = 5$ dB and $SIR_i = 10$ dB.

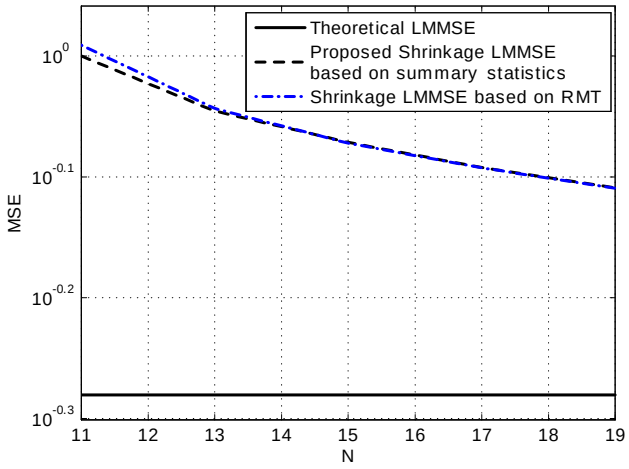

Figure 3.4: Performance comparison between proposed Shrinkage LMMSE based on summary statistics (3.7), theoretical LMMSE estimator (1.4) and Shrinkage LMMSE based on RMT (3.3) when M=10, $SNR = 5$ dB and $SIR_i = 10$ dB.

of the regularized sample LMMSE (3.25) is compared to the theoretical LMMSE (1.4), its sample implementation (1.9) and the shrinkage of the sample LMMSE (3.3). Moreover as a lower bound the optimal though unrealizable shrinkage of regularized sample LMMSE is considered. This is obtained considering the optimal shrinkage (3.21) in (3.20) and obtaining β_1 and β_2 that minimize the MSE (1.5) by means of a grid search. M is set to 16 and $N \in [16, 300]$, because (1.9) and (3.3) can not deal with $M > N$. The proposed shrinkage of the sample LMMSE and regularized sample LMMSE dramatically outperform the sample LMMSE in the small sample size regime. In the large sample size regime all the methods converge, as in this situation the SCM becomes a good estimate of $\mathbf{R}$. Moreover, the shrinkage of the regularized sample LMMSE (3.25) improves the shrinkage of the sample LMMSE (3.3), since it relies on the shrinkage of the SCM, which is a better estimate of $\mathbf{R}$ than the SCM, the one used by (3.3). Indeed (3.25) may be viewed as a generalized version of (3.3) to support $M/N \in (0, \infty)$.

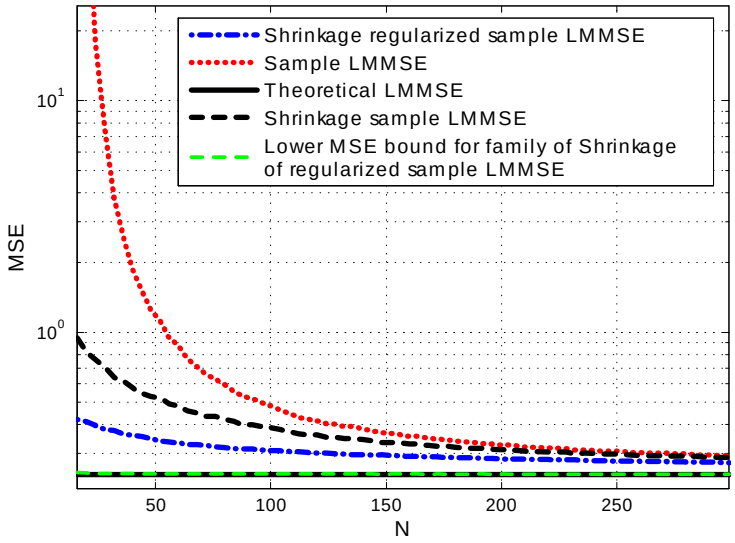

Figure 3.5: Performance comparison between proposed shrinkage of regularized sample LMMSE (3.25), theoretical LMMSE estimator (1.4), sample LMMSE (1.9) and shrinkage of sample LMMSE (3.3).

In figure 3.6, the proposed shrinkage of the regularized sample LMMSE (3.25) is compared to the DL-LMMSE (1.10) and the LW-LMMSE (1.16), i.e. to other methods that are robust to the small sample size and that support $M > N$, thus in this figure $M = 16$

and $N \in [4, 300]$. In order to implement the DL-LMMSE, the DL factor considered is the one proposed in [77], which arises from analyzing the estimation error of the covariance, i.e. δ is chosen to be equal to the standard deviation of the diagonal entries of the SCM, see section 1.4.1. Moreover as a lower bound the LMMSE (1.4) is considered and also the optimal though unrealizable shrinkage of regularized sample LMMSE. This is obtained considering the optimal shrinkage (3.21) in (3.20) and obtaining β_1 and β_2 that minimize the MSE (1.5) by means of a grid search. The estimator proposed clearly outperforms the other alternative methods, i.e. LW and DL implementations. This is due to the additional shrinkage of the filter governed by $\hat{\alpha}_o$ (3.25) and a design based on facing the minimization of the MSE in the estimation of $x(n)$. On the other hand, recall that the ad-hoc DL-LMMSE and LW-LMMSE aim to enhance the data covariance estimate which is not the final target. Moreover, LW-LMMSE outperforms DL-LMMSE as its regularization of $\hat{\mathbf{R}}$ is asymptotically optimal in terms of the MSE in the estimation of the data covariance.

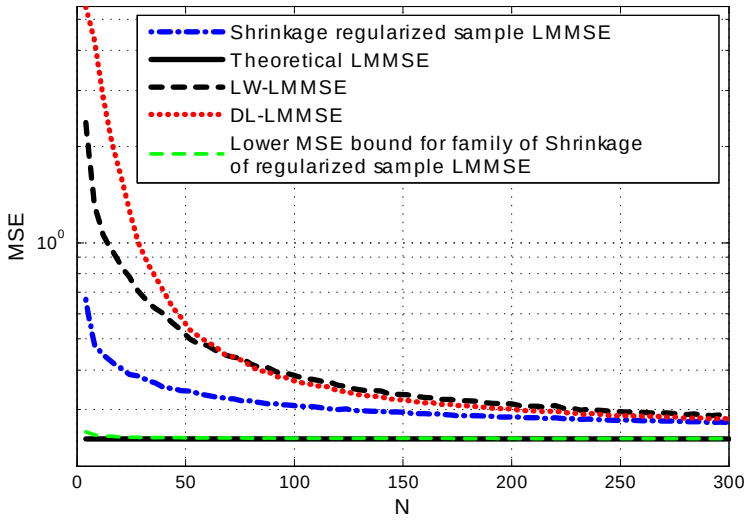

Figure 3.6: Performance comparison between shrinkage of regularized sample LMMSE (3.25), theoretical LMMSE estimator (1.4), LW-LMMSE (1.16) and DL-LMMSE (1.10).

In figure 3.7, the performance of the shrinkage of the sample LMMSE towards matched filter in (3.28) is evaluated for a small value of $M = 10$, whereas in figure 3.8 the same type of simulation than in figure 3.7 is carried out, but for a relative high value of $M = 50$. For comparison purposes the more basic shrinkage of the sample LMMSE proposed in (3.3) is considered as well as the theoretical LMMSE in (1.4) and the conventional sample LMMSE

in (1.9). The first aim of these two figures is to assess the performance degradation of the proposed methods for a small value of M compared to a large value of this parameter. This is interesting as the proposed methods are optimal for a large M, because they are (M, N)-consistent estimates. Comparing figures 3.7 and 3.8 one can see that the degradation is small and as a consequence the proposed shrinkage LMMSE methods offer a good performance even for rather small values of M. In fact one can observe that the performance of the shrinkage of the sample LMMSE towards matched filter is very close to the optimal LMMSE.

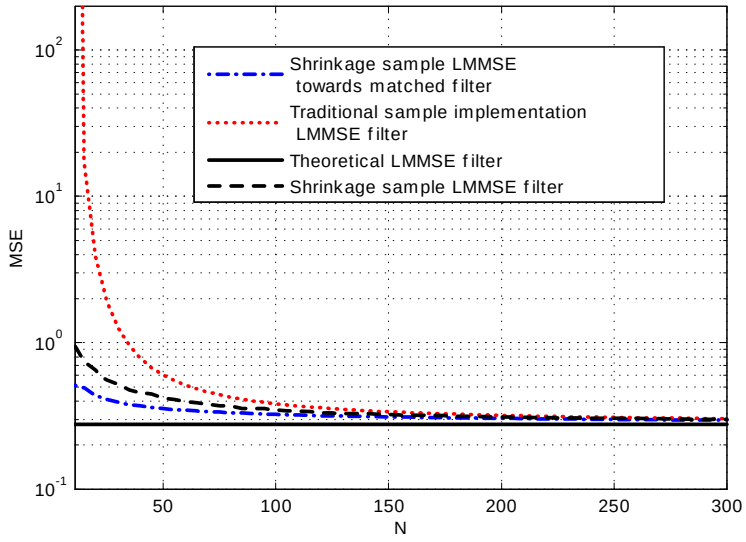

Figure 3.7: Performance comparison between the shrinkage of the sample LMMSE (3.3), the shrinkage of the sample LMMSE towards a matched filter (3.28), the LMMSE (1.4) and the sample LMMSE (1.9), when SNR=5 dB, SIR_i=0 dB and M=10.

Furthermore, figures 3.7 and 3.8 highlight that the proposed methods are robust to the small sample size regime thanks to the shrinkage correction and the design based on random matrix theory. Indeed they outperform the sample LMMSE for any sample size, specially in the small sample size regime where the improvement is huge. It is also interesting to observe that for N growing large, compared to M, the shrinkage, the theoretical and the sample estimators tend to converge. This is because in this case the SCM is the MVUE of $\mathbf{R}$ and it is well conditioned. As a consequence, the sample LMMSE tends to be optimal, i.e. tends to the LMMSE. The shrinkage estimators are aware of this situation and reflect it by means

of the shrinkage factors, which lead to obtain the sample LMMSE. This was noted in (3.29) and also in section 3.2. This behavior will be more clear in the upcoming figures. Finally, it is worth remarking the comparison between the two proposed shrinkage methods. Figures 3.7 and 3.8 highlight that the shrinkage of the sample towards the matched filter obtains better performance than the more simple direct shrinkage of the sample LMMSE. The reason is that the former incorporates more a priori information about the problem or is more general. This can be easily observed from the structure of both filters. The shrinkage of the sample LMMSE has the form $\mathbf{w} = \alpha\hat{\mathbf{R}}^{-1}\mathbf{s}$, which is a particular case of the shrinkage of the sample LMMSE towards a matched filter $\mathbf{w} = (\alpha_1\hat{\mathbf{R}}^{-1} + \alpha_2\mathbf{I})\mathbf{s}$. Note that this latter filter incorporates more a priori information, because it combines the sample LMMSE with a scaling of a matched filter. Namely, on the one hand, the sample LMMSE relies on an estimation of the inverse covariance obtained from the available samples, i.e. the inverse of the SCM. On the other hand, the scaling of the matched filter can be interpreted as an initial guess of the LMMSE where one assumes that only AWGN and the signal of interest are present in the scenario, thereby it ignores the available statistical samples.

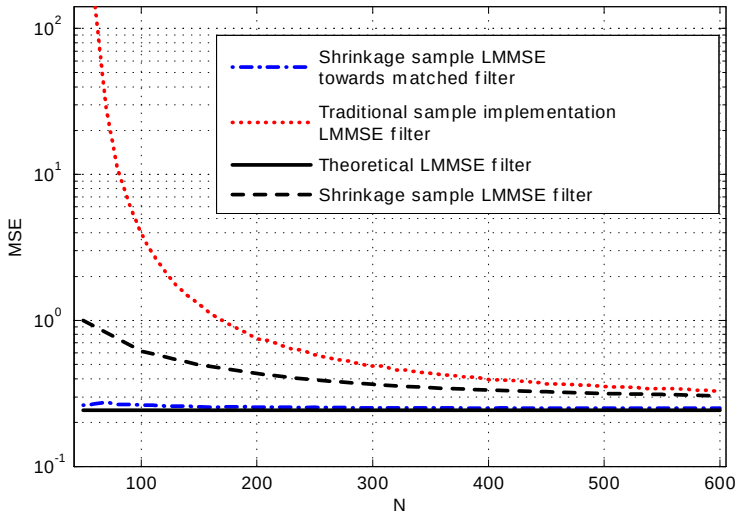

Figure 3.8: Performance comparison between the shrinkage of the sample LMMSE (3.3), shrinkage of the sample LMMSE towards a matched filter (3.28), the LMMSE (1.4) and the sample LMMSE (1.9), when SNR=5 dB, SIR_i=0 dB and M=50.

Next, in figure 3.9 the shrinkage effect is exemplified. Namely, in figure 3.9 we run a Monte Carlo simulation to plot $|\check{\alpha}_{lb,1}|^2$ and $|\check{\alpha}_{lb,2}|^2$ of the proposed shrinkage method in

(3.28). Recall that the proposed shrinkage filter reads $\check{\mathbf{w}}_{lb,s}=\check{\alpha}_{lb,1}\hat{\mathbf{R}}^{-1}\mathbf{s}+\check{\alpha}_{lb,2}\mathbf{s}$. 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, $\check{\mathbf{w}}_{lb,s}$ tends to give more weight to the sample LMMSE than to the matched filter. Indeed when $N \gg M$ the proposed filter $\check{\mathbf{w}}_{lb,s}$ tends to disregard the matched 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 (3.29). And effectively, figure 3.9 highlights this behavior, as $\frac{M}{N}$ decreases $|\check{\alpha}_{lb,1}|^2$ tends to increase whereas $|\check{\alpha}_{lb,2}|^2$ tends to decrease. On the other hand, as in general in the small sample size regime the matched filter yields better performance than the sample LMMSE, $\check{\mathbf{w}}_{lb,s}$ has the next behavior. As $\frac{M}{N}$ increases, $\check{\mathbf{w}}_{lb,s}$ tends to give more weight to the matched filter than to the sample LMMSE. Indeed in the extreme case where $\frac{M}{N}$ is close to 1, the proposed filter $\check{\mathbf{w}}_{lb,s}$ tends to disregard the sample LMMSE and give most of the weight to the matched filter, see (3.30). And effectively figure 3.9 highlights this behavior as well. Namely, As $\frac{M}{N}$ increases, $|\check{\alpha}_{lb,2}|^2$ tends to increase whereas $|\check{\alpha}_{lb,1}|^2$ tends to decrease.

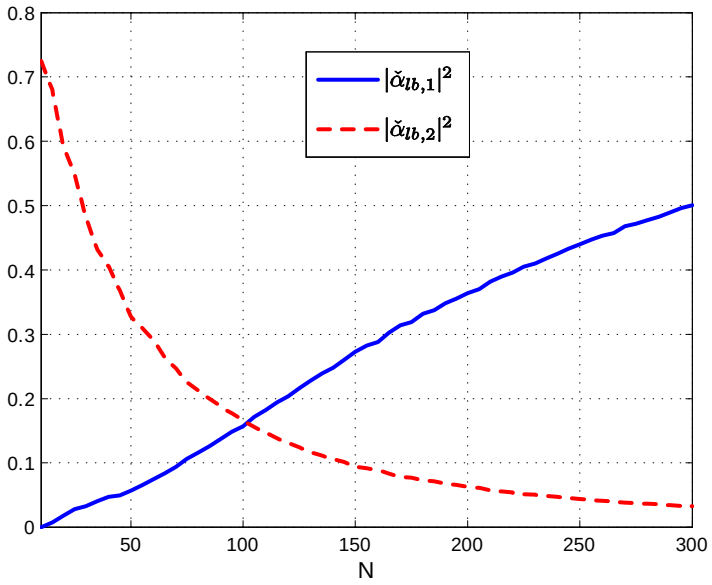

Figure 3.9: Shrinkage factors of the proposed shrinkage of the sample LMMSE towards mathced filter in (3.28) when $M = 10$, SNR=5 dB and SIR_i=0 dB.

Next, in figure 3.10 the performance of the proposed shrinkage of the sample LMMSE methods in (3.3) and (3.28) is compared to the LW and ad-hoc DL implementations of the

LMMSE, see (1.10) and (1.16), which are robust to the small sample size as well. M is set to 30 and $N \in (30, 600)$. In these plots we consider the DL factor proposed in [77] based on analyzing the estimation error of the covariance, i.e. δ is chosen to be equal to the standard deviation of the diagonal entries of the SCM, see section 1.4.1. Figure 3.10 shows that the proposed shrinkage of the sample LMMSE towards a matched filter (3.28) outperforms the DL and LW implementations of the LMMSE provided that $M < N$, indeed in some cases the gain is significant. The reason for this improvement is as follows.

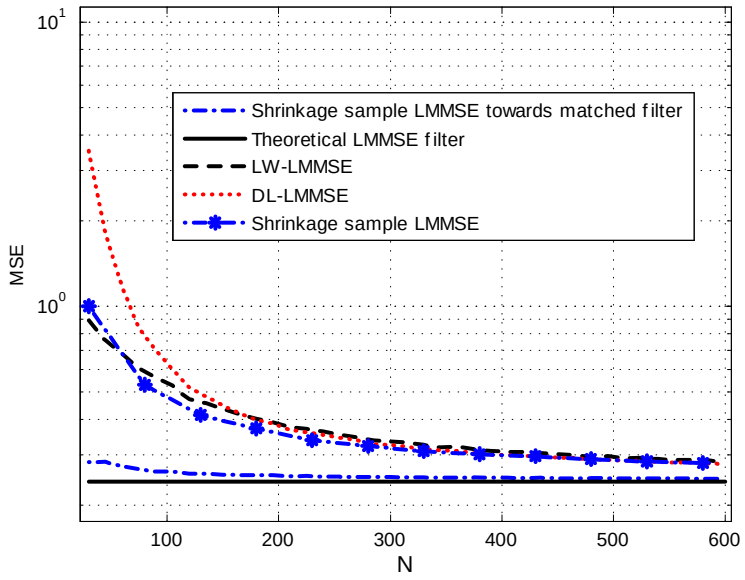

Figure 3.10: Performance comparison between the proposed shrinkage of the sample LMMSE towards a mathced filter (3.28), the shrinkage sample LMMSE (3.3), the LMMSE (1.4), the LW-LMMSE (1.16) and the DL-LMMSE (1.10), implemented with δ equal to the standard deviation of the diagonal entries of $\hat{\mathbf{R}}$. SNR=5 dB, SIR_i=0 dB and M=30.

On the one hand, both DL and LW implementations seek to enhance the covariance estimate, but they do not deal directly with the estimation of the parameter of interest. DL regularizes the SCM by analyzing the error bounds in the estimation of the covariance, whereas LW proposes a shrinkage of the SCM that seeks to optimize asymptotically the MSE in the estimation of the covariance. On the other hand, the method suggested herein faces directly the estimation of the parameter of interest x by obtaining an $(M, N)-$consistent estimate of the MSE optimal, though unrealizable, estimator of x

in lemma 3.5. Actually the same rationale can be applied to explain why the more basic shrinkage of the sample LMMSE (3.3) also obtains performance gains compared to the ad-hoc DL.

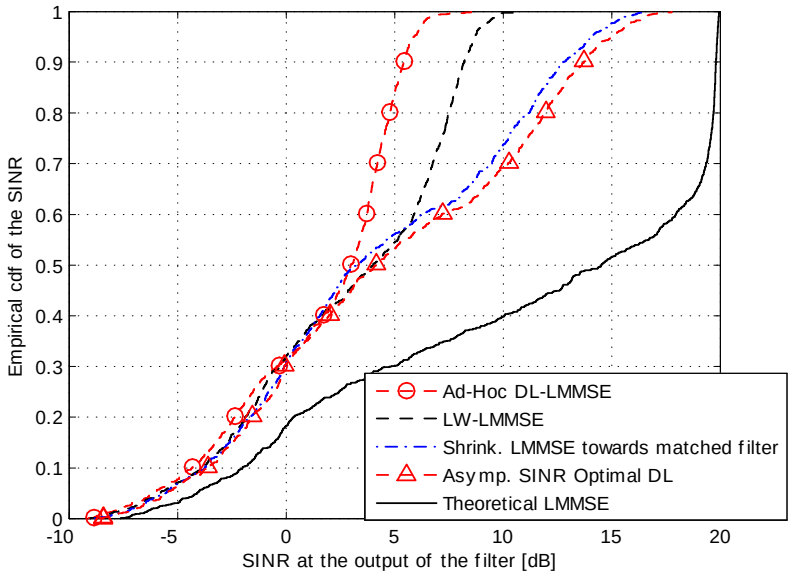

Figure 3.11: Empirical cdf of the SINR at the output of the next beamformers. The proposed shrinkage of the sample LMMSE towards a matched filter (3.28), the theoretical LMMSE (1.4), the LW-LMMSE (1.16), the ad-hoc DL-LMMSE implemented with δ equal to the standard deviation of the diagonal entries of $\hat{\mathbf{R}}$ (1.13) and the DL-LMMSE implemented with the asymptotically SINR optimal DL factor (1.15) in [9]. SNR=20 dB, SIR_i=0 dB, M=50 and N=70.

Figure 3.11 assesses the SINR achieved at the output of the proposed shrinkage filter in (3.28) as well as other related work methods. As it was commented above the SINR is an important metric widely used in beamforming an other signal processing and communications applications. Thereby it is interesting to give insights on the performance of the proposed shrinkage filter in (3.28) in terms of SINR. To this end, the proposed method is compared to the DL method in [9]. This is an important state-of-the-art method, as it obtains, for the type of diagonally loaded filters, the DL factor which optimizes asymptotically the SINR at the output of the filter, see section 1.4.1 for further details. In order to complete the comparison the ad-hoc DL method based on the DL factor (1.13) and the LW-LMMSE (1.16) are incorporated in the simulation as well.

The simulation conditions are similar than the ones of the previous figures, but taking into account the values in [9] to allow a fair comparison for the DL method in [9]. Thereby, in figure 3.11 $M = 50$ and $N = 70$, which corresponds to a small sample size regime. Moreover, 29 interferers are considered, whose power is the same than the one of the signal of interest, and the SNR=20 dB. At each iteration of the simulation the DOA of the interferers and the SOI are generated according to a uniform distribution within the range $[-90°, 90°]$. This permits to allow variability in the scenario and permits to generalize the results of the previous figures, which considered fixed values of DOA for both the interferers and the SOI.

Given these simulation conditions, figure 3.11 displays the empirical cdf of the SINR at the output of the filters. It can be observed that the proposed shrinkage outperforms both the ad-hoc DL and the LW method. This corroborates the result obtained in figure 3.10 for a more general scenario in terms of DOA values, as figure 3.11 considers different DOA values at each iteration of the simulation, whereas in figure 3.10 they were fixed. The proposed shrinkage LMMSE method obtains better performance than the ad-hoc DL and the LW algorithms because it deals directly with the estimation of the parameter of interest. That is, it obtains the shrinkage factors which optimize asymptotically the MSE in the estimation of the parameter of interest. On the contrary, as it was commented in the previous figure, the ad-hoc DL and the LW methods try to obtain a better estimation of the covariance, than the SCM, by analyzing or optimizing metrics related to the covariance, which is not the final target herein.

Figure 3.11 highlights that the proposed shrinkage LMMSE obtains a performance which is close to the one of the asymptotically optimal DL method proposed in [9]. Moreover, in chapter 5 it is shown that the proposed shrinkage LMMSE (3.28) outperforms the DL method in [9], in terms of SINR, when there is an uncertainty in the steering vector of the SOI. The rationale is that the DL method in [9] obtains an asymptotically optimal DL factor to cope with the small sample size regime. However, if there is an uncertainty in the steering vector, an additional tuning of that DL factor is required. Otherwise the DL method undergoes a performance degradation due to the signal cancelation effect, where the filter may confuse the SOI with an interference and may tend to cancel it. On the contrary, the proposed shrinkage LMMSE filter in the small sample size regime tends to give more weight to the matched filter and to disregard the sample LMMSE. Thereby, it circumvents the degradation due to the signal cancelation effect when there is an uncertainty in the steering vector. That is, it may attenuate the SOI due to e.g. pointing errors, but it will not try to cancel the SOI as it was an interference.

In figure 3.12, in order to complement the previous figure, the same type of simulation is carried out, but in this case the MSE is considered as the performance metric, namely the

empirical cdf of the MSE related to each filter is displayed. This is a more fair comparison for the proposed shrinkage LMMSE in (3.28) as it obtains the shrinkage factors as the ones that optimize asymptotically the MSE in the estimation of the parameter of interest. Moreover, as it was comented above, the MSE is an important metric in applications where the complex amplitude of the signal of interest is important, e.g. in subband beamforming [17] see further details in section 3.2. Figure 3.12 shows that the proposed shrinkage LMMSE in (3.28) clearly outperforms the DL method proposed in [9]. This is because the proposed method designs the shrinkage factors to minimize the MSE, whereas the DL method in [9] obtains a DL factor which optimizes asymptotically the SINR, which does not guarantee a good estimation performance in terms of MSE, as it was pointed out in [17].

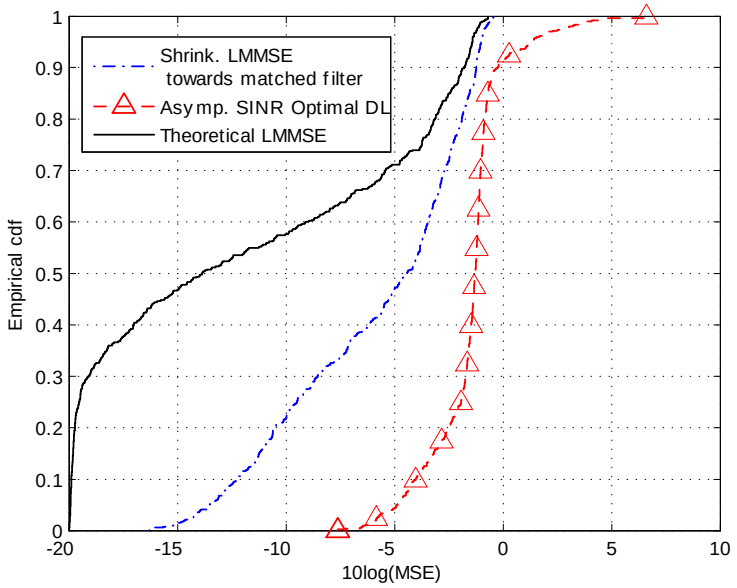

Figure 3.12: Empirical cdf of the MSE associated to the next beamformers. The proposed shrinkage of the sample LMMSE towards a matched filter (3.28), the theoretical LMMSE (1.4) and the DL-LMMSE implemented with the asymptotically SINR optimal DL factor (1.15) in [9]. SNR=20 dB, SIR_i=0 dB and.M=50, N=70.

Figures 3.13 to 3.15 assess the performance of the shrinkage of the regularized sample LMMSE towards a matched filter proposed in (3.38). Recall that this filter can be viewed as a generalization of the shrinkage filter proposed in (3.28) to support the scenarios where M is higher than N. Thereby it is the more elaborate filter proposed in this chapter. In

order to study the performance of the proposed filter similar type of simulations than in the previous figures are used. That is, the DOA of the SOI and the interferences are generated according to a uniform distribution within the range $[-90°, 90°]$ at each iteration and then the cdf of the MSE and the SINR achieved by the filters are displayed. In the next plots the SOI and the interferences have the same power and the SNR is set to 20 dB. The shrinkage of the regularized LMMSE towards a matched filter proposed in (3.38) is compared to the next methods. The theoretical LMMSE in (1.4), which is the lower bound in terms of MSE or the upper bound in terms of SINR. The method proposed in [9], which obtains the asymptotically SINR optimal DL factor δ in (1.15) for the type of DL-LMMSE filters given by $\mathbf{w} = (\hat{\mathbf{R}} + \delta\mathbf{I})^{-1}\mathbf{s}$. Finally for comparison purposes, the method proposed in [79] is also considered, as it obtains the asymptotically MSE optimal shrinkage factors τ_1, τ_2 which characterize the type of regularized LMMSE filters given by $\mathbf{w} = (\tau_1\hat{\mathbf{R}} + \tau_2\mathbf{I})^{-1}\mathbf{s}$.

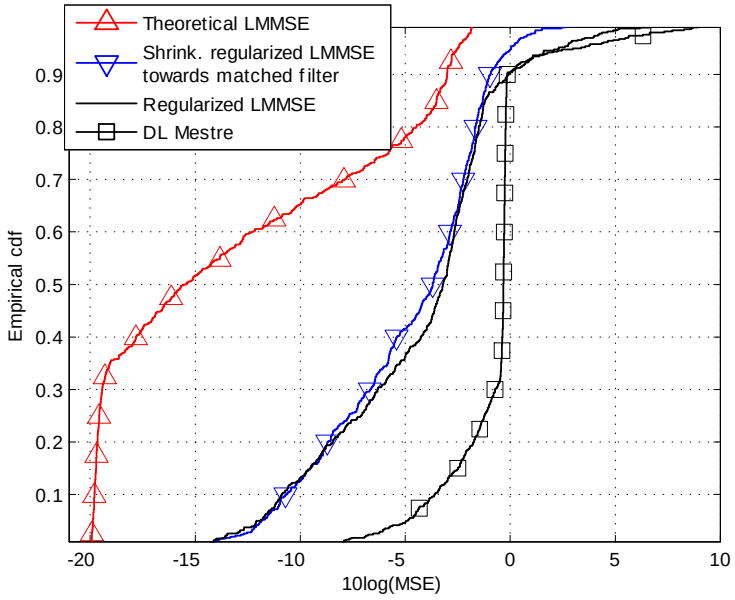

Figure 3.13: Empirical cdf of the MSE associated to the next beamformers. The proposed shrinkage of the regularized sample LMMSE towards a matched filter (3.38), the theoretical LMMSE (1.4), the DL-LMMSE implemented with the asymptotically SINR optimal DL factor (1.15) proposed in [9] and the regularized LMMSE in [79], which obtains the asymptotically MSE optimal shrinkage of the SCM. SNR=20 dB, SIR_i=0 dB, M=20 and N=19.

Thereby, first in figures 3.13 and 3.14 the cdf of the MSE is taken into account as

a performance metric. Namely, in figure 3.13 $M = 20$, $N = 19$ and the number of interferences plus the SOI, i.e. K, is set to 12, whereas in figure 3.14 $M = 10$, $N = 9$ and K is set to 6. In both figures it can be observed that the proposed shrinkage of the regularized sample LMMSE towards a matched filter clearly outperforms the DL-LMMSE method proposed in [9].

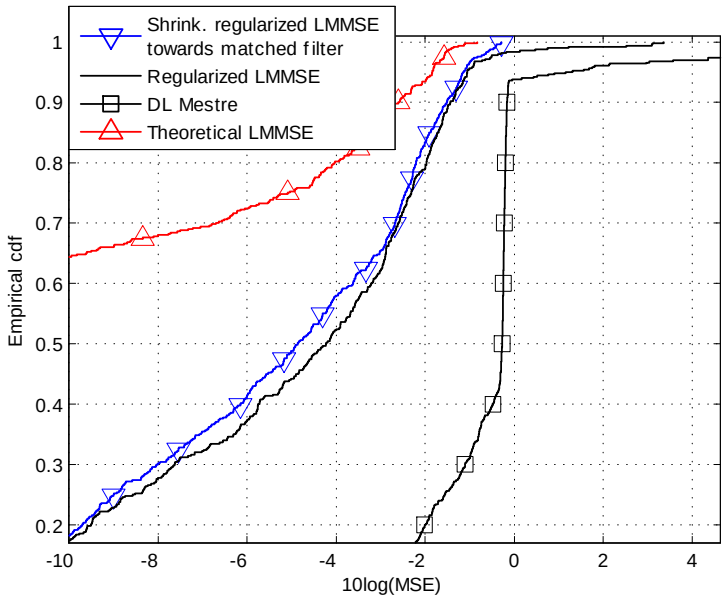

Figure 3.14: Empirical cdf of the MSE associated to the next beamformers. The proposed shrinkage of the regularized sample LMMSE towards a matched filter (3.38), the theoretical LMMSE (1.4), the DL-LMMSE implemented with the asymptotically SINR optimal DL factor (1.15) proposed in [9] and the regularized LMMSE in [79], which obtains the asymptotically MSE optimal shrinkage of the SCM. SNR=20 dB, SIR_i=0 dB, M=10 and N=9.

The rationale is that the proposed method seeks to optimize the MSE whereas the DL-LMMSE method proposed in [9] focuses on the optimization of the SINR. Moreover, figures 3.13 and 3.14 show that the proposed shrinkage in (3.38) obtains some performance gains compared to the regularized LMMSE proposed in [79], which also seeks to optimize the MSE. The rationale relies on the structure of the filters, as the proposed filter is more general than the regularized LMMSE in [79] from the next point of view. Namely, the proposed filter in (3.38) has the structure $\mathbf{w} = (\alpha_1(\hat{\mathbf{R}} + \delta\mathbf{I})^{-1} + \alpha_2\mathbf{I})\mathbf{s}$, thereby it considers the shrinkage of the inverse sample covariance, which is properly regularized to permit to

deal with situations where $M > N$. On the other hand [79] only considers the regularization of the SCM and it does not consider the shrinkage of the inverse sample covariance matrix.

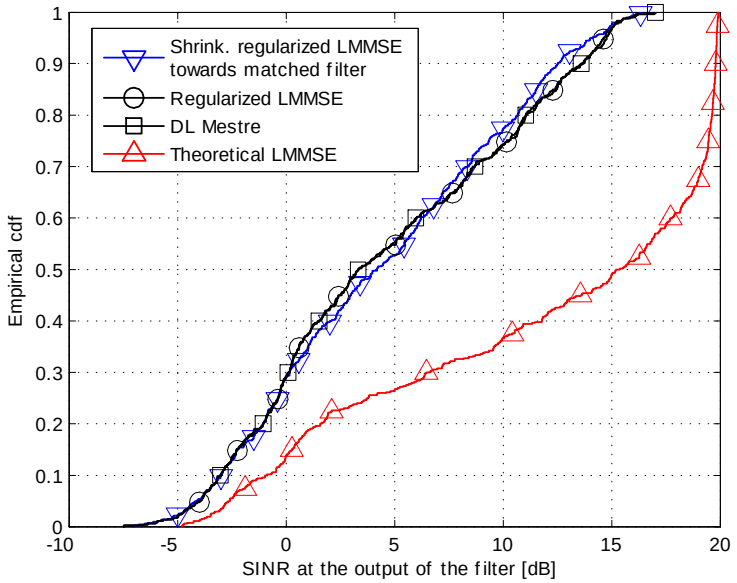

Figure 3.15: Empirical cdf of the SINR at the output of the next beamformers. The proposed shrinkage of the regularized sample LMMSE towards a matched filter (3.38), the theoretical LMMSE (1.4), the DL-LMMSE implemented with the asymptotically SINR optimal DL factor (1.15) proposed in [9] and the regularized LMMSE in [79], which obtains the asymptotically MSE optimal shrinkage of the SCM. SNR=20 dB, SIR_i=0 dB, M=20 and N=19.

Next, in figure 3.15 the same type of simulation than in figure 3.13 is considered, though in this case the cdf of the SINR is displayed as a performance metric instead of the cdf of the MSE. It can be observed that in this case the proposed method and the other robust methods display almost the same performance. In this regard, more insights are given in chapter 5 for the situations where the signature vector of the SOI is not known precisely. Namely, in those situations, it will be shown in chapter 5 that the shrinkage of the regularized LMMSE towards a matched filter proposed in (3.38) clearly outperforms both the DL-LMMSE proposed in [9] and the regularized LMMSE proposed in [79].

3.8 Appendix: proofs

Proof of Lemma 3.1

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

$$\alpha_l{=}\arg\min_{\alpha}\ \mathrm{MSE}\left(\mathbf{w}=\boldsymbol{\psi}\alpha\right) = \arg\min_{\alpha}\ \mathbb{E}\left[\left|x(n)-(\boldsymbol{\psi}\alpha)^H\mathbf{y}(n)\right|^2 \mid \hat{\mathbf{R}}\right]$$

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,

$$\alpha_l{=}\arg\min_{\alpha}\ \alpha^*\boldsymbol{\psi}^H\mathbf{R}\boldsymbol{\psi}\alpha{+}\gamma(1{-}\ \alpha^*\boldsymbol{\psi}^H\mathbf{s}-\mathbf{s}^H\boldsymbol{\psi}\alpha) \tag{3.42}$$

At this point, the optimal solution is found by setting to zero the derivative of the argument in (3.42) with respect to α^*, i.e. it is the solution to $\frac{\partial\,\mathrm{MSE}(\alpha)}{\partial\alpha^*}=0$, which yields,

$$\alpha_l = \left(\boldsymbol{\psi}^H\mathbf{R}\boldsymbol{\psi}\right)^{-1}\gamma\boldsymbol{\psi}^H\mathbf{s} \tag{3.43}$$

Now, recalling that $\boldsymbol{\psi}=\hat{\mathbf{R}}^{-1}\mathbf{s}$, (3.43) leads to the next expression,

$$\alpha_l = \frac{\gamma\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}} \tag{3.44}$$

Which concludes the proof as (3.44) coincides with (3.2).

Proof of Theorem 3.1

The statement that the estimator is realizable is clear from the expression of its shrinkage factors (3.3), as they do not depend on any unknown parameter. With regard to the consistency, it suffices to proof that the shrinkage factor $\check{\alpha}_l$ in (3.3) is a consistent estimate of the theoretical factors α_l in (3.2), i.e. we have to demonstrate that $\check{\alpha}_l \asymp \alpha_l$ within the general asymptotics framework where $M, N \to \infty$ at a constant rate $M/N \to c \in (0,1)$, where recall that $\asymp$ denotes almost sure convergence. To this end, let use the asymptotic equivalences presented in Lemma 2.1, in the theoretical LMMSE shrinkage vector (3.2). Namely, the results $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp (1-c)^{-3}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}$ and $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} \asymp (1-c)^{-1}\mathbf{s}^H\mathbf{R}^{-1}\mathbf{s}$ are considered. This yields the next asymptotic deterministic equivalent expression for α_l,

$$\alpha_l \asymp \gamma(1-c)^2 \tag{3.45}$$

Which concludes the proof as according to (3.3) $\check{\alpha}_l = \gamma(1-c)^2$ and as a consequence $\alpha_l \asymp \check{\alpha}_l$.

Proof of Lemma 3.3

Let us define $\delta = \beta_2/\beta_1$, then one can express the numerator and the denominator in (3.21) as $\beta_1^{-1}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$ and $\beta_1^{-2}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{R}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$, respectively. The convergence of these expressions was obtained in [9, appendix I] in terms of the eigenvectors, $\mathbf{e}_i$, and eigenvalues, λ_i, of $\mathbf{R}$. A sketch of the proof for the convergence of $\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{R}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$ is also found in the appendix of chapter 2. Thereby, in [9, appendix I] it was shown that considering the asymptotic regime where $M, N \to \infty$ and $M/N \to c \in (0, \infty)$, the desired quantities converge in probability to the next deterministic expressions,

$$\beta_1^{-1}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \asymp \beta_1^{-1}\sum_{i=1}^{M}\frac{(1+cb)\left|\mathbf{s}^H\mathbf{e}_i\right|^2}{\lambda_i+\rho}$$

$$\beta_1^{-2}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{R}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \asymp \beta_1^{-2}((1+cb)^2+cb')\sum_{i=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_i\right|^2\lambda_i}{(\lambda_i+\rho)^2}$$

where $\rho \triangleq \delta(1+cb)$, $b \triangleq b(z)\mid_{z=0}$ is the positive solution to the next equation, $b' = \frac{db(z)}{dz}\mid_{z=0}$ is defined next and $b(z)$ is defined in [9, eq. 25],

$$b = \frac{1}{M}\sum_{i=1}^{M}\frac{\lambda_i(1+cb)}{\lambda_i+\rho}$$

$$b' = \left(1-\frac{1}{M}\sum_{i=1}^{M}\frac{c\lambda_i^2}{(\lambda_i+\rho)^2}\right)^{-1}\frac{1}{M}\sum_{i=1}^{M}\frac{\lambda_i^2(1+cb)^2}{(\lambda_i+\rho)^2}.$$

Finally, bearing in mind the next equalities,

$$\sum_{i=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_i\right|^2}{\lambda_i+\rho} = \mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$$

$$\sum_{i=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_i\right|^2\lambda_i}{(\lambda_i+\rho)^2} = \mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$$

one obtains that the numerator and the denominator in (3.21) convergence to the next expressions,

$$\beta_1^{-1}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \asymp \beta_1^{-1}(1+cb)\mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$$
$$\beta_1^{-2}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{R}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} \asymp \beta_1^{-2}((1+cb)^2+cb')$$
$$\cdot\mathbf{s}^H(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{R}(\mathbf{R}+\rho\mathbf{I})^{-1}\mathbf{s}$$

Substituting these expressions in (3.21), and after some manipulations, one obtains (3.22).

Proof of Lemma 3.4

The consistent estimates of these parameters is given in [124, App.A] and it is based on expressing b, ξ, η_n and η_d in terms of the next real Stieltjes transforms $t(x)$ and $s(x)$,

$$\xi = 1 - s(x)\,\mid_{x=\rho^{-1}} + \rho^{-1}\frac{ds(x)}{dx}\mid_{x=\rho^{-1}}$$
$$b/(1+cb) = 1 - s(\delta^{-1}(1+cb)^{-1})$$
$$\eta_d = \rho^{-1}t(\rho^{-1}), \quad \eta_n = -[x^2 dt(x)/dx]\,\mid_{x=\rho^{-1}}$$

$$t(x) = \sum_{k=1}^{M}\frac{\left|\mathbf{s}^H\mathbf{e}_k\right|^2}{1+x\lambda_k}, \;\; s(x) = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{1+x\lambda_k}, \;\; x>0$$

The (M,N)-consistent estimators of $t(x)$ and $s(x)$ were obtained in [101], see [124, App.A], considering the asymptotic regime where $M,N\to\infty$ and $M/N\to c\in(0,\infty)$, and are given by

$$\hat{t}(x) = \sum_{k=1}^{M}\frac{\left|\mathbf{s}^H\hat{\mathbf{e}}_k\right|^2}{1+\theta(x)\hat{\lambda}_k}, \hat{s}(x) = \frac{1}{M}\sum_{k=1}^{M}\frac{1}{1+\theta(x)\hat{\lambda}_k}$$

where $\theta(x)$ is the positive solution to the next equation,

$$\theta(x)[1-c+c\frac{1}{M}\,\mathrm{Tr}[(\mathbf{I}+\theta(x)\hat{\mathbf{R}})^{-1}]] = x, x>0$$

Thereby, this paves the way to obtain the next (M,N)-consistent estimates of b, ξ, η_n and η_d,

$$\hat{b} = \frac{1 - \frac{\delta}{M}\,\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]}{1 - c(1 - \frac{\delta}{M}\,\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}])}$$

$$\hat{\eta}_d = (1 - c + c\tfrac{\delta}{M}\,\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}])\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$$

$$\hat{\xi} = \frac{\frac{1}{M}\,\mathrm{Tr}[\hat{\mathbf{R}}^2(\hat{\mathbf{R}}+\delta\mathbf{I})^{-2}] - \frac{c}{M^2}\,\mathrm{Tr}^2[\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]}{1 - c + c\delta^2\frac{1}{M}\,\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-2}]}$$

$$\hat{\eta}_n = \frac{(1 - c(1 - \frac{\delta}{M}\,\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}]))^2}{1 - c + \frac{c}{M}\,\mathrm{Tr}[(\delta^{-1}\hat{\mathbf{R}} + \mathbf{I})^{-2}]}\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s}$$

Proof of Lemma 3.5

Let us define $\boldsymbol{\alpha} \triangleq (\alpha_1\ ,\ \alpha_2)^T$ and $\boldsymbol{\Omega} \triangleq \left(\hat{\mathbf{R}}^{-1}\mathbf{s}\ ,\ \mathbf{s}\right)$ and let $\boldsymbol{\alpha}_{lb}$ be the vector of optimal shrinkage factors, then the MSE optimization problem in (3.26) may be reformulated as,

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

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,

$$\boldsymbol{\alpha}_{lb}{=}\arg\min_{\boldsymbol{\alpha}}\ \boldsymbol{\alpha}^H\boldsymbol{\Omega}^H\mathbf{R}\boldsymbol{\Omega}\boldsymbol{\alpha}{+}\gamma(1{-}\boldsymbol{\alpha}^H\boldsymbol{\Omega}^H\mathbf{s} - \mathbf{s}^H\boldsymbol{\Omega}\boldsymbol{\alpha}) \tag{3.46}$$

Observe that in (3.46) 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, [14]. Indeed, the optimization problem in (3.46) 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,

$$\boldsymbol{\alpha}_{lb} = \left(\boldsymbol{\Omega}^H\mathbf{R}\boldsymbol{\Omega}\right)^{-1}\gamma\boldsymbol{\Omega}^H\mathbf{s} \tag{3.47}$$

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 [14], the expression (3.47) yields,

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

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,

$$\boldsymbol{\alpha}_{lb} = \frac{\gamma \begin{pmatrix} \mathbf{s}^H\mathbf{R}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \\ \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} \end{pmatrix}}{\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}\mathbf{s}^H\mathbf{R}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}\mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}} \tag{3.48}$$

which concludes the proof as (3.48) coincides with (3.27).

Proof of Theorem 3.5

The proof for Theorem 3.5 is readily obtained from Lemma 3.5 and Lemma 2.1. Namely, the claim that the estimator is realizable is evident from the expression of its shrinkage factors in (3.28). With regard to the consistency of the estimator it suffices to prove that the shrinkage factor $\check{\boldsymbol{\alpha}}_{lb}$ in (3.28) is a consistent estimate of the optimal shrinkage vector $\boldsymbol{\alpha}_{lb}$ in (3.27), i.e. $\check{\boldsymbol{\alpha}}_{lb} \asymp \boldsymbol{\alpha}_{lb}$ within the general asymptotics framework where $M, N \to \infty$ at a constant rate $M/N \to c \in (0, 1)$.

In order to prove that $\check{\boldsymbol{\alpha}}_{lb} \asymp \boldsymbol{\alpha}_{lb}$, let us use the asymptotic equivalences presented in Lemma 2.1, in the theoretical LMMSE shrinkage vector (3.27). This yields the next asymptotic equivalent expression for $\check{\boldsymbol{\alpha}}_{lb}$,

$$\boldsymbol{\alpha}_{lb} \asymp \frac{\gamma \begin{pmatrix} \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} - (1-c)^{-1} \\ (1-c)^{-2}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} - \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(1-c)^{-1} \end{pmatrix}}{(1-c)^{-2}\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} - (1-c)^{-2}} \tag{3.49}$$

And, after straightforward manipulations, expression (3.49) may be rewritten as follows,

$$\boldsymbol{\alpha}_{lb} \asymp \frac{\gamma \begin{pmatrix} (1-c)^2\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} - (1-c) \\ c\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} \end{pmatrix}}{\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} - 1} = \check{\boldsymbol{\alpha}}_{lb}. \tag{3.50}$$

This concludes the proof, as it highlights that the shrinkage factor in (3.28) is a consistent estimate of the optimal shrinkage factor in (3.27).

Chapter 4

Shrinkage of the sample MVDR method to deal with the small sample size degradation

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 estimators which are robust to the small sample size regime and that preserve 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 shrinkage target consisting of a matched filter or conventional beamformer. In other words a linear combination of the sample MVDR and the conventional beamformer. As it has been mentioned in the previous chapters, this structure implements a correction of the sample MVDR filter in the small sample size regime, whose aim is to diminish the overall estimation error by optimizing the bias variance tradeoff. The proposed filter combines the information obtained by the observations, which is embedded in the sample MVDR, with a priori information, which is represented by the conventional beamformer. The optimal shrinkage factors, i.e. the coefficients of the linear combination, are obtained as the ones which maximize the SINR associated to the proposed filter. Unfortunately, this optimal

shrinkage factors depend 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 factors 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. That is, 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. In section 4.5, the shrinkage method proposed in sections 4.3 is generalized to support small sample size situations where $M > N$. To this end, a regularization of the SCM is considered in the shrinkage of the sample MVDR towards the matched filter. That is the type of filter $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, where $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I}$, is proposed. The design of the shrinkage factors α_1, α_2 and the regularization parameter δ relies on optimizing the SINR at the output of the filter and then applying RMT tools to obtain (M, N)-consistent estimates of the uknown terms within the regime where $(M, N) \to \infty$ and $M/N \to (0, \infty)$. Provided that $M/N \in (0, 1)$, the numerical simulations highlight that the first shrinkage method proposed in section 4.3 outperforms the conventional sample MVDR method and also other robust techniques to the small sample size regime explained in section 1.4.1. Namely, the ad-hoc implementations of the DL technique and the LW-MVDR method, which is based on shrinking the SCM by minimizing the MSE in the estimation of the data covariance. In addition it displays almost the same performance than the DL technique [9] which asymptotically optimizes the SINR, which is also exposed in section 1.4.1. Moreover, regarding the more general shrinkage method proposed in section 4.5, the numerical simulations highlight that it permits to extend the contributions of the shrinkage method in 4.3 to $M/N \in (0, \infty)$. That is, on the one hand it clearly outperforms the LW-MVDR method, which had better performance than the ad-hoc implementations of the DL technique in the previous simulations. And on the other hand, the shrinkage method proposed in section 4.5 obtains the same performance than the asymptotically optimal DL technique proposed in [9]. Moreover, as it is shown in chapter 5, the shrinkage MVDR methods proposed in sections 4.3 and 4.5 outperform the asymptotically optimal DL method proposed in [9], when there is an uncertainty in the signature of the SOI or steering vector $\mathbf{s}$ of the data model (1.1).

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 $\hat{\mathbf{R}}^{-1}$. This approach relies 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. [33], this means that when N is comparable to M and $M \leqslant N$ inverting $\hat{\mathbf{R}}$ dramatically amplifies the estimation error. Moreover, when $M > N$ the sample correlation matrix is not even invertible. On the other hand, when $N \gg M$ the sample MVDR is optimal as $\hat{\mathbf{R}}$ is the MVUE estimator of $\mathbf{R}$. In order to overcome the small sample size degradation, the first proposed estimator is the next shrinkage of the sample MVDR filter towards a matched filter,

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

A more general form of this estimator is tackled in section 4.5. 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. The rationale for the proposed filter is as follows. 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 linear combination coefficients α_1, α_2 and the term $\mathbf{s}$. That is, 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 $\mathbf{w} = \mathbf{s}$. This latter filter can be regarded as an initial guess of the MVDR where one assumes that only the SOI and additive white noise are present in the scenario, i.e. one disregards the available statistical samples and estimates the unknown noise covariance within $\mathbf{R}$ by just $\mathbf{I}$. After applying the Woodbury's identity this leads to a scaling of a matched filter. Another interesting interpretation of the proposed filter is obtained by rewriting the expression of the proposed filter, after applying the distributive property of matrices, as

$$\mathbf{w} = (\alpha_1\hat{\mathbf{R}}^{-1} + \alpha_2\mathbf{I})\mathbf{s} \triangleq \check{\mathbf{R}}^{-1}\mathbf{s}.$$

That is, the proposed method is implementing a correction of the sample MVDR which relies on a shrinkage of the inverse of the SCM, $\check{\mathbf{R}}^{-1} = \alpha_1\hat{\mathbf{R}}^{-1} + \alpha_2\mathbf{I}$ which is a better

estimate of $\mathbf{R}^{-1}$ than $\hat{\mathbf{R}}^{-1}$. Or even more precisely, consider the eigendecomposition of the SCM $\hat{\mathbf{R}}^{-1} = \hat{\mathbf{E}}\hat{\mathbf{\Lambda}}^{-1}\hat{\mathbf{E}}^H$, where $\hat{\mathbf{E}}$ is a matrix which stacks in its columns the eigenvectors of the SCM and $\hat{\mathbf{\Lambda}}$ is a diagonal matrix which contains the eigenvalues of the SCM in its main diagonal. Then, it is clear that the proposed filter can be expressed as $\mathbf{w} = \hat{\mathbf{E}}(\alpha_1\hat{\mathbf{\Lambda}}^{-1} + \alpha_2\mathbf{I})\hat{\mathbf{E}}^H\mathbf{s}$. Therefore the proposed filter is implementing a correction of the sample MVDR which consists of a shrinkage of the eigenvalues of $\hat{\mathbf{R}}^{-1}$.

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) in terms of the optimization of the SINR at the output of the filter. That is, the aim is to solve the problem stated in (1.29), when $f(\text{MSE}(\mathbf{w})) = \text{MSE}(\mathbf{w})$ and when the constraint $\mathbf{w}^H\mathbf{s} = 1$ is imposed, which is equivalent to optimize the SINR at the output of the filter. This is more clear if one observes the expressions of the MSE and the SINR

$$\begin{aligned}\text{MSE}(\mathbf{w}) &= \mathbf{w}^H\mathbf{R}_n\mathbf{w} + \gamma|\mathbf{w}^H\mathbf{s} - 1|^2 \\ \text{SINR}(\mathbf{w}) &= \frac{\gamma|\mathbf{w}^H\mathbf{s}|^2}{\mathbf{w}^H\mathbf{R}_n\mathbf{w}}.\end{aligned}$$

Moreover, 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 matched 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, which is equivalent to maximize the SINR at the output of $\mathbf{w}$ and it permits to avoid the knowledge about the second moment of $x(n)$. This problem is mathematically formulated as follows,*

$$\begin{aligned}\hat{x}_{c,s}(n) = \mathbf{w}_{c,s}^H\mathbf{y}(n); \quad \mathbf{w}_{c,s} = \arg\min_{\mathbf{w}} \quad & \text{MSE}(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{w}^H\mathbf{s} = 1, \mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2\mathbf{s}\end{aligned} \tag{4.1}$$

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. That is, $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}$ 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 [130]. 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, 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. That is, 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 Asymptotically optimal shrinkage of the 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 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,*

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

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

Proof: See section 4.4.

■

This *(M,N)*-consistent estimator, shows explicitly 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. Namely, recall that as it is explained above, in the previous section, the proposed shrinkage structure implements a correction in the form of a bias variance tradeoff that permits to diminish the overall MSE in the estimation of the SOI. Also another interpretation is that a shrinkage of the inverse of the SCM, i.e. $\check{\alpha}_{c,1}\hat{\mathbf{R}}^{-1} + \check{\alpha}_{c,2}\mathbf{I}$, is implemented by the proposed filter which is a better estimate than $\hat{\mathbf{R}}^{-1}$, in the small sample size regime. Moreover, in this regard, note the shrinkage coefficients $\check{\alpha}_{c,1}\ ,\ \check{\alpha}_{c,2}$ are designed to maximize the SINR at the output of the filter. On the other hand, as we will see in the proof of this theorem, the proposed filter relies on random matrix theory, i.e. the consistency is obtained within the framework of general asymptotics, that embraces the small sample size scenario where N can be comparable to M. 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 is close to 1. Moreover, the simulations also highlight that the proposed method outperforms robust techniques to the small sample size regime such as the LW-MVDR (1.17), provided that $M/N \in (0,1)$, which obtains an asymptotically optimal shrinkage of the SCM by minimizing the MSE in the estimation of the data covariance.

In fact, note that this performance metric is not the final target. On the contrary the proposed method focus on obtaining a good estimate the SOI $x(n)$, which is the actual target. The simulations also highlight better performance of the proposed method than the ad-hoc implementations of the DL-MVDR (1.11) explained in section 1.4.1, provided that $M/N \in (0, 1)$. This is because the proposed approach is more analytical, as it deals directly with the optimization of a metric related to the estimation of the parameter of interest. Finally, it is worth mentioning that the proposed shrinkage MVDR displays almost the same performance than the DL technique [9] which optimizes asymptotically the SINR for the type of DL methods. In this regard, an important contribution, as it will be shown in chapter 5, is that the proposed shrinkage MVDR outperforms the DL method [9], in situations where the signature vector of the SOI $\mathbf{s}$ in (1.1) is not precisely known.

At this point, it is interesting to study the values of the shrinkage factors in (4.4) when c approaches its extremes values, i.e. when $c \to 1$ and $c \to 0$. This permits to get more insights on the effects of the shrinkage effect on the proposed filter. When $c \to 1$, after straightfordward manipulations of (4.4), the next expression is obtained,

$$c \to 1 \Rightarrow \check{\boldsymbol{\alpha}}_c \to (0, 1)^T. \tag{4.5}$$

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 matched 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 factors in (4.4) tend to the next expression,

$$c \to 0 \Rightarrow \check{\boldsymbol{\alpha}}_c \to \left(\frac{1}{\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}}, 0 \right)^T. \tag{4.6}$$

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. That is, 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 usually 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 factors give more weight to a filter built only from "a priori" information, i.e. a type of matched filter $\mathbf{w} \propto \mathbf{s}$. 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 logical that the shrinkage factors give more importance to the sample MVDR filter.

4.4 Proofs of Lemma 4.1 and Theorem 4.1

In this section, the proofs of Lemma 4.1 and Theorem 4.1 are provided. Recall that on the one hand the proof of Lemma 4.1 permits to obtain the optimal, though unrealizable, shrinkage factors α_1, α_2 for the type of proposed shrinkage MVDR filter $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$. On the other hand, Theorem 4.1 yields an $(M, N)-$consistent estimate of the shrinkage factors obtained in Lemma 4.1, i.e. it obtains the asymptotically optimal shrinkage factors.

Proof of Lemma 4.1

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

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

Next, the constraint $\mathbf{w}^H \mathbf{s} = 1$ is applied to reduce the objective function to $\mathbf{w}^H \mathbf{R}_{\mathbf{n}} \mathbf{w}$. This leads to the next optimization problem, which highlights that the shrinkage factors α_1, α_2 are designed to optimize the SINR at the output of the type of filters $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$,

$$\begin{aligned} \mathbf{w}_{c,s} = \arg\min_{\mathbf{w}} \quad & \mathbf{w}^H \mathbf{R}_{\mathbf{n}} \mathbf{w} \\ \text{s.t.} \quad & \mathbf{w}^H \mathbf{s} = 1, \mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s} \end{aligned}$$

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 $\mathbf{w}^H \mathbf{R} \mathbf{w}$. Namely note that, $\mathbf{w}^H \mathbf{R} \mathbf{w} = \mathbf{w}^H (\gamma \mathbf{s}\mathbf{s}^H + \mathbf{R}_{\mathbf{n}}) \mathbf{w}$ and due to the constraint $\mathbf{w}^H \mathbf{s} = 1$, this is equivalent to optimize the objective function $\gamma + \mathbf{w}^H \mathbf{R}_{\mathbf{n}} \mathbf{w}$, which on its turn is equivalent to optimize $\mathbf{w}^H \mathbf{R}_{\mathbf{n}} \mathbf{w}$. Therefore, (4.7) can be reformulated as,

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

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

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

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 [131], we readily obtain that the optimum shrinkage factors are,

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

At this point, applying to (4.10) the property of multiplication of partitioned matrices [14], 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,

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

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 matched 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 where $M, N \to \infty$ and $M/N \to c \in (0, 1)$. 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 the RMT results in Lemma 2.1 are considered in the theoretical MVDR shrinkage vector (4.2). Namely, the first equivalence that is considered is $\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s} \asymp \mathbf{s}^H \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} \asymp (1-c)^{-1}$, the second one is $\mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{R} \hat{\mathbf{R}}^{-1} \mathbf{s} \asymp (1-c)^{-3} \mathbf{s}^H \mathbf{R}^{-1} \mathbf{s} \asymp (1-c)^{-2} \mathbf{s}^H \hat{\mathbf{R}}^{-1} \mathbf{s}$ and the third equivalence is $\mathbf{s}^H \mathbf{R} \mathbf{s} \asymp \mathbf{s}^H \hat{\mathbf{R}} \mathbf{s}$. 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 $M/N \to c \in (0,1)$,

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

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

4.5 Shrinkage and regularization of the sample MVDR

In this section the shrinkage of the sample MVDR proposed in section 4.2 is generalized to the case where the observation dimension M can be higher than the sample size N. Those situations provoke that the sample covariance is not invertible. In order to cope with this problem a regularization of the shrinkage MVDR method of section 4.2 is proposed. Thereby, the new filter proposed in this section has the next expression,

$$\mathbf{w} = \alpha_1(\hat{\mathbf{R}} + \delta\mathbf{I})^{-1}\mathbf{s} + \alpha_2\mathbf{s} \tag{4.14}$$

In fact, this type of filter also helps to improve the performance of the method proposed in section 4.2 in the interval where $N > M$. This is because the unknown $\mathbf{R}^{-1}$, stemming from the MVDR, is estimated using the regularization $(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}$, which is a better estimate of $\mathbf{R}^{-1}$ than just considering $\hat{\mathbf{R}}^{-1}$, when N is comparable to M and $N > M$. More insights are given in the numerical simulations section comparing figures 4.6 and 4.8. Another interesting interpretation of (4.14) is that it can be viewed as a generalization of the type of DL filters $\mathbf{w} = (\hat{\mathbf{R}} + \delta\mathbf{I})^{-1}\mathbf{s}$. In this regard, the numerical simulations of chapter 5 show that the shrinkage in (4.14) leads to improve the type of DL filters when there is an uncertainty in $\mathbf{s}$. The rationale is that (4.14) promotes the shrinkage towards the matched filter $\mathbf{w} \propto \mathbf{s}$ which circumvents the signal cancelation effect of the MVDR when there is an uncertainty in $\mathbf{s}$. That is, the MVDR may tend to cancel the SOI when either the sample size is small or there is an uncertainty in $\mathbf{s}$, as it may interpret the SOI as an interference. The type of DL methods that deal with a finite sample size and assume a known $\mathbf{s}$, e.g. [9], alleviate the signal cancelation effect, but they still undergo a degradation when $\mathbf{s}$ is not

precisely known. On the other hand, (4.14) promotes the shrinkage towards the matched filter which may imply some attenuation of the SOI, but it avoids the signal cancelation effect. Moreover, bearing in mind an array processing application, note that the main beam of the matched filter is wider than the one of the MVDR. Thereby the linear combination of the regularized sample MVDR with a matched filter in (4.14) can show more robustness to an uncertainty in $\mathbf{s}$ than considering directly a regularized sample MVDR.

The design of the proposed filter implies to obtain an expression for the parameters α_1, α_2 and δ. To this end, the proposed approach is to optimize SINR at the output of the filter. Given our data model in (1.1), the SINR is given by

$$\text{SINR} = \frac{\gamma |\mathbf{w}^H \mathbf{s}|^2}{\mathbf{w}^H \mathbf{R_n} \mathbf{w}}.$$

Taking into account that the data covariance has the expression $\mathbf{R} = \gamma \mathbf{s}\mathbf{s}^H + \mathbf{R_n}$, the SINR can be rewritten as follows,

$$\text{SINR} = \left(\frac{\mathbf{w}^H \mathbf{R} \mathbf{w}}{\gamma |\mathbf{w}^H \mathbf{s}|^2} - 1 \right)^{-1}.$$

Therefore, the design of the shrinkage parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$ and the regularization parameter δ is obtained by optimizing the next expression

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{\mathbf{w}^H \mathbf{R} \mathbf{w}}{|\mathbf{w}^H \mathbf{s}|^2} \\ \text{s.t.} \quad & \mathbf{w} = \alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \mathbf{s} + \alpha_2 \mathbf{s}. \end{aligned}$$

In order to solve this optimization problem a two step procedure is followed. First the optimal $\boldsymbol{\alpha}$ is obtained, for any given δ, by optimizing the SINR. Second the optimal $\boldsymbol{\alpha}$ is substituted in the expression of the SINR, which leads to obtain the optimal δ. This is summarized in the next lemma.

Lemma 4.2 *Assume that a set of observations $\{\mathbf{y}(n)\}_{n=1}^N$ fulfilling the model in (1.1), with assumptions (a)-(c) is available. Given $\{\mathbf{y}(n)\}_{n=1}^N$, consider the problem of estimating the unknown $x(n)$ in (1.1), based on maximizing the SINR when the estimator $\hat{x}(n) = \mathbf{w}^H \mathbf{y}(n)$ is a shrinkage of a regularized sample MVDR i.e. $\mathbf{w} = \alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \mathbf{s} + \alpha_2 \mathbf{s}$. Considering the typical distortionless constraint of the SOI $\mathbf{w}^H \mathbf{s} = 1$, this problem can be mathematically expressed as*

$$\hat{x}(n) = \mathbf{w}^H \mathbf{y}(n); \quad \mathbf{w} = \arg\min_{\mathbf{w}} \frac{\mathbf{w}^H \mathbf{R} \mathbf{w}}{|\mathbf{w}^H \mathbf{s}|^2} \quad s.t. \quad \mathbf{w} = \alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \mathbf{s} + \alpha_2 \mathbf{s} \quad \mathbf{w}^H \mathbf{s} = 1. \tag{4.15}$$

Then, defining $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I}$, *the optimal shrinkage factor* $\boldsymbol{\alpha}_o \triangleq (\alpha_1 \ , \ \alpha_2)^T$ *for any given regularization* δ *is given by,*[1]

$$\boldsymbol{\alpha}_o = \begin{pmatrix} \mathbf{s}^H \mathbf{R} \mathbf{s} \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{s} - \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s} \\ \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{R} \check{\mathbf{R}}^{-1} \mathbf{s} - \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{s} \mathbf{s}^H \mathbf{R} \check{\mathbf{R}}^{-1} \mathbf{s} \end{pmatrix} \tag{4.16}$$

Moreover, the optimal regularization δ_o *is obtained by means of the next optimization,*

$$\delta_o = \arg\min_{\delta} \frac{\boldsymbol{\alpha}_o^H \begin{pmatrix} \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{R} \check{\mathbf{R}}^{-1} \mathbf{s} & \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s} \\ \mathbf{s}^H \mathbf{R} \check{\mathbf{R}}^{-1} \mathbf{s} & \mathbf{s}^H \mathbf{R} \mathbf{s} \end{pmatrix} \boldsymbol{\alpha}_o}{\left| \boldsymbol{\alpha}_o^H \begin{pmatrix} \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{s} \\ \mathbf{s}^H \mathbf{s} \end{pmatrix} \right|^2} \tag{4.17}$$

Proof: The proof for $\boldsymbol{\alpha}_o$ follows easily from the proof for lemma 4.1, just considering $\check{\mathbf{R}}$ instead of $\hat{\mathbf{R}}$. The proof for lemma 4.1 is detailed in section 4.4. On the other hand, δ_o is obtained after easy manipulations by considering the type of filter $\mathbf{w} = \alpha_1 (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \mathbf{s} + \alpha_2 \mathbf{s}$ and by substituting the optimal shrinkage $\boldsymbol{\alpha}_o$ in the expression of the SINR.

■

The optimal values of the shrinkage $\boldsymbol{\alpha}_o$ and regularization parameters δ_o in Lemma 4.2 highlight the dependence on the unknown data covariance $\mathbf{R}$. Namely, $\boldsymbol{\alpha}_o$ and δ_o depend on the next quantities,

$$\begin{matrix} \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{R} \check{\mathbf{R}}^{-1} \mathbf{s} \\ \mathbf{s}^H \mathbf{R} \check{\mathbf{R}}^{-1} \mathbf{s} \\ \mathbf{s}^H \check{\mathbf{R}}^{-1} \mathbf{R} \mathbf{s} \\ \mathbf{s}^H \mathbf{R} \mathbf{s} \end{matrix} \tag{4.18}$$

which must be estimated to obtain a realizable estimator. To this end, the framework dealt with in this thesis must be taken into account. That is, the estimation must bear in

[1] In fact (4.15) yields the expression (4.16) times a scalar term, but the scalar is irrelevant in terms of SINR.

mind that the number of samples N may be scarce compared to the observation dimension M. Thereby, a RMT approach is proposed to obtain the estimations, as it deals implicitly with the small sample size regime. Namely, to estimate the expressions in (4.18) lemma 2.2 is considered, which establishes the converge in probability for the next terms, within the asymptotic framework where $(M,N)\to\infty$ and $M/N\to c\in(0,\infty)$,

$$\begin{aligned}
\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} &\asymp \frac{1}{\left(1-\frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}]\right)^2}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s}\\
\mathbf{s}^H\mathbf{R}\check{\mathbf{R}}^{-1}\mathbf{s} &\asymp \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\mathbf{s}^H\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s}\\
\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} &\asymp \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\mathbf{s}\\
\mathbf{s}^H\mathbf{R}\mathbf{s} &\asymp \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}.
\end{aligned} \tag{4.19}$$

Regarding the expressions in (4.19) recall that $\check{\mathbf{R}}=\hat{\mathbf{R}}+\delta\mathbf{I}$. At this point, the estimators for all the unknown quantities in (4.16) and (4.17) are available. Therefore, an estimator of the optimal shrinkage factor $\boldsymbol{\alpha}_o$ in (4.16) and the regularization parameter δ_o (4.17) is obtained, which is robust to the small sample size thanks to the RMT approach. This result is formally stated in the next theorem.

Theorem 4.2 *Let denote* $\check{\boldsymbol{\alpha}}_o=(\check{\alpha}_{o,1},\check{\alpha}_{o,2})^T=\hat{\boldsymbol{\alpha}}_{o|\delta=\hat{\delta}_o}$ *and* $\check{\mathbf{R}}=\hat{\mathbf{R}}+\delta\mathbf{I}$. *Then, a realizable and (M,N)-consistent estimate of the optimal shrinkage MVDR estimator in (4.15), within the general asymptotic framework where* $M,N\to\infty$ *and* $M/N\to c\in(0,\infty)$, *reads as follows,*

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

$$\hat{\boldsymbol{\alpha}}_o=\begin{pmatrix}\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}-\frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\mathbf{s}\\ \frac{1}{\left(1-\frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}]\right)^2}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s}-\mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s}\left(\frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\right)\mathbf{s}^H\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s}\end{pmatrix} \tag{4.20}$$

$$\hat{\delta}_o = \arg\min_{\delta} \frac{\hat{\boldsymbol{\alpha}}_o^H \begin{pmatrix} \frac{1}{\left(1-\frac{c}{M}\mathrm{Tr}[\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}]\right)^2}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s} & \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\mathbf{s}^H\check{\mathbf{R}}^{-1}\hat{\mathbf{R}}\mathbf{s} \\ \frac{1}{1-c+c\frac{\delta}{M}\mathrm{Tr}[\check{\mathbf{R}}^{-1}]}\mathbf{s}^H\hat{\mathbf{R}}\check{\mathbf{R}}^{-1}\mathbf{s} & \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s} \end{pmatrix} \hat{\boldsymbol{\alpha}}_o}{\left|\hat{\boldsymbol{\alpha}}_o^H \begin{pmatrix} \mathbf{s}^H\check{\mathbf{R}}^{-1}\mathbf{s} \\ \mathbf{s}^H\mathbf{s} \end{pmatrix}\right|^2} \tag{4.21}$$

Proof: The proof is based on lemma 2.2, as it provides the (M,N)-consistent estimates of the unknown quantities in lemma 4.2.

■

Remark 1: In (4.20) and (4.21) it has been used $\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}$ as an estimate of the unknown $\mathbf{s}^H\mathbf{R}\mathbf{s}$ in lemma 4.2. An alternative estimate for $\mathbf{s}^H\mathbf{R}\mathbf{s}$ is $\mathbf{s}^H(\hat{\mathbf{R}}+\hat{\delta}_o\mathbf{I})\mathbf{s}$. In numerical simulations it has been observed that this latter approach leads to slightly better performance in the estimation of the parameter of interest $x(n)$.

Remark 2: In order to find the optimal value for $\hat{\delta}_o$, a one dimensional search is needed, as $\hat{\delta}_o$ is the argument optimizing (4.21). This requires matrix inversions for each iteration of the search due to the expressions involved in (4.20) and (4.21). Fortunately, these matrix inversions can be avoided, which leads to reduce the computational cost of the numerical search. To achieve this aim, first the next identities can be considered, where $\hat{\lambda}_m$ and $\hat{\mathbf{e}}_m$ denote the m-th sample eigenvalue of $\hat{\mathbf{R}}$ and its associated eigenvector, respectively,

$$\begin{aligned}
\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} &= \sum_{m=1}^{M}\frac{|\mathbf{s}^H\hat{\mathbf{e}}_m|^2}{\delta+\hat{\lambda}_m} \\
\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} &= \sum_{m=1}^{M}\frac{|\mathbf{s}^H\hat{\mathbf{e}}_m|^2\hat{\lambda}_m}{(\delta+\hat{\lambda}_m)^2} \\
\mathrm{Tr}[\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}] &= \sum_{m=1}^{M}\frac{\hat{\lambda}_m}{\delta+\hat{\lambda}_m} \\
\mathrm{Tr}[(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}] &= \sum_{m=1}^{M}\frac{1}{\delta+\hat{\lambda}_m} \\
\mathbf{s}^H(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\hat{\mathbf{R}}\mathbf{s} &= \sum_{m=1}^{M}\frac{|\mathbf{s}^H\hat{\mathbf{e}}_m|^2\hat{\lambda}_m}{(\delta+\hat{\lambda}_m)} \\
\mathbf{s}^H\hat{\mathbf{R}}(\hat{\mathbf{R}}+\delta\mathbf{I})^{-1}\mathbf{s} &= \sum_{m=1}^{M}\frac{|\mathbf{s}^H\hat{\mathbf{e}}_m|^2\hat{\lambda}_m}{(\delta+\hat{\lambda}_m)}.
\end{aligned} \tag{4.22}$$

Moreover, note that $\{\hat{\lambda}_m\}_{m=1}^{M}$ and the expression $|\mathbf{s}^H\hat{\mathbf{e}}_m|^2$ need to be computed and stored only once, i.e. there is no need to recompute these values for each iteration of the numerical search.

4.6 Numerical simulations

4.6.1 Performance assessment of the proposed shrinkage MVDR in (4.4).

In this section the performance of the shrinkage MVDR estimator proposed in (4.4) is compared to the conventional sample MVDR estimator (1.9), the theoretical MVDR method (1.7) and the techniques that are robust to the small sample size regime, which were explained in section 1.4.1. Namely, these are the LW-MVDR (1.17), which relies on a shrinkage estimation of the SCM and whose shrinkage factors seek to minimize the MSE of the data covariance in the asymptotic case where $M, N \to \infty$ and $M/N \in (0, \infty)$. Another robust method that is considered for comparison purposes is the DL-MVDR (1.11). More specifically, several loading factors are considered to implement the DL-MVDR (1.11). On the one hand the heuristic DL factor (1.12), which arises from the analysis of the ratio between the SINR of DL-MVDR filters and the SINR of the optimal MVDR. The other rather ad-hoc DL factor is (1.13), which recall that relies on proposing bounds for the DL factor and it selects in an ad-hoc way the lower bound. Recall that the bounds where obtained from the analysis of the estimation error of the data covariance. Finally, the asymptotically SINR optimal loading factor (1.15) is considered. The methods are compared both in terms of the MSE in the estimation of the SOI and the SINR at the output of the filter. The rationale is that the SINR is a popular metric in a wide variety of applications such as beamforming, which is the one considered below. Moreover, the MSE is a popular metric in the design and comparison of estimators and it is important in applications where the complex amplitude of the SOI is important, e.g. in subband beamforming [17], which is within the type of beamforming applications considered below. In this regard, note that the expression of the proposed shrinkage MVDR filter (4.4) depends on a scalar $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(\mathbf{s}^H\hat{\mathbf{R}}\mathbf{s}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}(1-c)^2 - 2(1-c) + 1)$ which does not affect the SINR, though it is important in terms of MSE. Or in other words, the proposed filter optimizes the SINR regardless of the considered scaling of the filter. On the other hand, the MSE is only optimized for the scaling obtained in (4.4).

In order to conduct the simulations, the same simulation environment than in chapter 3 is considered. That is, beamforming in the context of array signal processing is considered

as an application to specify the value of the simulation parameters 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 detail in section 3.7 and arises from considering a uniform linear array, a SOI, several interferences and an AWGN. In this regard, both the SOI and the interferences are considered to be point sources, i.e. no spatial spread is considered, and all of them are generated according to a zero mean complex gaussian distribution whose power depends on the SNR and SIR, which is specified below. Therefore, recall that the data covariance $\mathbf{R}$ and the interference plus noise covariance $\mathbf{R_n}$ have the next form

$$\begin{aligned}\mathbf{R} =&\gamma \mathbf{s}\mathbf{s}^H + \mathbf{R_n};\\ \mathbf{R_n} =& \mathbf{S}\mathbf{P}\mathbf{S}^H + \sigma^2\mathbf{I}.\end{aligned}$$

Moreover, $\mathbf{S}$ is the matrix of steering vectors of the interferences and $\mathbf{P}$, assumed to be diagonal, contains the power associated to each of them. The SNR is considered to be 5 dB in the next simulations and the interferences have the same power than the SOI, which is set to $\gamma = 1$ without loss of generalization. The DOA of the signal of interest is set to $0°$ and eight interferers are considered, whose DOAs are the same than in figure 3.7 to 3.10, i.e. $\left\{\theta_k \frac{180°}{\pi}\right\}_{k=1}^{8} = \{8°, -15°, 23°, -21°, 46°, -44°, -85°, 74°\}$, where recall that θ_k is the DOA of the k-th interferer in radians. Moreover, the number of antennas M is considered to be fixed, and specified below, and the sample size or number of snapshots N is variable to emulate any of the sample size regimes considered herein, i.e. $M/N \in (0, 1)$. For more specific details on the simulation conditions the reader is referred to section 3.7.

In figures 4.1 and 4.2, the performance of the proposed shrinkage MVDR estimator in (4.4) is compared to the one of the theoretical MVDR (1.7) and the sample MVDR (1.9) for a relatively small and high value of M, namely $M = 10$ and $M = 50$, respectively. The performance is shown both in terms of the SINR at the output of the filters and the MSE achieved by each filter in the estimation of the signal of interest. The proposed method is optimal for a large M, as it is an (M, N)-consistent estimate, thereby these simulations permit to assess their performance degradation for a small M. One can see in figures 4.1 and 4.2 that this degradation is not significant and as a consequence the proposed shrinkage MVDR offers a good performance even for rather small values of M. In fact one can observe that their performance is very close to the optimal MVDR.

Furthermore, figures 4.1 and 4.2 highlight that the proposed method is robust to the small sample size regime and that outperforms the sample MVDR 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. The improvement of performance in the intermediate sample size regime is remarkable as well. It is also interesting to observe that for $N \gg M$ the

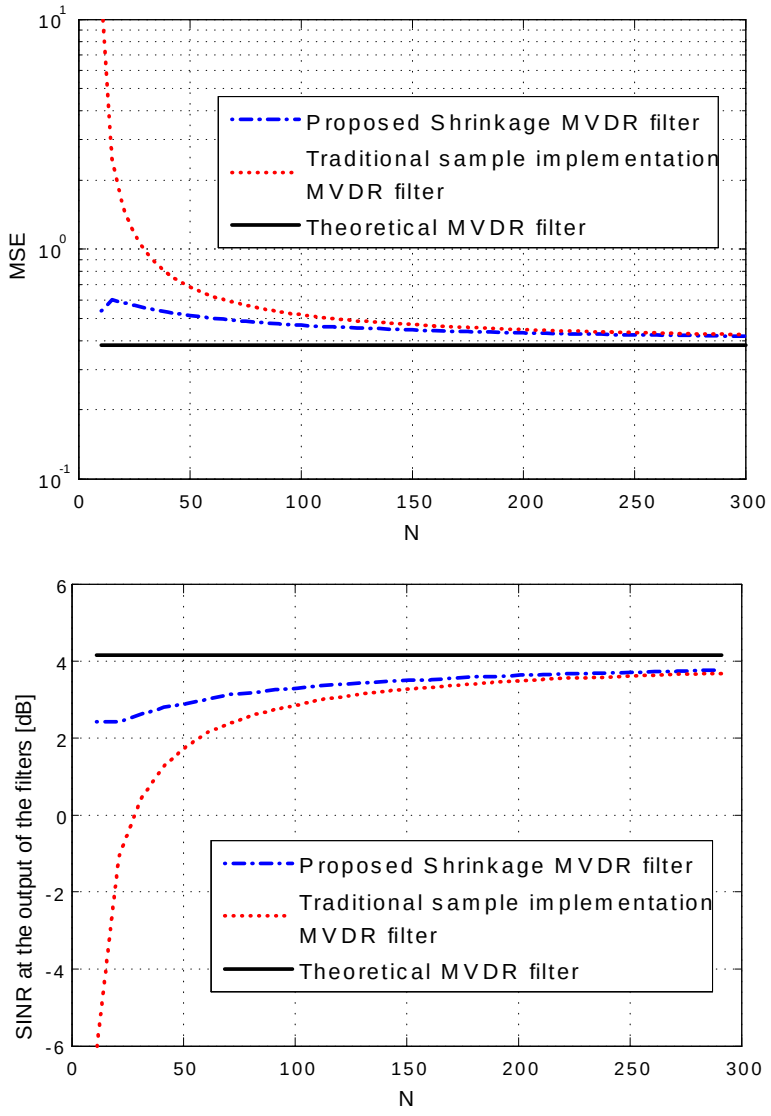

Figure 4.1: Performance comparison between proposed shrinkage MVDR estimator (4.4), MVDR (1.7) and sample MVDR (1.9), when SNR=5 dB and SIR_i=0 dB. M=10.

shrinkage, the theoretical and the sample estimators tend to converge. This is because in this case the SCM is the MVUE of $\mathbf{R}$ and it is well conditioned. As a consequence, the sample MVDR tends to be optimal, i.e. tends to the MVDR method. The shrinkage estimator is aware of this situation and reflects it by means of the shrinkage factors, which lead to obtain the sample MVDR, as it was noted in (4.6). This behavior will be more clear in the upcoming figures.

Next, in figure 4.3 the shrinkage effect in the proposed filter is exemplified. Namely, in figure 4.3 a Monte Carlo simulation is carried out to plot the coefficients $|\check{\alpha}_{c,1}|^2$ and $|\check{\alpha}_{c,2}|^2$ of the shrinkage method proposed in (4.4). The simulation conditions are the same than in the previous figures. Recall that the proposed shrinkage filter reads $\check{\mathbf{w}}_{c,s}=\check{\alpha}_{c,1}\hat{\mathbf{R}}^{-1}\mathbf{s}+\check{\alpha}_{c,2}\mathbf{s}$. 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, $\check{\mathbf{w}}_{c,s}$ tends to give more weight to the sample MVDR than to the matched filter. In fact, when $N \gg M$ the proposed filter tends to disregard the matched 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 4.3 highlights this behavior, as $\frac{M}{N}$ decreases $|\check{\alpha}_{c,1}|^2$ tends to increase whereas $|\check{\alpha}_{c,2}|^2$ tends to decrease. On the other hand, as in general in the small sample size regime the matched 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 matched filter than to the sample MVDR. Indeed in the extreme case where $\frac{M}{N}$ is close to 1, the proposed filter $\check{\mathbf{w}}_{c,s}$ tends to disregard the sample MVDR and give most of the weight to the matched filter. And effectively figure 4.3 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.

In figure 4.4 the proposed shrinkage MVDR is compared to another method that is robust to the small sample size. Namely, the DL implementation of the MVDR (1.11) with the conventional choice of the diagonal loading factor $\delta = 10\hat{\lambda}_{\text{min}}$ [14, p 748]. For simulation purposes the performance of the optimal MVDR (1.7) is displayed as well, $M = 30$ and the rest of simulation parameters are the same than above. The performance is presented in terms of MSE, though in terms of SINR similar results were observed and they are not displayed for the sake of the clarity. One can see, on the top figure, that the proposed shrinkage MVDR outperforms the DL-MVDR, namely in the small sample size regime the improvement is significant. The rationale of this behavior is explained by means of the bottom figures. Note that the DL choice $\delta = 10\hat{\lambda}_{\text{min}}$ depends on the sample estimation of the minimum eigenvalue of $\mathbf{R}$. Unfortunately, in the small sample size regime the distribution of the sample eigenvalues undergoes a spreading phenomenon, which has been studied in the RMT literature, see e.g. [73]. This can be observed in the bottom plots of figure 4.4. The real eigenvalues of $\mathbf{R}$ are $1.77, 1.55, 1.40, 1.37, 1.33, 1.28, 1.25, 1.05, 0.82$

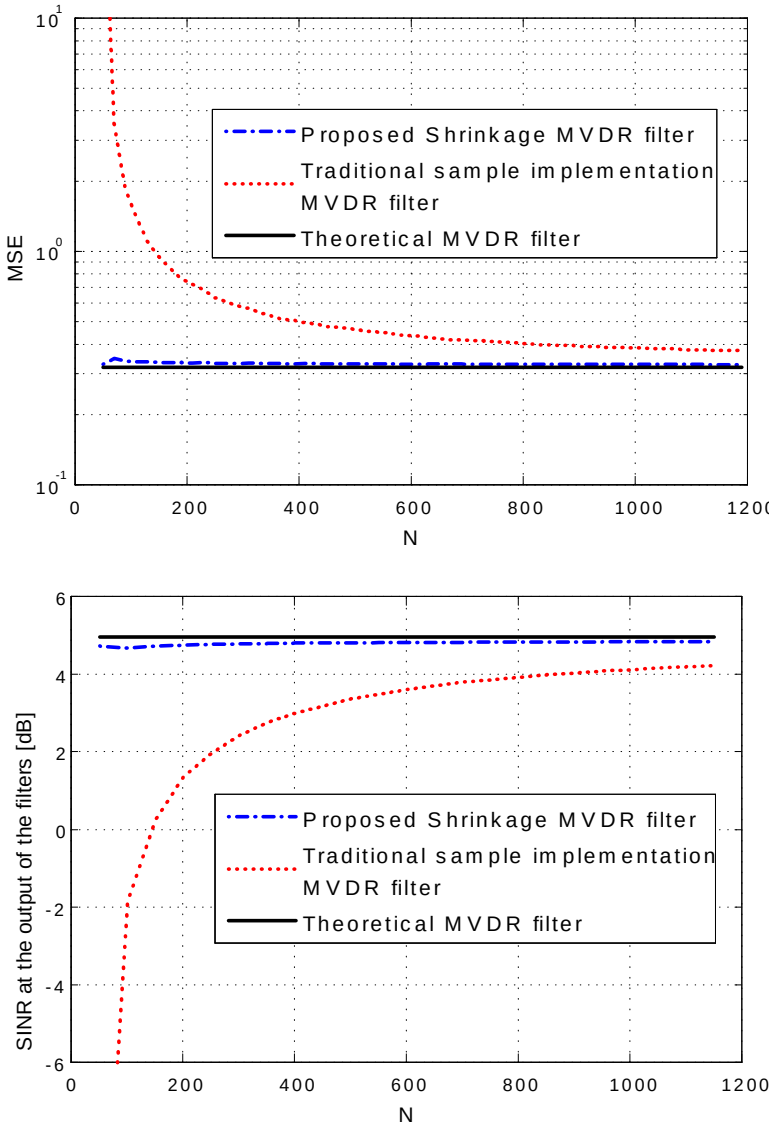

Figure 4.2: Performance comparison between proposed shrinkage MVDR estimator (4.4), MVDR (1.7) and sample MVDR (1.9), when SNR=5 dB and SIR_i=0 dB. M=50.

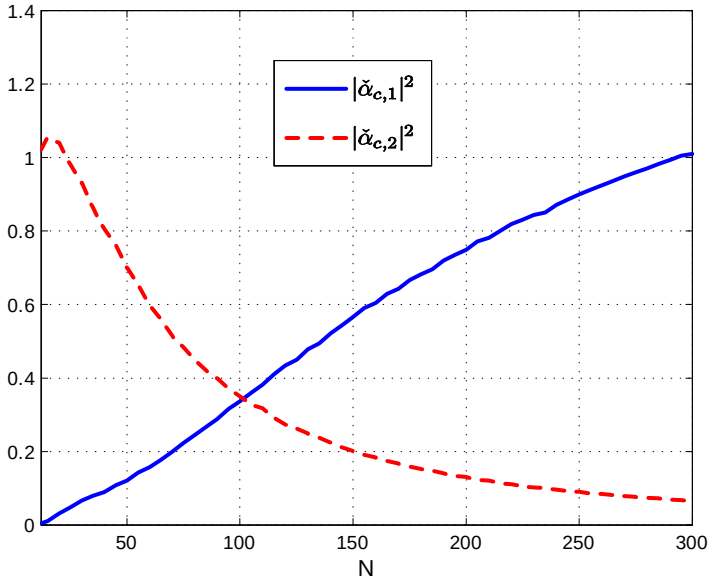

Figure 4.3: Shrinkage factors of the proposed shrinkage MVDR method in (4.4) when $M = 10$, SNR=5 dB and SIR_i=0 dB.

and 0.32 with multiplicity 21. In the large sample size regime, the sample eigenvalues are a good estimate of the real ones. Therefore, one can see in figure 4.4 that for a small M/N the histogram of the eigenvalues tend to be similar to the real eigenvalues. However, in the small sample size regime the sample eigenvalues are not a good estimate of the real ones. In fact, one can see that the larger M/N the more spread the histogram of the sample eigenvalues is. This provokes that in the small sample size regime $\hat{\lambda}_{\min}$ can be smaller than expected and the DL regularization of the SCM may have a limited effect, which leads to worse performance than expected for the DL method. On the other hand, the proposed method is aware of the sample eigenvalue spreading phenomenon and counteracts it thanks to the shrinkage structure and the proposed random matrix theory approach to obtain the shrinkage coefficients. This can be more clear if one takes into account that the proposed filter $\check{\mathbf{w}}_{c,s}=\check{\alpha}_{c,1}\hat{\mathbf{R}}^{-1}\mathbf{s}+\check{\alpha}_{c,2}\mathbf{s}$ is implementing a correction of the eigenvalues of $\hat{\mathbf{R}}^{-1}$. Namely, as it was mentionend in section 4.2, after carrying out the eigendecomposition of the SCM, the propopsed filter can be expressed as $\check{\mathbf{w}}_{c,s} = \hat{\mathbf{E}}(\check{\alpha}_{c,1}\hat{\mathbf{\Lambda}}^{-1} + \check{\alpha}_{c,2}\mathbf{I})\hat{\mathbf{E}}^H\mathbf{s}$, where $\hat{\mathbf{E}}$ contains in its columns the eigenvectors of the SCM and $\hat{\mathbf{\Lambda}}$ is a diagonal matrix containing the eigenvalues of $\hat{\mathbf{R}}$. Thereby, the expression $\check{\alpha}_{c,1}\hat{\mathbf{\Lambda}}^{-1} + \check{\alpha}_{c,2}\mathbf{I}$ highlights that the proposed

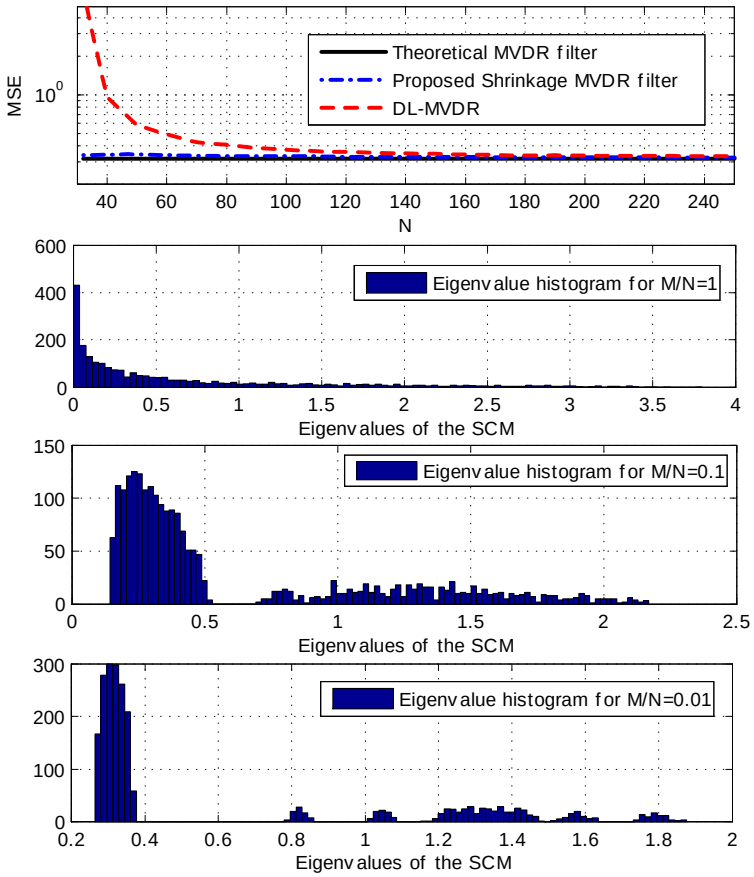

Figure 4.4: On the top figure, performance comparison between proposed shrinkage MVDR in (4.4) and DL-MVDR in (1.11) with $\delta = 10\lambda_{\min}$. Bottom figures, histogram of the SCM eigenvalues to illustrate the eigenvalue spread effect. M=30, SNR=5 dB and SIR_i=0 dB in all the figures.

filter implements a correction in terms of a shrinkage of the eigevalues of $\hat{\mathbf{R}}^{-1}$.

Next, in figure 4.5 the performance of the proposed shrinkage MVDR (4.4) is compared to the LW and DL implementations of the MVDR (1.17) and (1.11), respectively, which are robust to the small sample size regime as well. In these plots we consider the rather ad-hoc DL factor proposed in [77] based on analyzing the estimation error of the covariance, i.e. δ is chosen to be equal to the standard deviation of the diagonal entries of the SCM (1.13). Figure 4.5 shows that the proposed estimator herein outperforms the DL and LW implementations of the MVDR. The reason for this improvement is as follows. On the one hand, both DL and LW implementations seek to enhance the covariance estimate, but they do not deal directly with the estimation of the parameter of interest. DL regularizes the SCM by analyzing the error bounds in the estimation of the covariance, whereas LW seeks a shrinkage of the SCM to optimize asymptotically the MSE of the covariance. On the other hand, the method suggested herein faces directly the estimation of the parameter of interest x by obtaining an $(M, N)-$consistent estimate of the MSE optimal, though unrealizable, estimator of x in (4.2).

Next, the proposed shrinkage of the MVDR in (4.4) is compared to the DL proposed in [9], which is summarized above in section 1.4.1 through equations (1.14) to (1.15). As commented above in section 1.4.1, the importance of this state-of-the-art work is that it provides a methodology to find the loading factor that asymptotically maximizes the SINR. To obtain the numerical results, almost the same simulation conditions than in the previous figures are considered, though considering the values in [9] to allow a fair comparison. Namely, $M = 50$ and $N = 70$, to simulate a small sample size situation, and 29 interferers are considered. To allow variability in the scenario, the DOA of the interferers and the signal of interest are generated, at each iteration of the simulation, as independent random variables uniformly distributed on $[-90°, 90°]$. This setup permits to obtain a general view of the performance for a variable DOA scenario, which complements the previous figures. Moreover, the power of the interferers is the same than the one of the signal of interest, which is 20 dB above the noise. Moreover, to obtain the DL factor in (1.15), the grid search is implemented in the interval (in decibels) $[10 \log \hat{\lambda}_{\min} - 20, 10 \log \hat{\lambda}_{\min} + 40]$ as suggested in [9, p. 77], being $\hat{\lambda}_{\min}$ the minimum eigenvalue of the SCM. In order to obtain a more complete performance comparison the LW-MVDR (1.17), the ad-hoc DL-MVDR (1.13) and the theoretical MVDR (1.7) methods are included in the simulation as well.

The methods are compared in terms of SINR to obtain a fair comparison for [9], which proposes a DL factor maximizing asymptotically the SINR. Thereby, figure 4.6 displays the empirical cumulative distribution function (CDF) of the SINR at the output of the filters. Figure 4.6 shows that the proposed method achieves almost the same SINR performance than [9]. Moreover, as it will be shown in the numerical results of chapter 5, the proposed

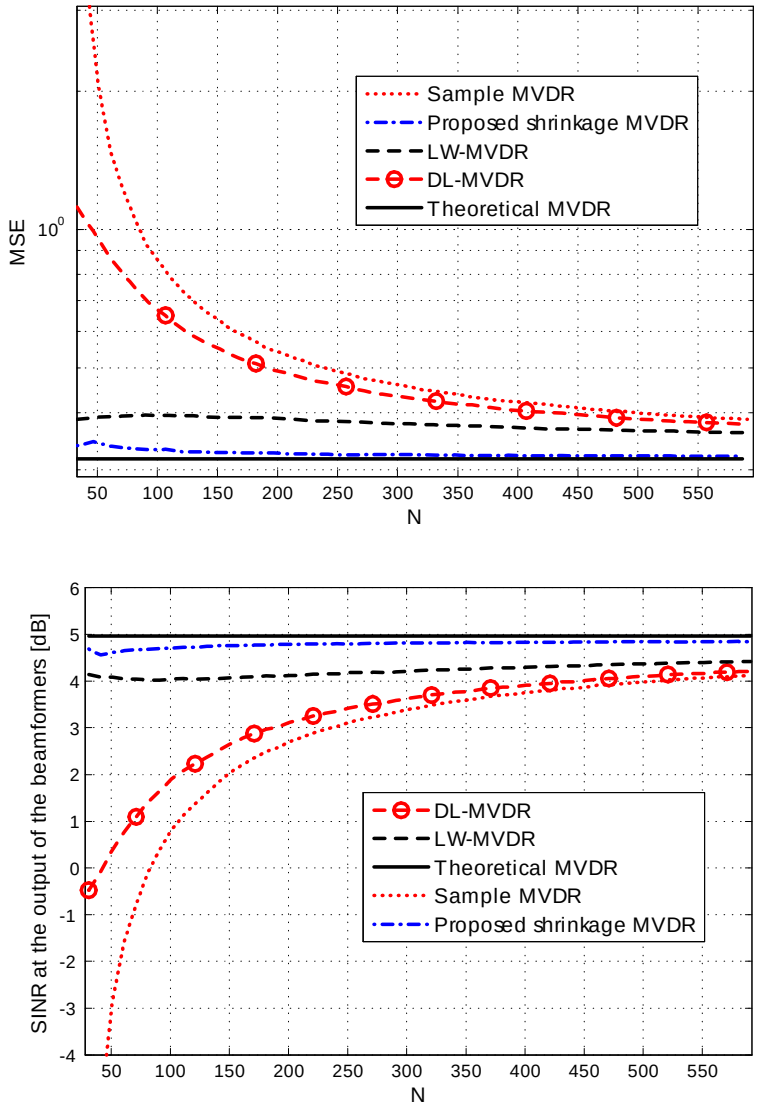

Figure 4.5: Performance comparison between proposed shrinkage MVDR estimator (4.4), MVDR (1.7), LW-MVDR (1.17), sample MVDR (1.9) and DL-MVDR (1.11), implemented with δ equal to the standard deviation of the diagonal entries of $\hat{\mathbf{R}}$, i.e. δ in (1.13). SNR=5 dB, SIR_i=0 dB and M=30.

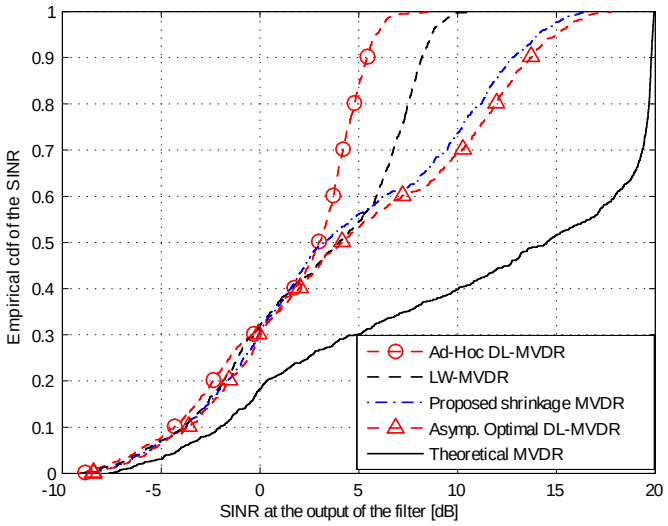

Figure 4.6: CDF of the SINR at the output of the DL-MVDR, implemented with the asymptotically optimal loading factor (1.15), the proposed shrinkage MVDR (4.4), the theoretical MVDR (1.7), LW-MVDR (1.17) and ad-hoc DL-MVDR (1.11), implemented with δ equal to the standard deviation of the diagonal entries of $\hat{\mathbf{R}}$, i.e. δ in (1.13).

shrinkage MVDR method outperforms the DL-MVDR [9] when there is an uncertainty in the signature vector of the signal of interest $\mathbf{s}$. This is because in the small sample size the proposed filter tends to disregard the contribution of the sample MVDR and give more weight to the conventional beamformer or matched filter and thus to avoid the signal cancellation effect provoked in the MVDR due to an uncertain $\mathbf{s}$. On the contrary the DL-MVDR regularizes the SCM through the loading factor, though only taking into account the small sample size degradation. Thereby, as an additional loading factor may be needed due to the uncertainties in $\mathbf{s}$, this may lead to undergoing performance degradation related to the signal cancellation effect. Figure 4.6 also shows that the proposed shrinkage MVDR outperforms the LW-MVDR and ad-hoc DL-MVDR. This corroborates the results obtained in figure 4.5 in an scenario where the DOA of both the SOI and the interferers is variable rather than fixed. As it was commented in figure 4.5, the proposed method obtains better performance than the LW-MVDR and ad-hoc DL-MVDR thanks to an approach based on addressing directly the estimation of the parameter of interest. On the contrary LW-MVDR and ad-hoc DL-MVDR try to obtain an estimate of the data covariance, that is better than the SCM, by optimizing or analyzing certain metrics related to the covariance.

4.6.2 Performance assessment of the proposed shrinkage and regularization of the sample MVDR in Theorem 4.2.

In the next set of simulations the performance associated to the shrinkage of the regularized sample MVDR method, proposed in Theorem 4.2, is assessed. Recall that the aim of this method is to extend or generalize the shrinkage MVDR in (4.4) to support the cases where $M \geqslant N$. That is, the new shrinkage method of Theorem 4.2 supports $M/N \in (0, \infty)$, whereas the shrinkage MVDR in (4.4) was constrained to $M/N \in (0, 1)$. The proposed method is compared to the theoretical MVDR (1.7) and to the state-of-the-art methods which have shown the best performance in the previous figures. These are the asymptotically optimal DL method proposed in [9], which is summarized above in section 1.4.1 and relies on the DL factor (1.15), and the LW-MVDR method in (1.17). All the simulation conditions are the same than in figure 4.6, except that in the next figures N is varied for a fixed value of $M = 50$. More specifically, in figure 4.7 $M = 50$ and $N = 25$ to simulate a situation where $M > N$, which was the motivation to propose the new shrinkage method of Theorem 4.2. Figure 4.7 shows that the new proposed shrinkage method outperforms clearly the LW-MVDR and it obtains the same performance than the asymptotically optimal DL-MVDR method. These results are validated for other sample size regimes in the next figures. Namely, in figure 4.8 $M = 50$ and $N = 70$, thereby the sample size is still rather small but now $N > M$. As in the previous figure the proposed shrinkage method outperforms the LW-MVDR and it leads to almost the same performance than the asymptotically optimal DL-MVDR. Moreover, comparing figure 4.8 with figure 4.6, it can be observed that the shrinkage of the regularized MVDR of Theorem 4.2 outperforms slightly the previous shrinkage MVDR in (4.4). This is because the shrinkage method of Theorem 4.2 is a generalized version of the previous shrinkage estimator in (4.4) due to the regularization of the SCM, which is a better estimator of $\mathbf{R}$ than just considering the SCM. Finally, in figure 4.9, $M = 50$ and $N = 500$, which corresponds to a sample size which is no longer small. The same conclusions than in previous figures can be extracted. On the one hand, the proposed shrinkage MVDR of Theorem 4.2 still outperforms the LW-MVDR method, though their performance are closer due to the increment of N. On the other hand both the shrinkage MVDR of Theorem 4.2 and the asymptotically optimal DL-MVDR obtain the same performance. As it was commented above, the proposed shrinkage MVDR obtains better performance than the LW-MVDR because it seeks to optimize a metric which is directly related to the parameter of interest x in (1.1), i.e. the SINR, whereas the LW-MVDR relies on a shrinkage estimator of the covariance whose aim is to optimize the MSE of the data covariance which is not directly related to the estimation of x.

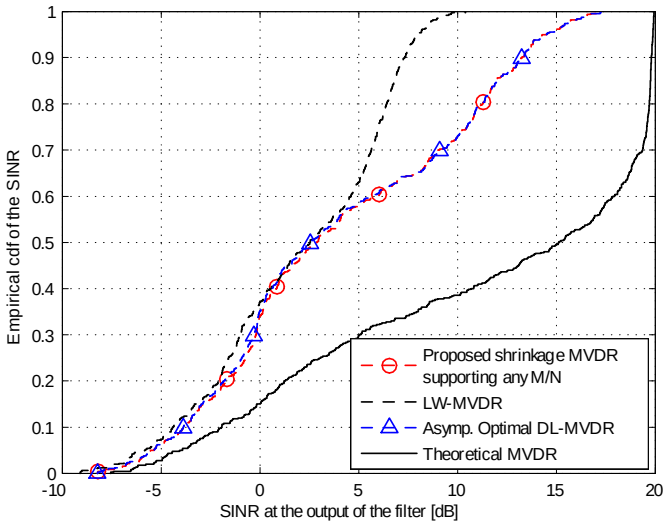

Figure 4.7: CDF of the SINR at the output of the DL-MVDR, implemented with the asymptotically optimal loading factor (1.15), the proposed shrinkage MVDR (4.20), the theoretical MVDR (1.7) and the LW-MVDR (1.17), when $M = 50$ and $N = 25$.

Moreover, at this point, it is important to comment that the proposed shrinkage MVDR of Theorem 4.2 outperforms the asymptotically optimal DL-MVDR when there is an uncertainty in the steering vector $\mathbf{s}$. This will be shown in chapter 5. As it was pointed above, in section 4.5, the rationale is that (4.14) promotes the shrinkage towards the matched filter $\mathbf{w} \propto \mathbf{s}$, which circumvents the signal cancelation effect of the MVDR, when there is an uncertainty in $\mathbf{s}$. And also, in terms of array processing, the matched filter has a wider main beam than the MVDR. Thereby, the linear combination of a regularized sample MVDR with a matched filter promoted by (4.14) can show more robustness to an uncertainty in $\mathbf{s}$ than considering directly a regularized sample MVDR.

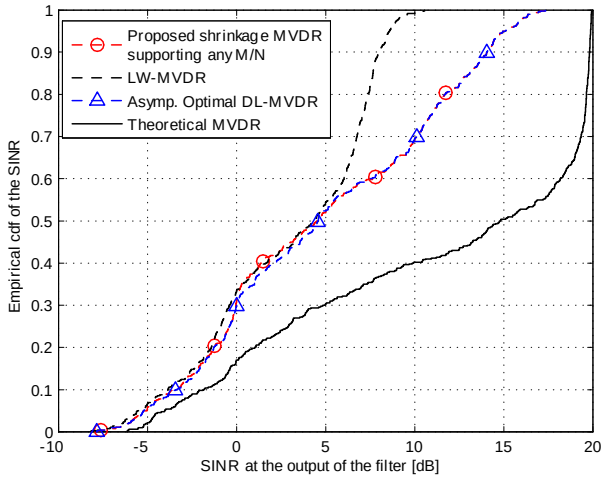

Figure 4.8: CDF of the SINR at the output of the DL-MVDR, implemented with the asymptotically optimal loading factor (1.15), the proposed shrinkage MVDR (4.20), the theoretical MVDR (1.7) and the LW-MVDR (1.17), when $M = 50$ and $N = 70$.

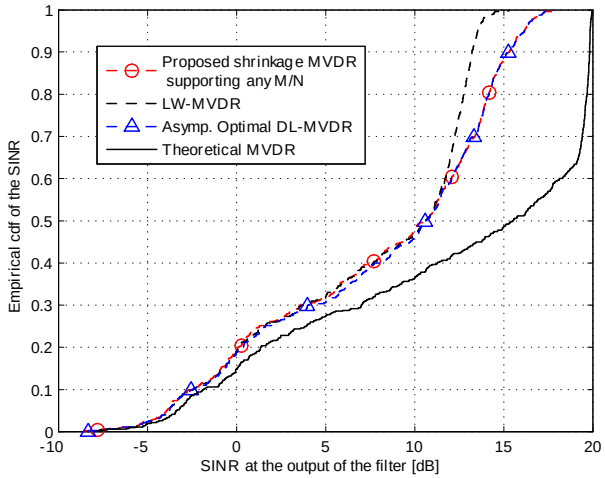

Figure 4.9: CDF of the SINR at the output of the DL-MVDR, implemented with the asymptotically optimal loading factor (1.15), the proposed shrinkage MVDR (4.20), the theoretical MVDR (1.7) and the LW-MVDR (1.17), when $M = 50$ and $N = 500$.

Chapter 5

Signature vector and covariance mismatch scenario: evaluation of shrinkage LMMSE/MVDR and a robust shrinkage MVDR

5.1 Introduction

The shrinkage filters presented in the previous chapters assume a known SOI signature vector. In this chapter not only a finite sample size scenario is considered but also a situation where there is a mismatch in the presumed signature vector associated to the SOI. Thereby, the objective of this chapter is twofold. On the one hand, the performance of the most relevant shrinkage LMMSE and MVDR methods proposed in chapters 3 and 4 is studied when there is both a mismatch in the covariance and the signature vector of the SOI. The proposed shrinkage LMMSE and MVDR methods show more robustness than the related work to an uncertainty in the presumed signature vector. On the other hand, the shrinkage MVDR proposed in section 4.2 is extended to be robust not only to the finite sample size regime, but also to uncertainties in the steering vector, i.e. the SOI signature vector. That is its design takes into account explicitly the finite sample size and the uncertainty of the SOI signature vector. The chapter is organized as follows, in section 5.2 the performance study of the shrinkage LMMSE and MVDR methods of chapters 3 and 4 is carried out. Then in section 5.3 the proposed approach to obtain the robust shrinkage MVDR is presented. Moreover, the challenges to obtain the desired shrinkage filter are

exposed. Namely, it involves solving a nonconvex optimization problem and there are quantities depending on the unknown theoretical covariance of the observations $\mathbf{R}$. Next, in section 5.4 the challenges posed by the robust shrinkage MVDR optimization problem are solved. On the one hand, the quantities depending on $\mathbf{R}$ are estimated relying on results from RMT, which is an approach that deals naturally with the small sample size regime. On the other hand, the original problem is reformulated in the form of a SOCP, which is a convex problem and thereby can be efficiently solved, i.e. in polynomial time, by means of interior point method software tools. Finally, in section 5.5 the proposed robust shrinkage MVDR method is compared to related work methods by means of numerical simulations.

5.2 Evaluation of the shrinkage LMMSE and MVDR methods of chapters 3 and 4 in a signature vector and covariance mismatch scenario.

The aim of this section is to carry out a performance assessment, by means of numerical simulations, of the most important shrinkage LMMSE and MVDR methods of chapters 3 and 4, when there is an uncertainty in the signature vector of the SOI and a small sample size situation. Recall that those methods assumed the next signal model (1.1)

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

Moreover, it was assumed that signature vector $\mathbf{s}$ is perfectly known. However, in practice the actual steering vector $\tilde{\mathbf{s}}$ may differ from the presumed $\mathbf{s}$ due to e.g. errors when pointing towards the signal of interest, see [37]. This signature vector mismatch leads to a performance degradation of the LMMSE and MVDR methods, as it was explained in chapter 1, because the SOI may be confused as an interference. The methods proposed in chapters 3 and 4 were designed to cope with a finite sample size regime and they do not consider an uncertain signature vector. Thereby, it is interesting to give more insights on their performance when not only the sample size is small but also when there is a mismatch in the presumed signature vector. The simulation conditions to carry out this study are the same than the ones of the numerical results in chapters 3 and 4, i.e. beamforming in an array processing context is considered. The only difference is that herein an error is introduced to model the mismatch between the actual and the presumed steering vectors. More specifically, in the next simulations, an error in the presumed DOA of the SOI is

introduced, which accounts in practice to a pointing error towards the signal of interest in an array signal processing context. Thereby, the actual steering vector $\tilde{\mathbf{s}} = \mathbf{s}(\theta_p + \theta_e)$, where $\mathbf{s}(\theta)$ is the presumed steering vector depending on a generic DOA θ, θ_p is the presumed DOA associated to the SOI and θ_e is an error in terms of DOA.

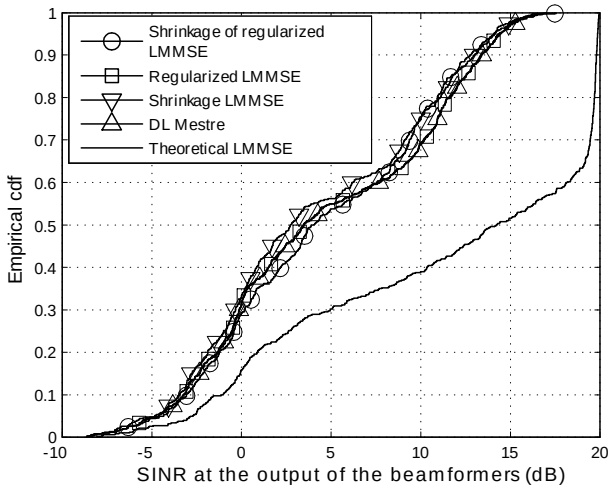

Figure 5.1: Comparison of the empirical cdf of the SINR at the output of the shrinkage LMMSE proposed in (3.28), the shrinkage of the regularized LMMSE proposed in Theorem 3.6, the regularized LMMSE in [79], the theoretical LMMSE (1.4) and Mestre-DL in [9] see (1.15), when M=50, N=55 and the DOA mismatch is 0.1°.

5.2.1 Performance of shrinkage LMMSE methods

In this section the performance of the most relevant shrinkage LMMSE methods proposed in chapter 3, as well as their related work, is studied. Namely, these are the shrinkage LMMSE proposed in (3.28) and its generalization to deal with $M > N$, i.e. the shrinkage of the regularized LMMSE proposed in Theorem 3.6. Regarding the related work the next methods are considered, as they demonstrated the best performance among the state-of-the-art methods in chapter 3. On the one hand, the regularized LMMSE in [79], which proposes a shrinkage of the sample covariance in the LMMSE method, whose shrinkage factors optimize asymptotically the MSE of the SOI. On the other hand, the DL method in [9], which is a shrinkage of the SCM in the LMMSE method and whose DL or shrinkage factor is designed to optimize asymptotically the SINR at the output of the filter. For

comparison purposes the theoretical LMMSE method (1.4) is considered in the simulations, i.e. the LMMSE implemented with the theoretical covariance matrix and the actual steering vector of the SOI.

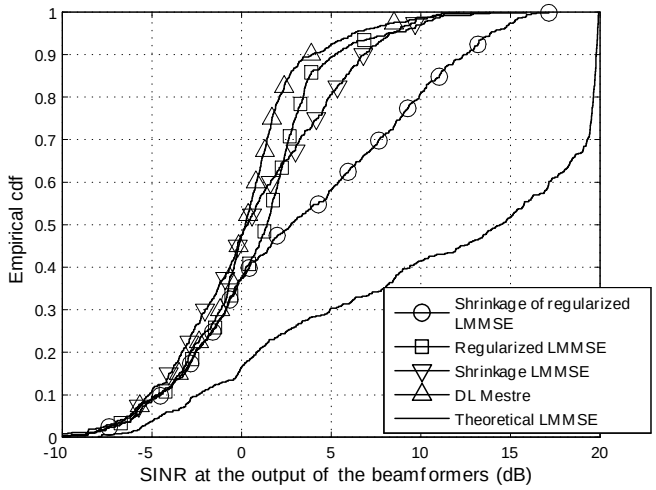

Figure 5.2: Comparison of the empirical cdf of the SINR at the output of the shrinkage LMMSE proposed in (3.28), the shrinkage of the regularized LMMSE proposed in Theorem 3.6, the regularized LMMSE in [79], the theoretical LMMSE (1.4) and Mestre-DL in [9] see (1.15), when M=50, N=55 and the DOA mismatch is 0.5°.

Figures 5.1 to 5.3 evaluate the performance of the filters when the DOA mismatch is increased and within a small sample size situation. More specifically the DOA mismatch in those figures is 0.1°, 0.5° and 1°, respectively, whereas $M = 50$ and $N = 55$. Moreover, the signal of interest and 29 interference signals are considered to be present in the scenario. It is assumed that the SOI and the interferers have the same power and their DOA are generated randomly at each iteration according to a uniform distribution between $-90°$ and $90°$. Moreover, an SNR=20 dB is considered in the simulations. Given these simulation conditions, the empirical cdf associated to the SINR at the output of the filters, is ploted in figures 5.1 to 5.3. It can be observed that as the DOA mismatch increases, the proposed shrinkage of the regularized LMMSE leads to obtain better performance than the rest of the methods under comparison. Moreover, the shrinkage LMMSE, which is a simplified version of the shrinkage of the regularized LMMSE, also obtains better performance than the DL method [9] and better performance than the regularized LMMSE [79] within some SINR intervals.

In order to understand the better performance of the proposed methods recall that the shrinkage of the regularized LMMSE and the shrinkage LMMSE rely on a filter of the type $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, where $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I}$ in the former and $\check{\mathbf{R}} = \hat{\mathbf{R}}$ in the latter. Thereby, the proposed shrinkage filters rely on a linear combination of the regularized LMMSE or the sample LMMSE with a matched filter. The matched filter on the one hand, has less spatial resolution than the type of LMMSE methods and thereby may be less sensitive to a DOA mismatch. On the other hand, under a DOA mismatch the LMMSE methods may confuse the SOI as an interference and may tend to cancel it, whereas the matched filter considers an scenario with a SOI and white noise. Thereby, although the matched filter will not point towards the correct direction, it will not try to cancel the SOI as it was an interference. Moreover, as the proposed shrinkage filters consider in their filter structure the sample or regularized LMMSE they can lead to reject better the interferers than just considering the matched filter.

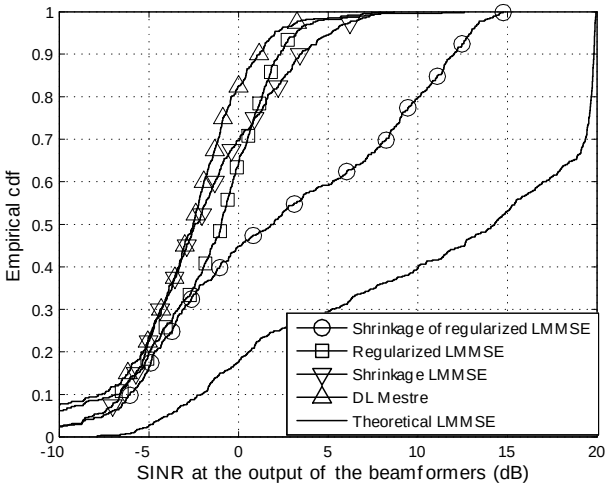

Figure 5.3: Comparison of the empirical cdf of the SINR at the output of the shrinkage LMMSE proposed in (3.28), the shrinkage of the regularized LMMSE proposed in Theorem 3.6, the regularized LMMSE in [79], the theoretical LMMSE (1.4) and Mestre-DL in [9] see (1.15), when M=50, N=55 and the DOA mismatch is 1°.

On the other hand, the regularized LMMSE method [79] and the DL method [9] undergo the signal cancelation effect due to the DOA mismatch. That is, they can confuse the SOI as an interference, which leads to try to cancel the SOI. They consider a regularization of the SCM to combat the small sample size regime, which could offer some benefits to combat

the DOA mismatch as well, see e.g. [37] and section 1.4.2. However, the regularization or DL of the SCM is designed to deal only with the small sample size regime in [79] and [9], assuming a known steering vector. Thereby, an additional tuning of the regularization or loading factor of the SCM is needed to deal properly with the DOA mismatch. Moreover, note that among the proposed methods, the shrinkage of the regularized LMMSE obtains better peformance than the shrinkage LMMSE. The rationale is that in the filter structure $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, the former considers the regularization $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I}$ whereas the latter considers $\check{\mathbf{R}} = \hat{\mathbf{R}}$ and the regularization leads to a better estimate in the small sample size and under a DOA mismatch than just considering the SCM.

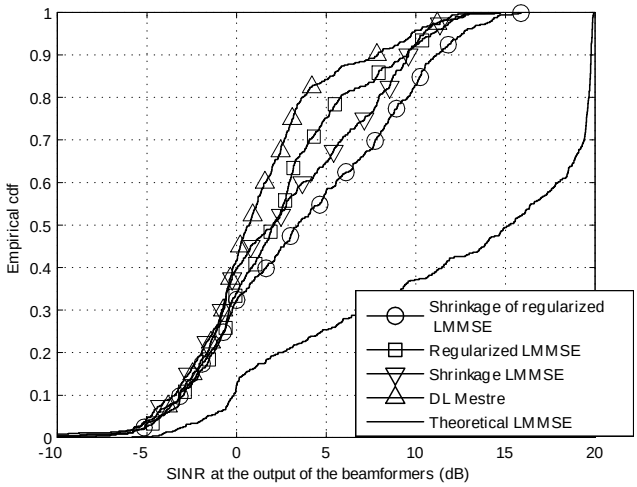

Figure 5.4: Comparison of the empirical cdf of the SINR at the output of the shrinkage LMMSE proposed in (3.28), the shrinkage of the regularized LMMSE proposed in Theorem 3.6, the regularized LMMSE in [79], the theoretical LMMSE (1.4) and Mestre-DL in [9] see (1.15), when M=20, N=22 and the DOA mismatch is 1°.

The next comments are related to figures 5.3, 5.4 and 5.5. The objective of these figures is to assess the impact in the performance of varying the dimension M of the observed signal, when there is a DOA mismatch and a small sample size situation. Note that in our case M is the number of antennas in the array. The simulation conditions are as follows. The DOA mismatch is set to 1° in all the simulations, whereas the sample size is set to maintain the same ratio M/N when varying M. Thereby, in figure 5.3 $M = 50$ and $N = 55$, in figure 5.4 $M = 20$ and $N = 22$ and in figure 5.5 $M = 10$ and $N = 11$. Moreover, the ratio K/M is maintained when varying M in the different simulations to allow a fair

comparison, where K is the number of interference signals plus the SOI. Namely, in figure 5.3 $M = 50$ and $K = 30$, in figure 5.4 $M = 20$ and $K = 12$ and in figure 5.5 $M = 10$ and $K = 6$. The rest of the simulation parameters are the same than in previous figures.

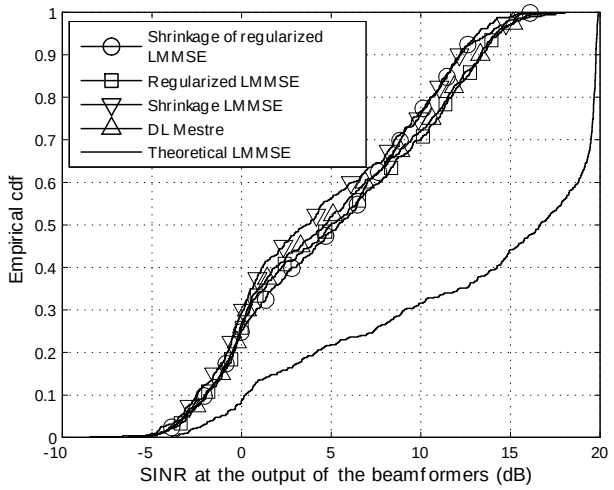

Figure 5.5: Comparison of the empirical cdf of the SINR at the output of the shrinkage LMMSE proposed in (3.28), the shrinkage of the regularized LMMSE proposed in Theorem 3.6, the regularized LMMSE in [79], the theoretical LMMSE (1.4) and Mestre-DL in [9] see (1.15), when M=10, N=11 and the DOA mismatch is 1°.

Figures 5.3, 5.4 and 5.5 highlight that increasing M leads the shrinkage of the regularized LMMSE, proposed in Theorem 3.6, to outperform the rest of the methods. Also the more simple shrinkage LMMSE proposed in (3.28), leads to some performance gains compared to the related work methods, when M is increased. The rationale for these results relies on a similar explanation than the one given in the previous figures. On the one hand, when M and K increase and there is a DOA mismatch, the signal cancelation effect on the LMMSE methods gets worse. This is because the filters tend to have more spatial resolution and it is more easy that the SOI is confused as it was an interference. The regularization of the SCM in the LMMSE may alleviate this effect. However, the regularized LMMSE [79] and the DL in [9] design the regularization to cope with the finite sample size, thereby under a DOA mismatch they may undergo a performance degradation. On the other hand, the shrinkage of the regularized LMMSE relies on a linear combination of a regularized LMMSE and a matched filter. Under a DOA mismatch the matched filter will not point exactly towards the SOI but it will not try to cancel the SOI as it was an interference.

This permits the shrinkage of the regularized LMMSE to be more robust to the DOA mismatch than the regularized LMMSE and the DL method. On the other hand, the proposed shrinkage also considers the regularization of the LMMSE, which leads to better rejection of the interference than just considering a matched filter. A similar rationale can be applied to explain why the shrinkage LMMSE in (3.28) obtains some gains compared to the regularized LMMSE and the DL method. Moreover, as it was comented before, the shrinkage of the regularized LMMSE obtains better performance than the shrinkage LMMSE because it relies on a regularization of the SCM, which is a better estimate than just considering the SCM. Moreover, in the small sample size regime the shrinkage LMMSE tends to disregard the LMMSE and give more weight to the matched filter, thereby the rejection against the interference is better in the shrinkage of the regularized LMMSE.

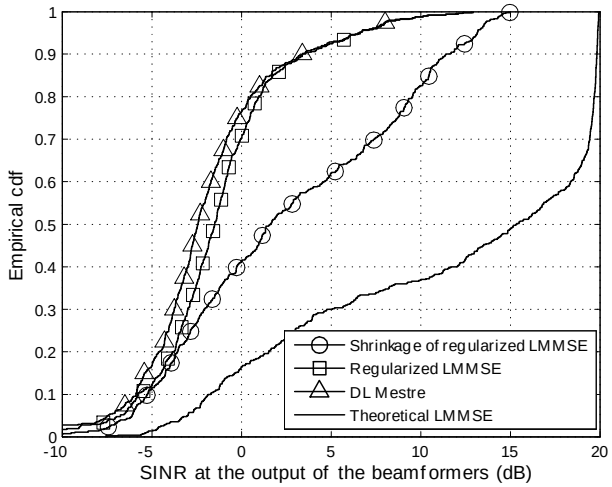

Figure 5.6: Comparison of the empirical cdf of the SINR at the output of the shrinkage LMMSE proposed in (3.28), the shrinkage of the regularized LMMSE proposed in Theorem 3.6, the regularized LMMSE in [79], the theoretical LMMSE (1.4) and Mestre-DL in [9] see (1.15), when M=50, N=30 and the DOA mismatch is 1°.

In figures 5.3, 5.6 and 5.7 the impact of several small sample size situations is studied, in an scenario where a DOA mismatch is present as well. Namely, in figure 5.3 $M = 50$ and $N = 55$, whereas in figure 5.6 $M = 50$ and $N = 30$ and in figure 5.7 $M = 50$ and $N = 10$. In all the simulations the DOA mismatch is 1° and $K = 30$. The rest of the simulation parameters are the same than in previous figures. Note that in figures 5.6 and 5.7 $M > N$, thereby the shrinkage LMMSE is not considered, as it does not

support this scenario. The figures highlight that the proposed shrinkage of the regularized LMMSE in Theorem 3.6 clearly outperforms the regularized LMMSE in [79] and DL in [9] when $M = 50$ and $N = 55$ or $N = 30$. The rationale is the same than in the previous simulations. Thanks to the linear combination $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, the shrinkage of the regularized LMMSE profits the fact that the matched filter is less sensitive to a DOA mismatch than the LMMSE. Moreover, it also profits the fact that the regularized LMMSE has better interference rejection capabilities than a matched filter. On the other hand, the regularization considered in [79] and [9] is designed to deal with the small sample size situation, but the protection against a DOA mismatch is not guaranteed. In fact note that decreasing the sample size (i.e. N for a fixed M) leads to increase the regularization of the SCM considered in [79] and [9] which also leads to more protection against a DOA mismatch. This can be observed in figure 5.7 where all the methods tend to give the same performance.

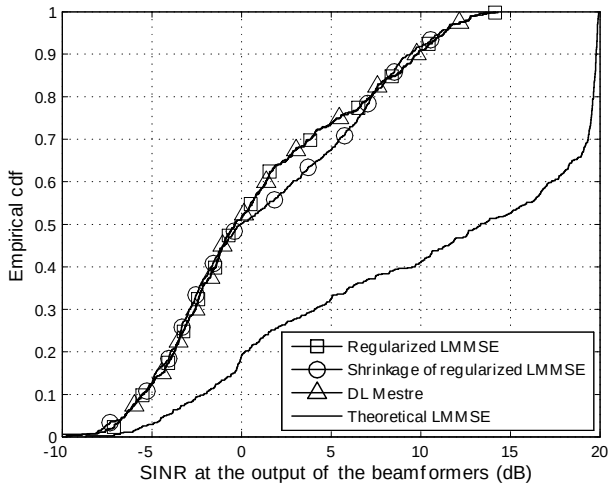

Figure 5.7: Comparison of the empirical cdf of the SINR at the output of the shrinkage LMMSE proposed in (3.28), the shrinkage of the regularized LMMSE proposed in Theorem 3.6, the regularized LMMSE in [79], the theoretical LMMSE (1.4) and Mestre-DL in [9] see (1.15), when M=50, N=10 and the DOA mismatch is 1°.

5.2.2 Performance of shrinkage MVDR methods

In this section a performance evaluation of the methods related with the MVDR technique will be carried out. More specifically, on the one hand the proposed shrinkage MVDR in (4.4) is considered. Moreover, the shrinkage of the regularized MVDR proposed in Theorem 4.2 is incorporated in the simulations. Recall that this is a generalization of the shrinkage MVDR in (4.4) to support the cases where $M > N$ and in general to improve its performance due to the regularization of the SCM. For the comparison purposes, the asymptotically SINR optimal DL method, proposed in [9], which is summarized above in section 1.4.1 is included in the simulations. The rationale is that this method was the best state-of-the-art algorithm in the numerical results of chapter 4. Finally, the theoretical MVDR (1.7) is considered as an upperbound in the simulations, as it considers the true covariance and signature vector in its implementation.

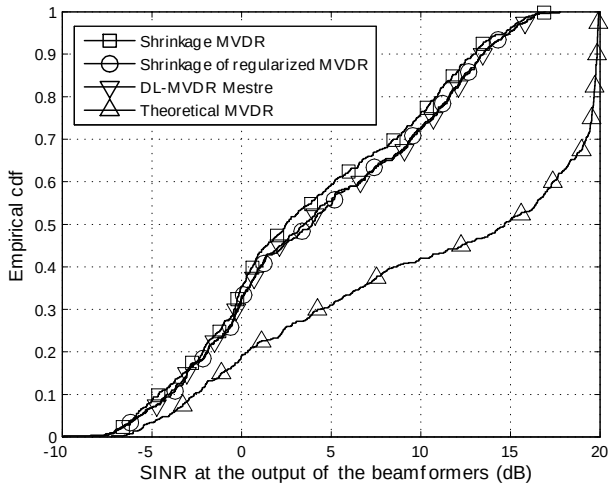

Figure 5.8: Comparison of the empirical cdf of the SINR at the output of the shrinkage MVDR proposed in (4.4), the shrinkage of the regularized MVDR proposed in Theorem (4.2) the theoretical MVDR (1.7) and Mestre-DL in [9] see (1.15), when M=50, N=55 and the DOA mismatch is 0.1°.

In figures 5.8 to 5.10 the effect of a DOA mismatch is considered. Namely, the DOA mismatch in these figures is 0.1°, 0.5° and 1°, respectively. Moreover, a small sample size situation is considered, namely $M = 50$ and $N = 55$. Thereby, in this situation the methods have to cope both with a mismatch in the presumed steering vector and a small

sample size situation. In this simulation the SOI and 29 interference signals are considered to be in the scenario. Both the SOI and the interferences have the same power and their DOA are generated randomly at each iteration according to a uniform distribution between $-90°$ and $90°$. Moreover, the SNR=20 dB.

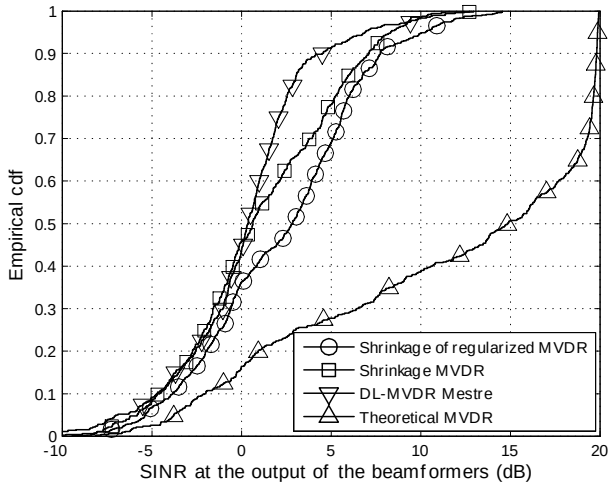

Figure 5.9: Comparison of the empirical cdf of the SINR at the output of the shrinkage MVDR proposed in (4.4), the shrinkage of the regularized MVDR proposed in Theorem (4.2) the theoretical MVDR (1.7) and Mestre-DL in [9] see (1.15), when M=50, N=55 and the DOA mismatch is $0.5°$.

It can be observed in figure 5.8 that for a very small DOA mismatch, the performance of the methods is similar to the case where the signature vector is perfectly known, see chapter 4. However, when the error in the presumed DOA increases, it can be observed in figures 5.9 and 5.10 that both the shrinkage MVDR proposed in (4.4) and the shrinkage of the regularized MVDR proposed in Theorem (4.2) obtain better performance than the DL-MVDR method proposed in [9]. The rationale is that the DL-MVDR in [9] relies on an asymptotically optimal DL factor which is designed to deal with the small sample size, though it assumes that the steering vector is pefectly known. Thereby, when there is a mismatch between the presumed and actual signature vectors, a degradation is expected in this method. The signature mismatch leads to the signal cancellation effect, where the SOI is interpreted as an interference and the filter tends to cancel it. As the MVDR is a method with a high spatial resolution this can provoke a significant degradation for a DOA mismatch situation. In fact the DL-MVDR would require an additional amount of

loading factor to deal with the steering vector uncertainty.

On the other hand, both the shrinkage MVDR and the shrinkage of the regularzied MVDR rely on a filter structure of the type $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, where $\check{\mathbf{R}} = \hat{\mathbf{R}}$ in the former and $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I}$ in the latter. Thereby these filters combine the regularized or sample MVDR with a matched filter, which is known to have less spatial resolution than the MVDR and thereby is less sensitive to a DOA mismatch and it is less prone to undergo the signal cancellation effect. Another interesting interpretation is that the matched filter can be viewed as a DL-MVDR where the loading factor tends to infinity [9]. These are the reasons behind the better performance of the proposed shrinkage methods compared to the DL-MVDR in [9]. Moreover, note that the shrinkage of the regularized MVDR is better than the shrinkage MVDR due to its regularization of the SCM which permits to cope more properly with the small sample size situation, when considered in the shrinkage filter $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$.

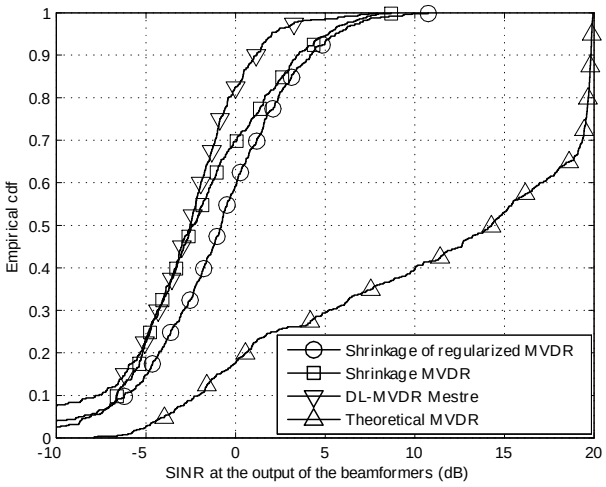

Figure 5.10: Comparison of the empirical cdf of the SINR at the output of the shrinkage MVDR proposed in (4.4), the shrinkage of the regularized MVDR proposed in Theorem (4.2) the theoretical MVDR (1.7) and Mestre-DL in [9] see (1.15), when M=50, N=55 and the DOA mismatch is 1°.

In figures 5.10, 5.11 and 5.12 a simulation is carried out, whose aim is to assess the effect of varying M in the performance of the filters, when there is a DOA mismatch and a small sample size situation. More specifically, the DOA mismatch is set to 1° in all the simulations, whereas the sample size is set to maintain the same ratio M/N when varying

M. Recall that in figure 5.10 $M = 50$ and $N = 55$, thereby in figure 5.11 $M = 20$ and $N = 22$ and in figure 5.12 $M = 10$ and $N = 11$. Moreover, it is important to mention that the ratio K/M is maintained when varying M, where K is the number of interference signals plus the SOI. In figure 5.10 $M = 50$ and $K = 30$. Thereby, in figure 5.11 $M = 20$ and $K = 12$ and in figure 5.12 $M = 10$ and $K = 6$. The rest of the simulation parameters are the same than in the previous simulations.

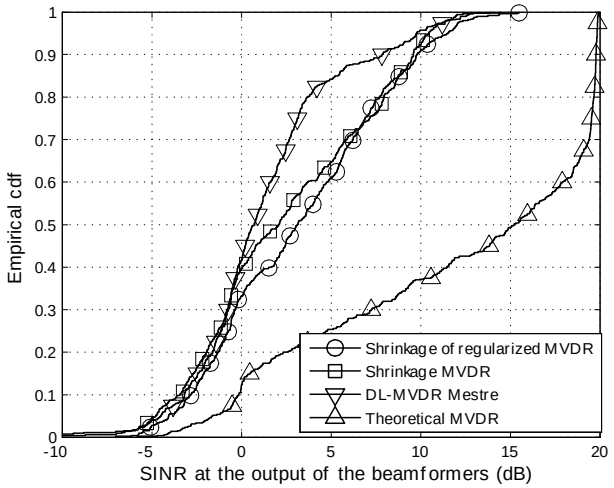

Figure 5.11: Comparison of the empirical cdf of the SINR at the output of the shrinkage MVDR proposed in (4.4), the shrinkage of the regularized MVDR proposed in Theorem (4.2) the theoretical MVDR (1.7) and Mestre-DL in [9] see (1.15), when M=20, N=22 and the DOA mismatch is 1°.

Figures 5.10, 5.11 and 5.12 highlight that when M is increased the proposed shrinkage MVDR and shrinkage of the regularized MVDR tend to outperform the DL-MVDR. The rationale is similar than in the previous simulation. That is, as M increases, the type of MVDR filters (i.e. the sample MVDR and the regularized or DL-MVDR) tend to be more sharp. As a consequence a DOA mismatch leads to exacerbate the consequences of the signal cancelation effect and to deteriorate the performance. In the shrinkage filters this effect is alleviated because they combine the sample MVDR or the regularized MVDR with a matched filter. This method, on the one hand is less resolutive and on the other hand does not lead to the signal cancelation effect. That is, due to the DOA mismatch the matched filter will not point exactly towards the SOI but it will not try to cancel it as it was an interference as it happens in the MVDR. On the other hand, it can be observed in

the simulations that when M is decreased all the filters tend to give the same performance as they have less spatial resolution, but this makes them more robust to a DOA mismatch.

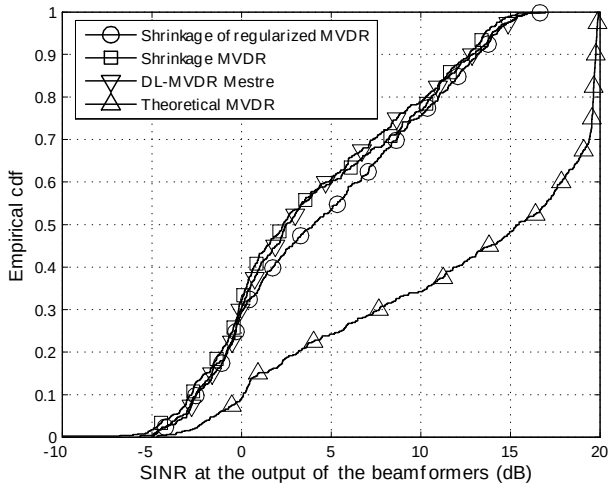

Figure 5.12: Comparison of the empirical cdf of the SINR at the output of the shrinkage MVDR proposed in (4.4), the shrinkage of the regularized MVDR proposed in Theorem (4.2) the theoretical MVDR (1.7) and Mestre-DL in [9] see (1.15), when M=10, N=11 and the DOA mismatch is 1°.

In figure 5.13 the same type of simulation than in figure 5.10 is considered, but in this case the sample size is decreased. Namely, the DOA mismatch is considered to be 1° and two situations are considered, in the first one $M = 50$ and $N = 30$, whereas in the second one $M = 50$ and $N = 10$. Note that in this situation $M > N$, thereby the shrinkage MVDR is not considered, as it does not support this scenario. However, its generalization which is the shrinkage of the regularized MVDR is able to cope with this situation and it is the method considered in the simulation. The rest of simulation parameters are the same than in figure 5.10. It is interesting to observe in figure 5.13 that for $M = 50$ and $N = 30$ the same conclusion can be extracted than in the previous plot, i.e. figure 5.10. That is, the shrinkage of the regularized MVDR outperforms the DL-MVDR in [9] thanks to the linear combination of the regularized MVDR with the matched filter, which is less sensitive to a DOA mismatch. However, it can be observed that the shrinkage of the regularized MVDR and the DL-MVDR tend to get closer than in figure 5.10. This is because the filters tend to increase the loading factor as the sample size is smaller. This has a beneficial effect in terms of dealing with steering vector mismatch as it was pointed

out e.g. in [37] and section 1.4.2. This is confirmed for an even smaller sample size regime of $M = 50$ and $N = 10$, which leads to obtain the same performance in both the shrinkage of the regularized MVDR and the DL-MVDR. In fact better performance than for $M = 50$ and $N = 30$ is obtained because for $N = 10$ the loading factor is higher and the steering vector mismatch is treated more properly and it more than compensates that N is smaller.

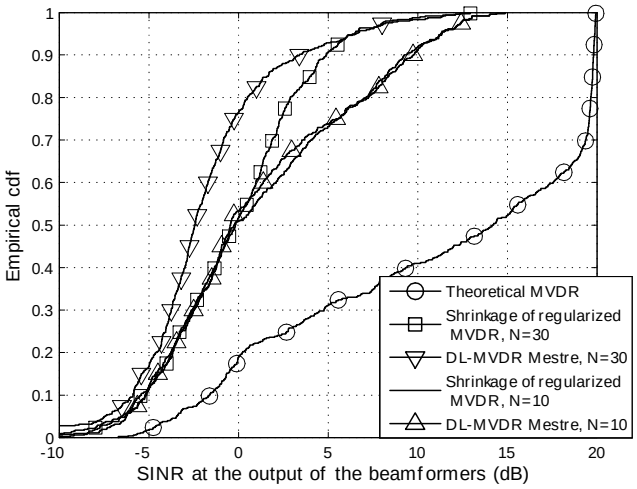

Figure 5.13: Comparison of the empirical cdf of the SINR (for different small sample sizes of N=30, N=10) at the output of the next filters. The shrinkage of the regularized MVDR proposed in Theorem (4.2) the theoretical MVDR (1.7) and Mestre-DL in [9] see (1.15), when M=50 and the DOA mismatch is 1°.

5.3 Robust shrinkage MVDR problem

In chapters 3 and 4 the type of shrinkage filter $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$ was proposed to counteract the degradation of the sample LMMSE and MVDR due to the finite sample size constraint. The methods proposed in those chapters assume that the SOI signature vector $\mathbf{s}$ is perfectly known. This is an assumption that has been considered in other related works, see e.g. [9]. On the other side, as it was explained in chapter 1, recall that another possible source of degradation of the LMMSE and MVDR methods is a mismatch between the presumed and the actual steering vector. Therefore, the assumption of a known $\mathbf{s}$ can be interpreted as focusing on the degradation that the small sample size provokes in

the MVDR and the LMMSE [9]. Another interpretation of that assumption is that one assumes that the small sample size degradation dominates compared to the steering vector mismatch degradation, see e.g. [49]. In this chapter, the shrinkage MVDR proposed in chapter 4 is improved to be robust not only to the finite sample size effect, but also to uncertainties in the presumed steering vector $\mathbf{s}$. In fact, the shrinkage of the LMMSE and the MVDR relying on $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, which were proposed in chapters 3 and 4, are equivalent in terms of SINR. Therefore, from the SINR viewpoint, the shrinkage method proposed in this chapter can be viewed also as a robust extension of the shrinkage LMMSE of chapter 3 to uncertainties in $\mathbf{s}$.

In this chapter, the actual steering vector, denoted by $\tilde{\mathbf{s}}$, is assumed to differ from the presumed steering vector $\mathbf{s}$ in the data model of the observations exposed in (1.1). Namely, recall that according to (1.1),

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

Herein the actual steering vector is assumed to be related to the presumed steering vector through the model

$$\tilde{\mathbf{s}} = \mathbf{s} + \boldsymbol{\Delta},$$

where $\boldsymbol{\Delta}$ is an unknown distortion vector. In practice, the distortions of the presumed steering vector arise from array calibration errors, errors when pointing towards the signal of interest, source spreading due to local scattering or source wavefront distortions due to environmental inhomogeneities, among other reasons [37]. Recall that the mismatch between $\tilde{\mathbf{s}}$ and $\mathbf{s}$ leads to a degradation in the MVDR due to the signal cancelation effect, which provokes that the MVDR may confuse the SOI with an interferer and thus may tend to cancel it, see [37] and references therein.

In the sequel, it is assumed that both $\mathbf{s}$ and $\tilde{\mathbf{s}}$ lie within an uncertainty set. Herein, it is assumed that the error between $\mathbf{s}$ and $\tilde{\mathbf{s}}$ describes a spherical uncertainty region, or in other words its norm is bounded. This yields the next uncertainty region, which was originally proposed in [37],

$$\mathcal{A}(\varepsilon) = \{\mathbf{a} | \mathbf{a} = \mathbf{s} + \mathbf{e}, \|\mathbf{e}\| \leq \varepsilon\}. \tag{5.1}$$

where $\varepsilon \in [0, \|\mathbf{s}\|)$ is a user parameter assumed to be known, see [37] [47]. Note that in the literature other types of uncertainty region have been proposed, e.g. [46] and [47] considered

ellipsoidal uncertainty sets, nonetheless this leads to assume a more sophisticated prior information, i.e. more a priori known model parameters [80]. Now recall that the shrinkage MVDR problem seeks to maximize the SINR for the type of shrinkage filter $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, which permits to deal with the small sample size degradation, and subject to a no distortion constraint towards the SOI signature vector. Therefore, the shrinkage MVDR can be made robust to uncertainties in $\mathbf{s}$ by modifying the no distortion constraint towards $\mathbf{s}$. Namely, the new robust shrinkage MVDR method proposed in this chapter arises from maximizing the SINR subject to a no distortion constraint for all the steering vectors that lie within the uncertainty set $\mathcal{A}(\varepsilon)$, for the type of filter $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$. This problem was stated rigorously in chapter 1 in (1.30) and it is given by,

$$\begin{aligned} \hat{x}(n) = \mathbf{w}^H \mathbf{y}(n); \quad \mathbf{w} = \arg \min_{\mathbf{w}} \quad & \mathbf{w}^H \mathbf{R} \mathbf{w} \\ \text{subject to} \quad & |\mathbf{w}^H \mathbf{a}| \geqslant 1 \text{ for all } \mathbf{a} \in \mathcal{A}(\varepsilon), \\ & \mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}. \end{aligned} \tag{5.2}$$

Note that this problem seeks to maximize the SINR for the type of the proposed shrinkage filters with the constraint that in the worst case the no distortion constraint will be maintained for any of the steering vectors belonging to the uncertainty region $\mathcal{A}(\varepsilon)$, i.e. for the particular vector $\mathbf{a}$ which yields the minimum value of $|\mathbf{w}^H \mathbf{a}|$. The type of no distortion constraint $|\mathbf{w}^H \mathbf{a}| \geqslant 1$ for all $\mathbf{a} \in \mathcal{A}(\varepsilon)$, which permits to treat the uncertainties in $\mathbf{s}$, was originally proposed in [37], see (1.23). However, unlike in [37] herein the small sample size is dealt with directly, as the type of filter $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$ is considered. Furthermore, in this regard, note that the theoretical covariance $\mathbf{R}$ is considered in the objective function, unlike in [37] which considered the SCM $\hat{\mathbf{R}}$, see (1.23). In fact, [37] has some robustness to the small sample size regime, as their filter can be interpreted in terms of a DL, $\mathbf{w} \propto (\hat{\mathbf{R}} + \lambda \varepsilon^2 \mathbf{I})^{-1}\mathbf{s}$, where the Lagrange multiplier λ cannot be obtained in closed form [37]. However, the robustness to the small sample size depends on the loading factor, which just depends on the parameter ε, which is designed to cope with the uncertainty of the SOI signature but not to deal directly with the small sample size. Thereby, [37] may require in some situations an additional parameter tuning of ε to counteract properly the small sample size. This parameter tuning is avoided thanks to the proposed approach as it will be shown below. Thereby, thanks to the proposed approach both the small sample size degradation and the uncertainties in $\mathbf{s}$ can be dealt with directly.

In order to proceed, consider the notation $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$, $\boldsymbol{\Omega} = (\hat{\mathbf{R}}^{-1}\mathbf{s}, \mathbf{s})$. This permits to express the filter $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$ as $\mathbf{w} = \boldsymbol{\Omega}\boldsymbol{\alpha}$. By taking this notation into account, the optimization problem (5.2) can be rewritten as

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{R} \boldsymbol{\Omega} \boldsymbol{\alpha} \\ \text{subject to} \quad & |\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{a}| \geqslant 1 \text{ for all } \mathbf{a} \in \mathcal{A}(\varepsilon), \\ & \mathcal{A}(\varepsilon) = \{\mathbf{a} | \mathbf{a} = \mathbf{s} + \mathbf{e}, \|\mathbf{e}\| \leq \varepsilon\}. \end{aligned} \tag{5.3}$$

Equation (5.3) highlights that the shrinkage coefficients in $\boldsymbol{\alpha}$ have two roles. First, they must implement the corrections of the sample MVDR that permit to deal with the small sample size regime. Second, they guarantee the no distortion constraint for all the steering vectors within the uncertainty region $\mathcal{A}(\varepsilon)$ and as a consequence permit to make robust the shrinkage MVDR to uncertainties in the presumed steering vector $\mathbf{s}$. On the other hand, note that in order to solve the problem (5.3) there are two issues to be faced. The first one, is that the objective function depends on the unknown $\mathbf{R}$. Or more precisely, the matrix $\boldsymbol{\Omega}^H \mathbf{R} \boldsymbol{\Omega}$ yields several quantities depending on $\mathbf{R}$. To counteract this problem, a RMT approach is proposed in the next section. This approach will permit to obtain consistent estimates of the unknown quantities within the asymptotic regime where $M, N \to \infty$ and $M/N \to c \in (0, 1)$. As it has mentioned in the previous chapters, this asymptotic regime permits to deal naturally with the small sample size situation. Thereby, thanks to the shrinkage structure of the proposed filter and the RMT approach, the proposed shrinkage MVDR filter in this chapter deals directly with the small sample size degradation. The second issue in (5.3) is that $|\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{a}| \geqslant 1$ is a nonlinear and nonconvex constraint on $\boldsymbol{\alpha}$ for each vector $\mathbf{a} \in \mathcal{A}(\varepsilon)$. Therefore, (5.3) belongs to the class of nonconvex quadratically constrained quadratic programming problems, which are known to be intractable [37]. In order to face this problem, in the next section it will be shown that (5.3) can be reformulated as a SOCP program and as a consequence it can be solved efficiently, or in polynomial time, by means of interior point methods [131].

5.4 Solution to the robust shrinkage MVDR based on RMT and the reformulation as a SOCP

In this section, all the necessary derivations to obtain the desired robust shrinkage MVDR filter are presented. To this end, the problems posed by (5.3) are faced next. The first problem that is treated is the dependence on the unknown data covariance $\mathbf{R}$ of the objective function in (5.3). In order to proceed, first the objective function is examined in more detail. Note that considering the definition of $\boldsymbol{\Omega} = (\hat{\mathbf{R}}^{-1}\mathbf{s}, \mathbf{s})$, the objective function can be expressed as

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

and using the property of multiplication of partitioned matrices, the last equation can be rewritten as follows,

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

This expression highlights that the objective function of the robust shrinkage MVDR problem in (5.3) depends on the unknown data covariance $\mathbf{R}$ through the quantities $\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}, \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s}, \mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s}$ and $\mathbf{s}^H\mathbf{R}\mathbf{s}$. In order to circumvent this problem and to obtain a solution that is still robust to the finite sample size, a RMT approach is proposed. Namely, the (M, N)-consistent estimate of these quantities is desired. This permits to obtain estimators that convergence to those functions of the unknown $\mathbf{R}$ within the asymptotic regime where $M, N \to \infty$ and $M/N \to c \in (0, 1)$, which deals implicitly with the finite sample size situation. To obtain the desired estimates, Lemma 2.1 is taken into account. Thereby, considering that $M, N \to \infty$ and $M/N \to c \in (0, 1)$ one obtains the next consistent estimate for the inner matrix of the objective function of (5.3)

$$\begin{pmatrix} \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} & \mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{R}\mathbf{s} \\ \mathbf{s}^H\mathbf{R}\hat{\mathbf{R}}^{-1}\mathbf{s} & \mathbf{s}^H\mathbf{R}\mathbf{s} \end{pmatrix} \asymp \begin{pmatrix} (1-c)^{-2}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} & (1-c)^{-1}\mathbf{s}^H\mathbf{s} \\ (1-c)^{-1}\mathbf{s}^H\mathbf{s} & \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s} \end{pmatrix} \triangleq \mathbf{B}. \tag{5.4}$$

Therefore, the first issue of problem (5.3), i.e. the dependence on $\mathbf{R}$, is solved by substituting the objective function by its (M, N)-consistent estimate obtained thanks to (5.4). This leads to the next optimization problem,

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \boldsymbol{\alpha}^H\mathbf{B}\boldsymbol{\alpha} \\ \text{subject to} \quad & |\boldsymbol{\alpha}^H\boldsymbol{\Omega}^H\mathbf{a}| \geqslant 1 \text{ for all } \mathbf{a} \in \mathcal{A}(\varepsilon), \\ & \mathcal{A}(\varepsilon) = \{\mathbf{a} | \mathbf{a} = \mathbf{s} + \mathbf{e}, \|\mathbf{e}\| \leq \varepsilon\}. \end{aligned} \tag{5.5}$$

Next, the second issue of (5.3) must be addressed, i.e. the fact that it is an intractable nonconvex quadratically constrained quadratic programming problem. To circumvent this issue the procedure is similar to the one proposed in [37]. Namely, first observe that the objective of (5.3) is to maximize the SINR subject to the no distortion constraint for all

the steering vectors $\mathbf{a} \in \mathcal{A}(\varepsilon)$, i.e. subject to the constraint $|\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{a}| \geqslant 1$ for all $\mathbf{a} \in \mathcal{A}(\varepsilon)$. This is equivalent to force that the no distortion constraint is fulfilled for the smallest value of $|\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{a}| \geqslant 1$. Therefore, the constraint $|\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{a}| \geqslant 1$ for all $\mathbf{a} \in \mathcal{A}(\varepsilon)$ can be rewritten in terms of this next expression,

$$\min_{\mathbf{a} \in \mathcal{A}(\varepsilon)} |\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{a}| \geqslant 1. \tag{5.6}$$

Moreover, recalling the definition of the uncertainty set $\mathcal{A}(\varepsilon) = \{\mathbf{a} | \mathbf{a} = \mathbf{s} + \mathbf{e}, \|\mathbf{e}\| \leq \varepsilon\}$, last equation can be rewritten as

$$\min_{\mathbf{e} \in \mathcal{D}(\varepsilon)} |\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{s} + \boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{e}| \geqslant 1, \text{where } \mathcal{D}(\varepsilon) = \{\mathbf{e} \mid \|\mathbf{e}\| \leq \varepsilon\}.$$

Next, in order to proceed, the Cauchy-Schwartz inequality, the triangle inequality and the relation $\|\mathbf{e}\| \leq \varepsilon$ are applied to obtain the next relations, [1]

$$|\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{s} + \boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{e}| \geqslant |\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{s}| - |\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{e}| \geqslant |\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{s}| - \varepsilon \|\boldsymbol{\Omega} \boldsymbol{\alpha}\| \tag{5.7}$$

Therefore, taking into account the equivalence between the constraint in (5.5) and (5.6) and the lower bound (5.7), the robust shrinkage MVDR problem (5.5) can be rewritten as follows,

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \boldsymbol{\alpha}^H \mathbf{B} \boldsymbol{\alpha} \\ \text{subject to} \quad & |\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{s}| - \varepsilon \|\boldsymbol{\Omega} \boldsymbol{\alpha}\| \geqslant 1. \end{aligned} \tag{5.8}$$

However, note that this problem is still nonconvex due to the presence of the absolute value operator. In order to circumvent this problem, note that the cost function in (5.8) is not changed if an arbitrary phase rotation of $\boldsymbol{\alpha}$ is forced. Thus, a new constraint can be incorporated in the problem to force a phase rotation of $\boldsymbol{\alpha}$ so that $|\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{s}|$ is a real number. Namely, the constraints $\text{Re}\{\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{s}\} \geqslant 0$ and $\text{Im}\{\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{s}\} = 0$ can be forced. This yields the next constraints for (5.8)

[1] The first inequality is obtained by observing that two complex numbers x and y fulfill the next relation, according to the triangular inequality, $|x| = |x + y - y| \leqslant |x + y| + |-y| = |x + y| + |y|$. Regarding the second inequality, using Cauchy-Scwartz inequality $-|\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{e}| \geqslant -\|e\| \|\boldsymbol{\Omega} \boldsymbol{\alpha}\|$ and noting that $\|\mathbf{e}\| \leq \varepsilon$ one obtains $-|\boldsymbol{\alpha}^H \boldsymbol{\Omega}^H \mathbf{e}| \geqslant -\varepsilon \|\boldsymbol{\Omega} \boldsymbol{\alpha}\|$.

$$\begin{aligned} &\boldsymbol{\alpha}^H\boldsymbol{\Omega}^H\mathbf{s} \geqslant \varepsilon\|\boldsymbol{\Omega}\boldsymbol{\alpha}\| + 1 \\ &\mathrm{Re}\{\boldsymbol{\alpha}^H\boldsymbol{\Omega}^H\mathbf{s}\} \geqslant 0 \\ &\mathrm{Im}\{\boldsymbol{\alpha}^H\boldsymbol{\Omega}^H\mathbf{s}\} = 0. \end{aligned} \tag{5.9}$$

Moreover, note that the second constraint in (5.9) is unnecessary if one takes into account the first constraint in (5.9). Therefore, taking into account this fact, (5.8) can be rewritten in the next form,

$$\begin{aligned} \underset{\boldsymbol{\alpha}}{\text{minimize}} \quad & \boldsymbol{\alpha}^H\mathbf{B}\boldsymbol{\alpha} \\ \text{subject to} \quad & \boldsymbol{\alpha}^H\boldsymbol{\rho} \geqslant \varepsilon\|\boldsymbol{\Omega}\boldsymbol{\alpha}\| + 1, \\ & \mathrm{Im}\{\boldsymbol{\alpha}^H\boldsymbol{\rho}\} = 0. \end{aligned} \tag{5.10}$$

Where, $\mathbf{B} = \begin{pmatrix} (1-c)^{-2}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} & (1-c)^{-1}\mathbf{s}^H\mathbf{s} \\ (1-c)^{-1}\mathbf{s}^H\mathbf{s} & \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s} \end{pmatrix}$, $\boldsymbol{\rho} = (\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}, \mathbf{s}^H\mathbf{s})^T$, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$ and $\boldsymbol{\Omega} = (\hat{\mathbf{R}}^{-1}\mathbf{s}, \mathbf{s})$. The next step is to express (5.10) in the form of a SOCP problem, to show that it can be solved efficiently, i.e. in polynomial time. Taking into account the eigendecomposition of $\mathbf{B}$

$$\mathbf{B} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^H \triangleq \mathbf{U}^H\mathbf{U},$$

one can express the objective function as $\boldsymbol{\alpha}^H\mathbf{B}\boldsymbol{\alpha} = \|\mathbf{U}\boldsymbol{\alpha}\|^2$, which is equivalent to minimize $\|\mathbf{U}\boldsymbol{\alpha}\|$ [37]. Thus, introducing a new positive variable τ, (5.10) can be expressed as,

$$\begin{aligned} \underset{\tau,\boldsymbol{\alpha}}{\text{minimize}} \quad & \tau \\ \text{subject to} \quad & \|\mathbf{U}\boldsymbol{\alpha}\| \leqslant \tau \\ & \varepsilon\|\boldsymbol{\Omega}\boldsymbol{\alpha}\| \leqslant \boldsymbol{\alpha}^H\boldsymbol{\rho} - 1 \\ & \mathrm{Im}\{\boldsymbol{\alpha}^H\boldsymbol{\rho}\} = 0. \end{aligned} \tag{5.11}$$

Now, the next variables are introduced to rewrite (5.11) in real valued form and to facilitate its interpretation as a SOCP,

$$\breve{\boldsymbol{\alpha}} = (\mathrm{Re}\{\boldsymbol{\alpha}\}^T, \mathrm{Im}\{\boldsymbol{\alpha}\}^T)^T \quad \breve{\mathbf{U}} = \begin{pmatrix} \mathrm{Re}\{\mathbf{U}\} & -\mathrm{Im}\{\mathbf{U}\} \\ \mathrm{Im}\{\mathbf{U}\} & \mathrm{Re}\{\mathbf{U}\} \end{pmatrix}$$
$$\breve{\boldsymbol{\Omega}} = \begin{pmatrix} \mathrm{Re}\{\boldsymbol{\Omega}\} & -\mathrm{Im}\{\boldsymbol{\Omega}\} \\ \mathrm{Im}\{\boldsymbol{\Omega}\} & \mathrm{Re}\{\boldsymbol{\Omega}\}\} \end{pmatrix} \quad \breve{\boldsymbol{\rho}} = (\mathrm{Re}\{\boldsymbol{\rho}\}^T, \mathrm{Im}\{\boldsymbol{\rho}\}^T)^T \tag{5.12}$$
$$\bar{\boldsymbol{\rho}} = (\mathrm{Im}\{\boldsymbol{\rho}\}^T, -\mathrm{Re}\{\boldsymbol{\rho}\}^T)^T$$

These, definitions permit to rewrite (5.11) as

$$\begin{aligned} \underset{\tau, \breve{\boldsymbol{\alpha}}}{\text{minimize}} \quad & \tau \\ \text{subject to} \quad & \|\breve{\mathbf{U}}\breve{\boldsymbol{\alpha}}\| \leqslant \tau \\ & \varepsilon\|\breve{\boldsymbol{\Omega}}\breve{\boldsymbol{\alpha}}\| \leqslant \breve{\boldsymbol{\alpha}}^T\breve{\boldsymbol{\rho}} - 1 \\ & \breve{\boldsymbol{\alpha}}^T\bar{\boldsymbol{\rho}} = 0. \end{aligned} \tag{5.13}$$

Now, let us introduce the next variables

$$\mathbf{y} = (\tau, \breve{\boldsymbol{\alpha}}^T)^T, \quad \mathbf{d} = (1, \mathbf{0}^T)^T. \tag{5.14}$$

These new variables permit to rewrite (5.13) as the next problem, which has the standard form of a SOCP [132],

$$\begin{aligned} \underset{\mathbf{y}}{\text{minimize}} \quad & \mathbf{d}^T\mathbf{y} \\ \text{subject to} \quad & \|(\mathbf{0}, \breve{\mathbf{U}})\mathbf{y}\| \leqslant \left(1, \mathbf{0}^T\right)\mathbf{y} \\ & \|(\mathbf{0}, \varepsilon\breve{\boldsymbol{\Omega}})\mathbf{y}\| \leqslant (0, \breve{\boldsymbol{\rho}}^T)\mathbf{y} - 1 \\ & (0, \bar{\boldsymbol{\rho}}^T)\mathbf{y} = 0. \end{aligned} \tag{5.15}$$

Where, recall that according to (5.14) $\mathbf{y} = (\tau, \breve{\boldsymbol{\alpha}}^T)^T \in \mathbb{R}^{5\times 1}$, $\mathbf{d} = (1, \mathbf{0}^T)^T \in \mathbb{R}^{5\times 1}$. Moreover, $\breve{\boldsymbol{\alpha}} \in \mathbb{R}^{4\times 1}$, $\breve{\mathbf{U}} \in \mathbb{R}^{4\times 4}$, $\breve{\boldsymbol{\Omega}} \in \mathbb{R}^{2M\times 4}$, $\breve{\boldsymbol{\rho}} \in \mathbb{R}^{4\times 1}$ and $\bar{\boldsymbol{\rho}} \in \mathbb{R}^{4\times 1}$ were defined in (5.12). Therefore, as (5.15) has the standard form of a SOCP (see (5.20) or [132]), it has been shown that the proposed robust shrinkage MVDR can be solved efficiently, in polynomial time, by means of software tools relying on interior point methods [131]. Namely, the computational cost of the proposed method is $\mathcal{O}(M^3)$, as it is shown in the appendix 5.A, which is the same than the cost of the sample MVDR (1.9) and the robust Capon filter in [37] based on a worst case performance optimization. Even more important, the cost is much smaller than the recent robust beamforming technique based on interference covariance

matrix reconstruction [58], which is robust to both steering uncertainties and covariance mismatches. The estimate of the interference covariance in that method, depends on evaluating an expression in a number of grid points S of the angular sector where the interference lies, which leads [58] to have a cost $\mathcal{O}(SM^2)$ where $S \gg M$. Moreover, note that thanks to the proposed shrinkage structure of the filter and to the RMT approach introduced through (5.4), the proposed approach deals directly with the small sample size regime. As a consequence, this permits to avoid the ad-hoc parameter tuning of ε that the robust Capon filter in [37] requires to counteract the covariance mismatches properly. That is, in the proposed robust Shrinkage MVDR, ε can concentrate on treating the steering vector mismatches, which was its original aim when it was introduced in the set $\mathcal{A}(\varepsilon)$ in (5.1) to model the uncertainty of the presumed steering vector.

5.5 Numerical Simulations

Next, the performance of the robust shrinkage MVDR filter proposed in this chapter, in the expression (5.15), is compared to other robust filters. To this end, the application considered is robust beamforming in array processing, thereby the proposed filter is called in the sequel robust shrinkage MVDR or Capon Beamformer (RSCB). For comparison purposes, the next robust beamformers are considered. First, the robust Capon beamformer (RCB) proposed in [37] and exposed above in (1.24). Recall that this beamformer is robust to uncertainties of the steering vector by incorporating an uncertainty region to the Capon beamformer, which leads to a worst-case optimization of the SINR, though it does not deal directly with the finite sample size effect. The second type of beamformers, which are considered for simulation purposes, are robust to the small sample size, though assuming a known steering vector. These are, on the one hand, the proposed shrinkage MVDR beamformer in (4.4), which is called Shrinkage MVDR in the sequel of this section. On the other hand, the DL which optimizes asymptotically the SINR, for a known steering vector, i.e. (1.15) proposed in [9]. Moreover, recall that the proposed RSCB and the RCB in [37] have the form of a convex optimization problem, namely they can be expressed as a SOCP. In order to implement these SOCP problems, the CVX software package was used in Matlab [133].

Next, the simulation conditions are described. A uniform linear array with $M = 50$ sensors spaced half wavelength is considered. Also, without loss of generality $\gamma = 1$ in (1.1) and a set of K narrowband interferers are considered. Thereby, $\mathbf{R_n}$ in (1.1) reads,

$$\mathbf{R_n} = \mathbf{APA}^H + \sigma^2\mathbf{I}$$

where σ^2 is the power of an AWGN, and its value will be given below. $\mathbf{A} \in \mathbb{C}^{M\times K}$ stacks in its columns the steering vector of the interferers, whose Direction Of Arrival (DOA) is specified below. $\mathbf{P}$ is a diagonal matrix with elements

$$\sigma_k^2 = \gamma 10^{-\mathrm{SIR}_k/10} \ \forall k = 1, \ldots, K.$$

Being SIR_k the ratio between the power of the SOI and the power of the k-th interferer and it is set to 0 dB. Finally, in order to generate the SCM, the data at the output of the array, the SOI, the signal of the $k-th$ interferer and the AWGN are generated, respectively, from the next iid complex gaussian distributions:

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

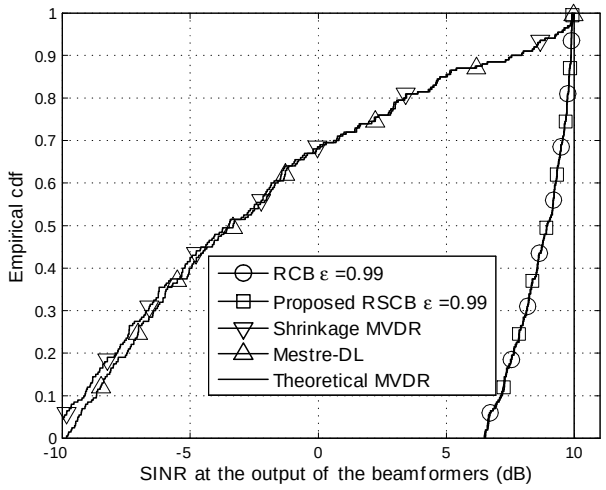

Figure 5.14: Comparison of the empirical cdf of the SINR at the output of the proposed RSCB (5.15), the RCB proposed in [37] see (1.24), the theoretical MVDR (1.7), Mestre-DL proposed in [9] see (1.15) and the shrinkage MVDR in (4.4), when the DOA mismatch lies within an uncertainty set and the sample size is large.

In figure 5.14 we study the effect of an error in the presumed steering vector when the errors due to the covariance mismatch are negligible and the steering vector error is assumed to be bounded within an uncertainty region. To this end, on the one hand a large

sample size regime is considered, namely $M = 50$ and $N = 5000$. Moreover, a steering error due to a DOA mismatch in the SOI is considered. More specifically, a maximum DOA mismatch of 1° is considered and at each iteration of the simulation the DOA mismatch is generated randomly according to a uniform distribution between 0° and 1°. Thereby, the DOA errors lead to an uncertainty set. The rest of the simulation parameters are SNR=10 dB, a SOI and three interferers whose DOA are 1°, 10°, -10° and 15°, respectively, and $\text{SIR}_i = 0$ dB for all the interferers. Finally, the parameter ε which defines the uncertainty set for the steering vector in both the proposed RSCB and the RCB, see (5.1) and [37], is set close to its maximum value, i.e. $\varepsilon = 0.99$.

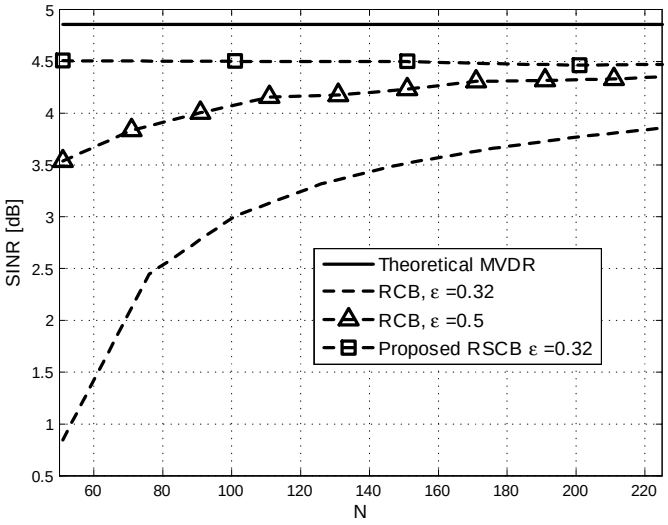

Figure 5.15: SINR comparison at the output of the proposed RSCB (5.15), the RCB proposed in [37] see (1.24) and the theoretical MVDR (1.7), when the error in the presumed DOA of the SOI is relatively small. Exemplification that, unlike the proposed RSCB, the RCB requires an additional heuristic tuning of ε to counteract properly the small sample size effect.

In figure 5.14 the empirical cumulative distribution function (CDF) of the SINR at the output of the beamformers is plotted. Note that the variability leading to plot the empirical cdf is due to consider a variable DOA mismatch at each iteration of the simulation. It can be observed that both the proposed RSCB (5.15) and the RCB deal properly with this situation as their design considers an error in the presumed steering vector which varies within an uncertainty set. Moreover, they obtain the same performance as the errors are just due to the steering vector mismatch. That is, recall that the advantage of the proposed RSCB over the RCB is that it can deal explicitly with both the mismatch in the covariance,

due to the small sample size, and the mismatch in the steering vector, whereas the RCB is designed to cope with the steering vector mismatch but its protection against the finite sample size is not directly tackled. On the other hand, the asymptotically optimal DL-MVDR [9] and the proposed shrinkage MVDR (4.4) undergo a performance degradation as they rely on a known steering vector. Moreover, note that both offer the same performance as in the large sample size regime both tend to the sample MVDR, which on its turn tends to an MVDR implemented with the true covariance matrix and the presumed steering vector, which differs from the actual steering vector.

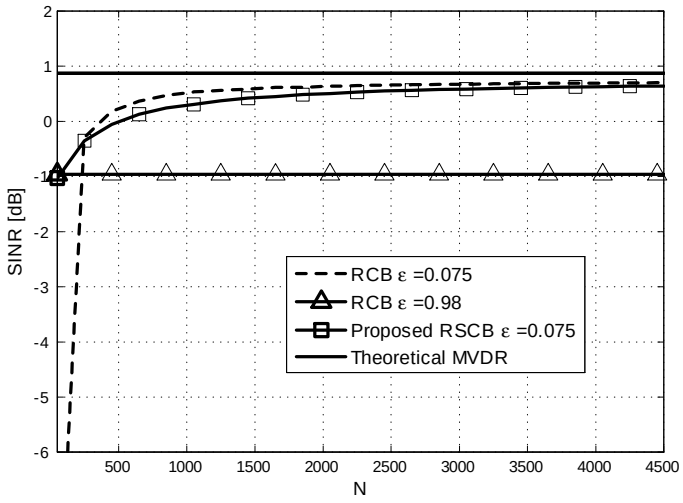

Figure 5.16: SINR comparison at the output of the proposed RSCB (5.15), the RCB proposed in [37] see (1.24) and the theoretical MVDR (1.7), when the error in the presumed DOA of the SOI is relatively small. Exemplification that choosing ε too high, to counteract the small sample size degradation, may lead to a saturation effect.

Next, in figure 5.15 to 5.17 the beamformers are compared when the error between the presumed and actual DOA of the SOI are relatively small. Namely, in figure 5.15, a scenario with the SOI and three interferers is considered, whose DOA are $1°$, $4°$, $-10°$ and $15°$ respectively and the SNR=5 dB. A mismatch of $0.2°$ in the look direction is considered, which corresponds to a theoretical norm of the error between the presumed and the actual steering vector of $\varepsilon_o = \|\tilde{\mathbf{s}} - \mathbf{s}\| = 0.31$. This figure exemplifies the limitation of the RCB commented above, i.e. ε is designed to counteract the steering errors and it is not clear whether it is robust enough to covariance mismatches. Thereby, ε requires an additional heuristic tuning. In effect, one can observe in figure 5.15 that setting $\varepsilon = 0.32$ in the RCB,

which is a close value to ε_o, leads to an important performance degradation in the small sample size and an additional tuning is needed to counteract the covariance mismatch. Namely, observe how the RCB with $\varepsilon = 0.5$ yields better performance.

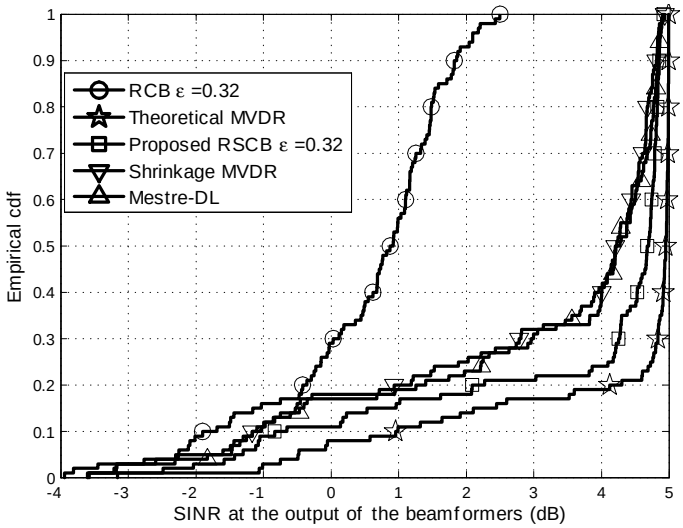

Figure 5.17: Comparison of the empirical cdf of the SINR at the output of the proposed RSCB (5.15), the RCB proposed in [37] see (1.24), the theoretical MVDR (1.7), Mestre-DL proposed in [9] see (1.15) and the shrinkage MVDR in (4.4), for a relatively small error in the presumed DOA of the SOI. The actual DOA of the SOI and the interferers are generated randomly at each iteration.

In fact, by observing figure 5.15, one may think that setting $\varepsilon \approx 1$, i.e. to its highest value, is always the correct choice, though in other situations this can provoke a saturation effect in the performance when N is increased, as it is shown in figure 5.16. The simulation parameters of figure 5.16 are a SOI and three interferers, whose DOA are 76°, 79°, 65° and 90° respectively, the SNR=5 dB and the look direction mismatch is 0.2°, which corresponds to $\varepsilon_o = 0.075$. On the other hand, as it can be observed in figures 5.15 and 5.16, the proposed RSCB does not require the tuning of ε, as it deals with the covariance mismatch thanks to the shrinkage structure of the beamformer and the RMT approach. These comments are confirmed in figure 5.17 for a more general scenario where the DOA of the desired signal and 7 interferers are generated randomly at each iteration, to this end a uniform distribution between $-90°$ and 90° is considered and the error in the presumed DOA of the SOI is 0.2° as in figures 5.15 and 5.16. In this simulation, the cdf of the SINR at the output of each beamformer is plotted for $M = 50$ and $N = 55$ (the rest of

the parameters are the same than in figure 5.15). For comparison purposes Mestre-DL (1.15) and the Shrinkage MVDR (4.4) have been included. They show small performance degradation as the error in the presumed DOA is relatively small and they are robust to the covariance mismatches.

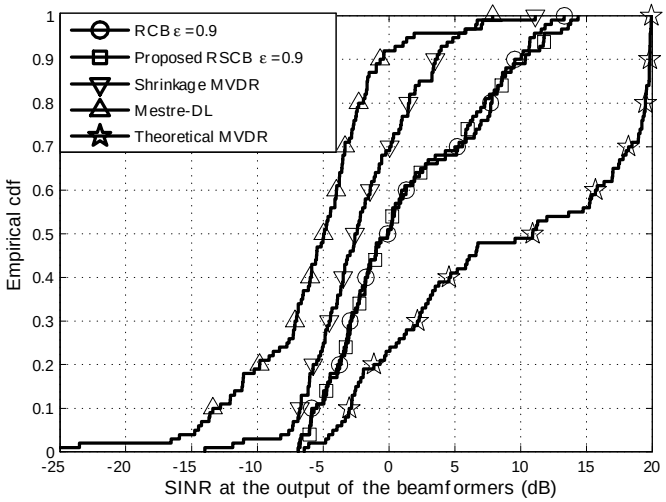

Figure 5.18: Comparison of the empirical cdf of the SINR at the output of the proposed RSCB (5.15), the RCB proposed in [37] see (1.24), the theoretical MVDR (1.7), Mestre-DL proposed in [9] see (1.15) and the shrinkage MVDR in (4.4), for a significant error in the presumed DOA of the SOI. The actual DOA of the SOI and the interferers are generated randomly at each iteration.

Nonetheless, as figure 5.18 highlights, if the error in the presumed DOA increases, Mestre-DL (1.15) and the Shrinkage MVDR (4.4) undergo a significant degradation. The simulation parameters of figure 5.18 are M=50, N=55, a SOI and 28 interferers whose DOA is generated randomly at each iteration according to a uniform distribution between $-90°$ and $90°$, the SNR=20 dB and a look direction mismatch of $1°$. Figure 5.18 highlights another advantage of the proposed RSCB, as it is more robust to steering errors than Mestre-DL and the Shrinkage MVDR. In fact this was one of the objectives of this chapter and it has been achieved thanks to the new formulation of the proposed RSCB, which incorporates an uncertainty region for the steering vector in the MVDR optimization problem and it requires a no distortion constraint for all the steering vectors within that uncertainty set. On the other hand, both Mestre-DL and the Shrinkage MVDR presume a known steering vector. Moreover, it is also interesting to observe how the shrinkage

MVDR proposed in this thesis in (4.4) is more robust to significant steering vector errors than Mestre-DL proposed in [9]. The reason is that Mestre-DL obtains an asymptotically optimal diagonal loading factor to deal for the small sample size regime, but it requires an additional tuning of the loading factor to deal with steering vector uncertainties. On the other hand, the proposed shrinkage MVDR in (4.4) tends to a scaling of the conventional beamformer in the small sample size regime. Therefore, as the main lobe of the conventional beamformer is wider than the one of the DL-MVDR it is less sensitive to uncertainties of the steering vector. A final important comment related to figure 5.18 is described next. When the steering vector errors increase, they may dominate over the covariance mismatches and the tuning of ε in the RCB is less significant. Both the proposed RSCB and the RCB deal properly with this situation with a high ε and they offer almost the same performance.

5.A Appendix: Computational Cost of the proposed method

Next, the computational cost of the proposed robust shrinkage MVDR method in (5.15) is analyzed. To this end, recall that (5.15) is a SOCP, thereby the cost of the proposed method is determined by the cost of solving a SOCP and the cost of forming the matrices that give form to that SOCP, i.e. to (5.15).

Regarding the cost of forming the matrices of (5.15), first observe that this optimization problem is formed by the next matrices, which were defined in (5.12) and (5.14)

$$\begin{array}{lll} \mathbf{y} = (\tau, \breve{\boldsymbol{\alpha}}^T)^T \in \mathbb{R}^{5\times 1}, & \mathbf{d} = (1, \mathbf{0}^T)^T \in \mathbb{R}^{5\times 1} & \breve{\boldsymbol{\alpha}} \in \mathbb{R}^{4\times 1}, \\ \breve{\mathbf{U}} \in \mathbb{R}^{4\times 4}, & \breve{\boldsymbol{\Omega}} \in \mathbb{R}^{2M\times 4}, & \breve{\boldsymbol{\rho}} \in \mathbb{R}^{4\times 1} \\ \bar{\boldsymbol{\rho}} \in \mathbb{R}^{4\times 1}. & & \end{array} \tag{5.16}$$

On its turn, according to (5.12), the expression of $\breve{\boldsymbol{\alpha}}, \breve{\mathbf{U}}, \breve{\boldsymbol{\Omega}}, \breve{\boldsymbol{\rho}}$ and $\bar{\boldsymbol{\rho}}$ is,

$$\begin{array}{ll} \breve{\boldsymbol{\alpha}} = (\mathrm{Re}\{\boldsymbol{\alpha}\}^T, \mathrm{Im}\{\boldsymbol{\alpha}\}^T)^T & \breve{\mathbf{U}} = \begin{pmatrix} \mathrm{Re}\{\mathbf{U}\} & -\mathrm{Im}\{\mathbf{U}\} \\ \mathrm{Im}\{\mathbf{U}\} & \mathrm{Re}\{\mathbf{U}\} \end{pmatrix} \\ \breve{\boldsymbol{\Omega}} = \begin{pmatrix} \mathrm{Re}\{\boldsymbol{\Omega}\} & -\mathrm{Im}\{\boldsymbol{\Omega}\} \\ \mathrm{Im}\{\boldsymbol{\Omega}\} & \mathrm{Re}\{\boldsymbol{\Omega}\}\} \end{pmatrix} & \breve{\boldsymbol{\rho}} = (\mathrm{Re}\{\boldsymbol{\rho}\}^T, \mathrm{Im}\{\boldsymbol{\rho}\}^T)^T \\ & \bar{\boldsymbol{\rho}} = (\mathrm{Im}\{\boldsymbol{\rho}\}^T, -\mathrm{Re}\{\boldsymbol{\rho}\}^T)^T \end{array} \tag{5.17}$$

Moreover, recall that the expressions for $\boldsymbol{\Omega}$ and $\boldsymbol{\rho}$ are given by

$$\boldsymbol{\Omega} = (\hat{\mathbf{R}}^{-1}\mathbf{s}, \mathbf{s}) \quad \boldsymbol{\rho} = (\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s}, \mathbf{s}^H\mathbf{s})^T. \tag{5.18}$$

And also remember that $\mathbf{U}$ arises from the eigendecomposition of $\mathbf{B}$, i.e. $\mathbf{B} = \mathbf{E}\boldsymbol{\Lambda}\mathbf{E}^H = \mathbf{U}^H\mathbf{U}$, where the expression for $\mathbf{B}$ is given by

$$\mathbf{B} = \begin{pmatrix} (1-c)^{-2}\mathbf{s}^H\hat{\mathbf{R}}^{-1}\mathbf{s} & (1-c)^{-1}\mathbf{s}^H\mathbf{s} \\ (1-c)^{-1}\mathbf{s}^H\mathbf{s} & \mathbf{s}^H\hat{\mathbf{R}}\mathbf{s} \end{pmatrix} \tag{5.19}$$

Therefore, by inspecting equations (5.16) to (5.19) one infers that the cost of forming the matrices required by the optimization problem (5.15) is due to the cost of computing $\hat{\mathbf{R}}^{-1}$, which is well known to be $\mathcal{O}(M^3)$.

Now, in order to know what is the overall cost of the proposed method in (5.15), it remains to know what is the cost of computing the SOCP involved in (5.15). To this end, bear in mind that the standard form of a SOCP is, see e.g. [132],

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{f}^T\mathbf{x} \\ \text{s.t.} \quad & \|\mathbf{A}_i\mathbf{x} + \mathbf{b}_i\| \leqslant \mathbf{c}_i^T\mathbf{x} + \mathbf{d}_i, \ i = 1, \ldots, N. \end{aligned} \tag{5.20}$$

Where $\mathbf{f} \in \mathbb{R}^n$, $\mathbf{A}_i \in \mathbb{R}^{n_i - 1 \times n}$, $\mathbf{b}_i \in \mathbb{R}^{n_i - 1}$, $\mathbf{c}_i \in \mathbb{R}^n$ and $\mathbf{d}_i \in \mathbb{R}$. According to [132], the cost of a SOCP is due to the cost of solving the system, which is $\mathcal{O}(n^3)$, and the cost of forming the matrix system, which is $\mathcal{O}(n^2 \sum_{i=1}^N n_i)$. In our case, observing (5.15) and comparing to the standard SOCP (5.20), it can be easily inferred that $n = 5, n_1 = 5, n_2 = 2M + 1$ and $n_3 = 2$. As a consequence, the cost of computing our SOCP is due to the cost of forming the matrix system, which is $\mathcal{O}(n^2 \sum_{i=1}^N n_i) = \mathcal{O}(5^2(5 + 2M + 1 + 2))$. Note, that this is smaller than the cost of forming the matrices of (5.15), which is $\mathcal{O}(M^3)$ according to equations (5.16) to (5.19). Therefore, the overall cost of the proposed robust shrinkage MVDR method in (5.15) is $\mathcal{O}(M^3)$.

Chapter 6

Conclusions and Future work

6.1 Conclusions

This thesis has dealt with the estimation of a parameter based on a linear filtering of the observations, when the parameter of interest is observed through a linear model. Moreover, this thesis considers the modern statistical framework, imposed by the current trend in e.g. array signal processing or wireless communications, where the dimension of the observations is comparable or even larger than the number of available statistical samples for the estimation purposes. Considering the MSE or SINR as the performance measures, the optimal linear estimators are the LMMSE or the MVDR, respectively. Unfortunately, these methods are not realizable as they depend on the inverse of the data covariance, i.e. the precision matrix. The conventional approach based on substituting the unknown covariance by its sample estimate, i.e. the SCM, leads to an important degradation within the statistical framework considered herein, where the sample size is small. Thereby, the aim of this thesis has been to propose corrections of the conventional sample LMMSE and MVDR filters to cope with the small sample size regime. To achieve this aim, several corrections based on the shrinkage estimation philosophy have been considered. The aim of those techniques is to reduce the MSE, of the methods to be corrected, by means of a linear transformation of the sample methods. Or more in general by means of an affine transformation which combines the sample methods with a priori information. Thereby, in this thesis shrinkage corrections of the sample LMMSE and MVDR filters have been considered, which may be summarized as particular cases of the general form of shrinkage filter $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, where $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I}$, $\hat{\mathbf{R}}$ is the SCM obtained from the observations, $\mathbf{s}$ is the signature vector associated to the SOI and α_1, α_2, δ are the parameters to be designed.

Namely, the corrections of the sample LMMSE that have been considered, from more simple to more elaborated, have been the following. A direct linear correction $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s}$ for the case that the sample size N and the observation dimension M fulfill the constraint $M/N \in (0,1)$. A linear correction of a regularized sample LMMSE, i.e. $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s}$, which permits to deal with any $M/N \in (0,\infty)$. Then, a more general shrinkage correction based on an affine transformation is considered. Namely, assuming that $M/N \in (0,1)$, the inverse of the SCM, which relies on the available statistical samples, is combined with an estimation of the precision matrix based on just a priori information, i.e. the identity matrix, which leads to $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s} = (\alpha_1 \hat{\mathbf{R}}^{-1} + \alpha_2 \mathbf{I})\mathbf{s}$. And finally, the previous filter is extended to deal with any $M/N \in (0,\infty)$ by considering a regularization of the SCM, i.e. the more general form of the shrinkage filter $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$ is treated. Regarding the sample MVDR, the shrinkage corrections that have been considered are based on the affine transformations mentioned above, i.e. the filters $\mathbf{w} = \alpha_1 \hat{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$ and $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$ have been considered.

The design of the shrinkage parameters α_1, α_2 and the regularization parameter δ is obtained as the optimization of the MSE, in the case of the LMMSE filters, or the maximization of the SINR, in the case of the MVDR, which can be viewed as the minimization of the MSE with a no distortion constraint. This leads to optimal α_1, α_2 and δ, though they depend on the unknown data covariance. To circumvent this problem, most of the proposed filters adopt a RMT approach which may be summarized as follows. First, an asymptotic regime which deals naturally with the small sample size regime is considered. Namely, this asymptotic regime considers that $M, N \to \infty$ and $M/N \to c$, where $c \in (0,1)$ in the more basic forms of shrinkage that consider $\delta = 0$ and $c \in (0,\infty)$ in the more general form of shrinkage filters with $\delta \neq 0$. Then, within this asymptotic regime, consistent estimators of the optimal shrinkage and regularization factors α_1, α_2 and δ are obtained. That is, the obtained estimators are asymptotically optimal as they tend to the optimal shrinkage estimators. This asymptotic approximation leads to good performance even for rather small values of M, N. Moreover, the RMT approach does not require any further assumption regarding the distribution of the observations.

The numerical simulations have shown that the proposed shrinkage filters clearly outperform the sample LMMSE and MVDR. Moreover, the next contributions are offered compared to the related work methods, which recall that are based on DL filters of the form $\mathbf{w} = (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1}\mathbf{s}$ or more in general on filters relying on a shrinkage of the SCM $\mathbf{w} = (\tau_1 \hat{\mathbf{R}} + \tau_2 \mathbf{I})^{-1}\mathbf{s}$, see section (1.4). Compared to the ad-hoc choices of the DL factor, the proposed shrinkage filters obtain better performance, as they obtain the shrinkage parameters that optimize asymptotically the metrics related to the estimation of the parameter of interest. This argument is also the reason to improve the filters based on the

LW shrinkage estimator of the SCM [33], as that method obtains a shrinkage which optimizes asymptotically the MSE in the estimation of the data covariance, which is not the final target. Compared to the DL method [9], which obtains the asymptotically optimal DL factor in terms of SINR, the next conclusions were observed. The proposed shrinkage filters based on an affine transformation of the inverse of the SCM, i.e. $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}+\alpha_2\mathbf{s}$ and $\mathbf{w} = \alpha_1\check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2\mathbf{s}$, obtain almost the same performance than [9] in terms of SINR and better performance than [9] in terms of MSE, when the proposed filters are designed as corrections of the sample LMMSE. Moreover, compared to the regularization of the LMMSE $\mathbf{w} = (\tau_1\hat{\mathbf{R}} + \tau_2\mathbf{I})^{-1}\mathbf{s}$ proposed in [79], which obtains the asymptotically optimal τ_1,τ_2 in terms of MSE, the next benefit was observed. The proposed shrinkage correction of the sample LMMSE, based on the filter $\mathbf{w} = \alpha_1\check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2\mathbf{s}$, obtains a performance gain in terms of MSE compared to [79], which is due to the additional shrinkage of the inverse covariance implemented through α_1, α_2. Both the proposed methods and the related work mentioned above assume that $\mathbf{s}$ is precisely known, though recall that another important source of degradation of the LMMSE and MVDR methods arises when there is an uncertainty in the signature vector $\mathbf{s}$. In those situations, the proposed shrinkage methods $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}+\alpha_2\mathbf{s}$ and $\mathbf{w} = \alpha_1\check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2\mathbf{s}$ outperform the related work methods [9] and [79]. The rationale is that the proposed methods consider the combination of the sample or regularized sample LMMSE and MVDR with a matched filter. Under and imprecise knowledge of $\mathbf{s}$, the matched filter may attenuate the SOI, but it will not try to cancel this signal as it was an interference, which is the behavior of the LMMSE and MVDR with a signature vector mismatch. On the contrary the DL [9] and regularization of the sample LMMSE [79] design the DL factor and shrinkage parameters to cope with the small sample size, but still undergo an important degradation due to the imprecise knowledge of $\mathbf{s}$. That is, they would require an additional tuning of the loading factor an shrinkage parameters to take into account the uncertainty in $\mathbf{s}$.

Finally, we propose a correction of the sample MVDR which is robust to both the degradation due to the small sample size and the degradation due to the imprecise knowledge of the signature vector **s.** To this end, our previous shrinkage filter $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}+\alpha_2\mathbf{s}$ is considered to deal with the small sample size impairment. Moreover, the MVDR problem formulation is modified by incorporating a no distortion constraint for all the vectors that are within an uncertainty region related to the signature vector. In this way, both the small sample size and the uncertainty in the signature vector are dealt with explicitly. The proposed filter outperforms our previous shrinkage MVDR filter $\mathbf{w} = \alpha_1\hat{\mathbf{R}}^{-1}\mathbf{s}+\alpha_2\mathbf{s}$ and the DL filter [9], which assume a known $\mathbf{s}$. Moreover, our work rather than [37] deals directly with the small sample size degradation, recall that [37] was the seminal work proposing the modification of the MVDR based on an uncertainty region for the signature vector.

Compared to [37], this leads to some performance gains in situations where the degradation due to the small sample size is more important than the degradation due to the uncertainty in $\mathbf{s}$.

6.2 Future work

This thesis has considered a statistical inference framework where the sample size support N is small or at most comparable to the system's dimension M. Within this context, topics for future research are described next.

6.2.1 Parameter estimation with low sample size support in a Massive MIMO context

Within the context of array signal processing and wireless communications one of the most promising and prominent technologies is massive MIMO [4], which consists of equipping the base stations (BS) with large arrays of antennas, namely several hundreds of antenna elements are envisaged. This will permit to cope with the increasing demand for higher spectral and energy efficiency in wireless systems [134]. Thereby, the statistical inference framework considered in this thesis fits perfectly within the massive MIMO context, as the sample size support for inference purposes is scarce compared to the system's dimension. More specifically, a possible problem to treat is the channel estimation based on an LMMSE methodology. In fact, pilot based channel estimation is a hot topic of research in massive MIMO [135], as due to the pilot contamination effect the performance of channel estimation and the spectral efficiency is known to be degraded. This effect refers to the spatially correlated intercell interference due to the pilot reuse in wireless cellular networks. The mathematical expression of the LMMSE channel estimator depends on the inverse covariance of the observations which on its turn depends on the covariance of the disturbance term, which includes the receiver noise and the intracell and intercell interference [135]. Thereby, these covariances must be estimated from a set of available snapshots, which will be small or comparable to the system's dimension. This paves the way to apply the statistical inference framework considered in this thesis.

The benefits of large antenna arrays for wireless cellular communications systems have been stressed by a great deal of research works in recent years. However, less attention has been paid to the use of this massive number of antennas at the fusion centre (FC) of sensor networks. In fact, massive MIMO can produce interesting benefits regarding the

energy efficiency of sensor networks. Namely, [136] shown that a constant detection and estimation performance can be achieved by decreasing the transmit power of the nodes by a factor $1/M$ and increasing to M the number of antennas at the fusion centre. Thereby, the channel estimation problem, mentioned above, is even more critical within a WSN framework characterized by a large number of antennas at the FC. Namely, note that the less the sensor nodes access the channel the better, as they are in general battery powered and the battery must have a duration of months or years. Thereby, the statistical inference framework dealt with in this thesis, where the sample size is small compared to the system's dimension is worth to be applied in this massive MIMO WSN framework.

6.2.2 Low complexity estimators

When the system's dimension increases it is important to device strategies to reduce the complexity of the proposed statistical inference methods. For instance, in massive MIMO or in general in other systems characterized by large arrays of antennas this is an important topic, see e.g. [135]. In fact, in that work they deal with the reduction of the LMMSE through a polynomial expansion technique, which has been widely applied in signal processing to reduce the complexity of the LMMSE in other systems such as in e.g. multiuser detection. The idea of this technique is to approximate large dimensional matrix inversions by an L-degree matrix polynomial, which leads to reduce the computational complexity from $\mathcal{O}(M^3)$ to $\mathcal{O}(LM^2)$. Another interesting technique is based on dimensionality reduction of the observed data through a random projection or random unitary transformation, see e.g. [8] [137]. Namely, the M-dimensional observed data is left multiplied by a random unitary matrix of dimensions $L \times M$, where $L \ll M$. In this regard, it is important to note that, according to the Johnson-Lindenstrauss lemma [138], an appropiately defined random projection preserves the length of the data vectors and the distance between vectors. A key question of this type of random projection techniques is to specify the distribution of the random unitary matrix. For instance in [8] they consider a isotropically random distribution or Haar measure whereas in [137] the entries of the random matrix are i.i.d. standard normal random variables, i.e. with zero mean and unit variance.

6.2.3 Non linear shrinkage estimation

As it has been mentioned in the conclusions section, the proposed shrinkage filters in this thesis can be summarized in the general form $\mathbf{w} = \alpha_1 \check{\mathbf{R}}^{-1}\mathbf{s} + \alpha_2 \mathbf{s}$, where $\check{\mathbf{R}} = \hat{\mathbf{R}} + \delta \mathbf{I}$, $\hat{\mathbf{R}}$ is the SCM obtained from the observations, $\mathbf{s}$ is the signature vector associated to the SOI

and α_1, α_2, δ are the shrinkage parameters to be designed. Moreover, the related work methods can be summarized as $\mathbf{w} = (\tau_1\hat{\mathbf{R}} + \tau_2\mathbf{I})^{-1}\mathbf{s}$. Thereby, after easy manipulations, based on the eigendecomposition of the SCM, it can be observed that both the proposed and the related work methods have the same eigenvectors than the SCM and apply a shrinkage correction to the sample eigenvalues. Namely, the proposed shrinkage filters can be rewritten as $\mathbf{w} = (\hat{\mathbf{E}}(\alpha_1(\hat{\mathbf{\Delta}} + \delta\mathbf{I})^{-1} + \alpha_2\mathbf{I})\hat{\mathbf{E}}^H)\mathbf{s}$ and the related work methods as $\mathbf{w} = (\hat{\mathbf{E}}(\tau_1\hat{\mathbf{\Delta}} + \tau_2\mathbf{I})^{-1}\hat{\mathbf{E}}^H)\mathbf{s}$. These expressions bring to light that the same shrinkage intensity is applied to every sample eigenvalue. A recent work [139], dealing with the estimation of large-dimensional covariance matrices, proposed to apply a different nonlinear shrinkage intensity to every sample eigenvalue. They showed that this approach provides important improvements over the linear shrinkage framework in situations where linear shrinkage does not improve enough the SCM, e.g. when the population eigenvalues are dispersed. Thereby, it would be interesting to study this type of nonlinear shrinkage estimation techniques in our framework. That is, our final target is to estimate a parameter which depends on functionals of the covariance, instead of the estimation of the covariance itself. Moreover, interesting enough, [139] not only builds on shrinkage estimation but also on large dimensional random matrix theory tools, which are the mathematical tools that have been used in this thesis. Namely, [139] focus on the type of estimators of the population covariance and the precision matrix which minimize the Frobenius norm by considering different shrinkage intensities for the sample eigenvalues. A brief description is provided next for the estimation of the covariance matrix, see [139] for the estimation of the precision matrix. Namely, the estimation is formulated as the next optimization problem,

$$\min_{\mathbf{D}} \|\hat{\mathbf{E}}\mathbf{D}\hat{\mathbf{E}}^T - \mathbf{R}\|_F$$

where $\hat{\mathbf{E}}$ denotes the matrix containing in its columns the sample eigenvectors $\{\hat{\mathbf{e}}_i\}_{i=1}^M$ and $\mathbf{D}$ a diagonal matrix containing the estimation of the eigenvalues. The optimal solution to this problem is given by $\mathbf{D} = \text{diag}(d_1, \ldots, d_M)$ where $d_i = \hat{\mathbf{e}}_i^T\mathbf{R}\hat{\mathbf{e}}_i$ for $i = 1, \ldots, M$. Building on recent RMT results which generalize the Marčenko-Pastur equation and considering that $M, N \to \infty$ and $M/N \to c \in (0, 1)$, it was shown in [140] that the optimal d_i can be approximated by the next quantity,

$$d_i^{or} = \frac{\hat{\lambda}_i}{|1 - c - c\hat{\lambda}_i\check{m}_F(\hat{\lambda}_i)|^2}$$

where $\hat{\lambda}_i$ is the i-th sample eigenvalue and $\check{m}_F(\hat{\lambda}_i)$ is defined in terms of the Stieltjes

transform of the limiting spectral distribution associated to the SCM $m_F(z)$ as follows, $\lim_{z \in \mathbb{C}^+ \to \lambda} m_F(z) = \check{m}_F(\lambda)\ \forall \lambda \in \mathbb{R} - \{0\}$. Note that $\check{m}_F(\hat{\lambda}_i)$ is not known as it is related to the limiting distribution of the sample eigenvalues not the observed one. However, the main contribution of [139] is precisely to provide a consistent estimation of $\check{m}_F(\hat{\lambda}_i)$ and thereby to provide a consistent estimation of d_i^{or}. Finally, observe that the estimation of the eigenvalues d_i^{or} is effectively a nonlinear shrinkage of the sample eigenvalues and that a different shrinkage intensity is applied for each of the eigenvalues.

References

[1] D.L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," in *Math challenges of the 21st century. American Mathematical Society*, 2000.

[2] A. Wiesel and A.O. Hero, "Distributed covariance estimation in gaussian graphical models," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 211–220, January 2012.

[3] R. Couillet and W. Hachem, "Fluctuations of spiked random matrix models and failure diagnosis in sensor networks," *IEEE Trans. Information Proc.*, vol. 59, no. 1, pp. 509–525, January 2013.

[4] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up mimo: Opportunities and challenges with very large arrays," *IEEE Signal Proces. Mag.*, vol. 30, no. 1, pp. 40–46, January 2013.

[5] L. Lu, G. Ye Li, A. Lee Swindlhehurst, A. Ashikhmin, and R. Zhang, "An overview of massive mimo: Benefits and challenges," *IEEE J. Selec. Topics Sign. Proc.*, vol. 8, no. 5, pp. 742–758, October 2014.

[6] J. Schäfer and K. Strimmer, "A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics," *Statist. Appl. Genet. Molec. Biol.*, vol. 4, no. 1, November 2005.

[7] A. Berge, A.C. Jensen, and A.H.S. Solberg, "Sparse inverse covariance estimates for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1399–1407, May 2007.

[8] T.L. Marzetta, G.H. Tucci, and S.H. Simon, "A random matrix-theoretic approach to handling singular covariance estimates," *IEEE Trans. Inform. Theory*, vol. 57, no. 9, pp. 6256–6270, September 2011.

[9] 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.

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

[11] 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.

[12] 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.

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

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

[15] Y. Eldar and A. Nehorai, "Mean squared error beamforming for signal estimation: a competitive approach," in *Robust Adaptive Beamforming*, J. Li and P. Stoica, Eds., pp. 259–298. Wiley-Interscience, John Wiley and Sons, 2005.

[16] Andrea Goldsmith, *Wireless Communications*, Cambridge University Press, New York, NY, USA, 2005.

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

[18] Y. Rong, Y.C. Eldar, and A.B. Gershman, "Performance tradeoffs among adaptive beamforming criteria," *IEEE Journal on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 651–659, December 2007.

[19] 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.

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

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

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

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

[24] 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.

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

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

[27] 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.

[28] 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.

[29] 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.

[30] 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.

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

[32] 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.

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

[34] 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.

[35] 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.

[36] Z.D. Bai and J. Silverstein, "No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices," *Annals of Probability*, vol. 26, no. 1, pp. 316–345, 1998.

[37] S. Vorobyov, A. Gershman, and Z.-Q. Luo, "Robust adaptive beamforming using worst-case performance optimization: a solution to the signal mismatch problem," *IEEE Transactions on Signal Processing*, vol. 51, no. 2, pp. 313–324, February 2003.

[38] 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.

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

[40] 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.

[41] 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.

[42] Y. Abramovich, "A controlled method for adaptive optimization of filters using the criterion of maximum snr," *Radiotekh. Elektron. (Radio Engineering Electronic Physics)*, vol. 26, no. 3, pp. 87–95, 1981.

[43] O. Cheremisin, "Efficiency of adaptive algorithms with regularized sample covariance matrix," *Radiotekh. Elektron. (Radio engineering electronic physics)*, vol. 27, no. 10, pp. 69–77, 1982.

[44] B. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Trans. Aerosp. Electron.*, vol. 24, no. 3, pp. 397–401, 1988.

[45] M. Ganz, R. Moses, and S. Wilson, "Convergence of the smi and the diagonally loaded smi algorithms with weak interference," *IEEE Trans. Antennas Propag.*, vol. 38, no. 3, pp. 394–399, Mar. 1990.

[46] R.G. Lorenz and S.P. Boyd, "Robust minimum variance beamforming," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[47] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1702–1715, July 2003.

[48] H. Cox, R.M. Zeskind, and M.H. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acous. Speech, Signal Proc.*, vol. ASSP–35, pp. 1365–1376, Oct. 1987.

[49] O. Besson and F. Vincent, "Performance analysis of beamformers using generalized loading of the covariance matrix in the presence of random steering vector errors," *IEEE Trans. Sig. Proc.*, vol. 53, no. 2, pp. 452–459, Feb. 2005.

[50] O. Besson, A.A. Monakov, and C. Chalus, "Signal waveform estimation in the presence of uncertainties about the steering vector.," *IEEE Trans. Sig. Proc.*, vol. 52, no. 9, pp. 2432–2440, Sep. 2004.

[51] K.L. Bell, Y. Ephraim, and H.L. Van Trees, "A bayesian approach to robust adaptive beamforming," *IEEE Trans. Sig. Proc.*, vol. 48, no. 2, pp. 386–398, Feb. 2000.

[52] D.D. Feldman and L.J. Griffiths, "A projection approach to robust adaptive beamforming," *IEEE Trans. Sig. Proc.*, vol. 42, pp. 867–876, Apr. 1994.

[53] L. Chang and C.C. Yeh, "Performance of dmi and eigenspace-based beamformers," *IEEE Trans. Antennas Prop.*, vol. 40, pp. 1336–1347, Nov. 1992.

[54] J.K. Thomas, L.L. Scharf, and D.W. Tufts, "The probability of a subspace swap in the svd," *IEEE Transactions on Signal Processing*, vol. 43, no. 3, pp. 730–736, 1995.

[55] M. Hawkes, A. Nehorai, and P. Stoica, "Performance breakdown of subspace-based methods: prediction and cure," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2001, 7-11 May, 2001, Salt Palace Convention Center, Salt Lake City, Utah, USA, Proceedings*, 2001, pp. 4005–4008.

[56] A. Hassanien, S.A. Vorobyov, and K.M. Wong, "Robust adaptive beamforming using sequential quadratic programming: an iterative solution to the mismatch problem," *IEEE Signal Process. Lett.*, vol. 15, pp. 733–736, November 2008.

[57] M. Joham, W. Utschick, and J.A. Nossek, "Robust adaptive beamforming based on steering vector estimation with as little as possible prior information," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2974–2987, June 2012.

[58] Y. Gu and A. Leshem, "Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3881–3885, July 2012.

[59] 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.

[60] Y. Chen, A. Wiesel, and A.O. Hero, "Robust shrinkage estimation of high-dimensional covariance matrices," *IEEE Trans. Signal Proc.*, vol. 59, no. 9, pp. 4097–4107, September 2011.

[61] A.P. Dempster, "Covariance selection," *Biometrics*, vol. 28, pp. 157–175, 1972.

[62] P.J. Bickel and E. Levina, "Covariance regularization by thresholding," *Annals of Statistics*, vol. 36, no. 6, pp. 2577–2604, 2008.

[63] P.J. Bickel and E. Levina, "Regularized estimation of large covariance matrices," *Annals of Statistics*, vol. 36, no. 1, pp. 199–277, 2008.

[64] T. Cai, C.H. Zhang, and H. Zhou, "Optimal rates of convergence for covariance matrix estimation," *Annals of Statistics*, vol. 38, no. 4, pp. 2118–2144, 2010.

[65] T. Cai and H. Zhou, "Minimax estimation of large covariance matrices under l1 norm," *Statistica Sinica to be published.*

[66] N. Meinshausen and P. Bühlmann, "High dimensional graphs and variable selection with the lasso," *Annals of Statistics*, vol. 34, pp. 1436–1462, 2006.

[67] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. Royal Statist. Soc.*, vol. 58, pp. 267–288, 1996.

[68] M. Yuan, "High dimensional inverse covariance matrix estimation via linear programming," *J. Mach. Learn. Res.*, vol. 11, pp. 2261–2286, 2010.

[69] E. Candes and T. Tao, "The dantzig selector: statistical estimation when p is much larger than n," *Ann. Statistics*, vol. 35, pp. 2313–2351, 2007.

[70] J. Friedman, T.R. Hastie, and R. Tibshirani, "Sparse inverse covariance estimation with the graphical lasso," *Biostatistics*, vol. 9, no. 3, pp. 432–441, 2008.

[71] O. Banerjee, L. E. Ghaoui, and A. dŠAspremont, "Model selection through sparse maximum likelihood estimation for multivariate gaussian or binary data," *Journal of Machine Learning Research*, vol. 9, pp. 485–516, 2008.

[72] A. Wiesel, Y.C. Eldar, and A.O. Hero, "Covariance estimation in decomposable gaussian graphical models," *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1482–1492, 2010.

[73] 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.

[74] L. Li, A. Tulino, and S. Verdú, "Design of reduced-rank MMSE multiuser detectors using random matrix methods," *IEEE Transactions on Information Theory*, vol. 50, no. 6, pp. 986–1008, June 2004.

[75] Sh. Yang and L. Hanzo, "Fifty years of MIMO detection: The road to large-scale MIMOs," *IEEE Commun. surveys & Tutorials*, vol. 17, no. 4, pp. 1941–1988, Fourth Quarter 2015.

[76] M. Joham, W. Utschick, and J.A. Nossek, "Linear transmit processing in MIMO communications systems," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2700–2712, August 2005.

[77] N. Ma and J. Goh, "Efficient method to determine diagonal loading value," in *Proc. International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 2003, vol. V, pp. 341–344.

[78] R. Abrahamsson, Y. Selén, and P. Stoica, "Enhanced covariance matrix estimators in adaptive beamforming," in *Proc. International Conference on Acoustics Speech and Signal Processing (ICASSP) 2007*, 2007, pp. 969–972.

[79] Ch.K. Wen, J.Ch. Chen, and P. Ting, "A shrinkage linear minimum mean square error estimator," *IEEE Signal Process. Letters*, vol. 20, no. 12, pp. 1179–1182, Dec. 2013.

[80] S.A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *Signal Proc. Elsevier*, vol. 93, pp. 3264–3277, May 2013.

[81] P. Stoica and R.L. Moses, *Spectral Analysis of Signals*, vol. 1, Prentice Hall, 2005.

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

[83] 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.

[84] 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.

[85] 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.

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

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

[88] 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.

[89] 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.

[90] 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.

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

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

[93] 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.

[94] L. Pastur and V. Vasilchuk, "On the law of addition of random matrices," *Communications in Mathematical Physics*, vol. 214, no. 2, pp. 249–286, 2000.

[95] W. Hachem, O. Khorunzhy, Ph. Loubaton, J. Najim, and L.A. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Trans. Info. Theory*, vol. 54, no. 9, pp. 3987–4004, 2008.

[96] 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.

[97] 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.

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

[99] Z.D. Bai and J.W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Springer-Verlag New York, New York, 2 edition, 2010.

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

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

[102] 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.

[103] 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.

[104] 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.

[105] 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.

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

[107] I.E. Telatar, "Capacity of multi antenna gaussian channels," *Bell Labs, Technical memorandum*, pp. 585–595, 1995.

[108] D.N.C. Tse and S.V. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Info. Theory*, vol. 45, no. 2, pp. 641–657, March 1999.

[109] 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.

[110] 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.

[111] P. Bianchi, M. Debbah, M. Maida, and J. Najim, "Performance of Statistical tests for Single-source Detection Using Random Matrix Theory," *IEEE Trans. Info. Theory*, vol. 57, no. 4, pp. 2400–2419, April 2011.

[112] 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.

[113] 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.

[114] 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.

[115] 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.

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

[117] 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.

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

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

[120] 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.

[121] 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.

[122] 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.

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

[124] X. Mestre and M.A. Lagunas, "Estimating the optimum loading factor against the finite sample size effect in minimum variance beamformers," in *Proc. Sensor Array and Multichannel Signal Processing Workshop (SAM) 2004*, 2004, pp. 357–361.

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

[126] 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.

[127] X. Mestre, "On the asymptotic behavior of quadratic forms of the resolvent of certain covariance-type matrices.," Tech. Rep. CTTC/RC/2006-01, CTTC. Available http://www.cttc.es/people/xmestre/, 2006.

[128] 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.

[129] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Trans. Signal Processing*, vol. 13, no. 4, pp. 67–94, June 1996.

[130] 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.

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

[132] M. Sousa Lobo, L. Vandenberghe, S. Boyd, and H. Lebret, "Applications of second-order cone programming," *Linear Algebra and its Applications Elsevier*, pp. 193–228, 1998.

[133] Michael Grant and Stephen Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," http://cvxr.com/cvx, Mar. 2014.

[134] H.Q. Ngo, E. Larsson, and T. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, April 2013.

[135] N. Shariati, E. Björnson, M. Bengtsson, and M. Debbah, "Low-complexity polynomial channel estimation in large-scale MIMO with arbitrary statistics," *IEEE Journal Selected Topics Signal Processing*, vol. 8, no. 5, pp. 815–830, October 2014.

[136] F. Jiang, J. Chen, A. Lee Swindlehurst, and J.A. López-Salcedo, "Massive MIMO for wireless sensing with a coherent multiple access channel," *IEEE Trans. Signal Proc.*, vol. 63, no. 12, pp. 3005–3017, June 2015.

[137] Q. Ding and E.D. Kolaczyk, "A compressed pca subspace method for anomaly detection in high-dimensional data," *IEEE Trans. Info. Theory*, vol. 59, no. 11, pp. 7419–7433, November 2013.

[138] W. Johnson and J. Lindenstrauss, "Extensions of lipschitz mappings into a hilbert space," *Contemp. Math.*, vol. 26, pp. 189–206, 1984.

[139] O. Ledoit and M. Wolf, "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *The Annals of Statistics*, vol. 40, no. 2, pp. 1024–1060, 2012.

[140] O. Ledoit and S. Péché, "Eigenvectors of some large sample covariance matrix ensembles," *Probab. Theory Related Fields*, vol. 151, pp. 233–264, 2011.