2.1 INTRODUCTION

Adaptive beamforming is a versatile approach to detect and estimate the signal-of-interest at the output of a sensor array by means of data-adaptive spatial filtering and interference rejection. It has a long and rich history of interdisciplinary theoretical research [1–8] and practical applications to numerous areas such as sonar [9–14], radar and remote sensing [15–18], wireless communications [19–23], global positioning [24–26], radio astronomy [27, 28], microphone array speech processing [29–31], seismology [32, 33], biomedicine [34, 35], and other fields. In a few recent years, there has been a renewed interest to this area in application to wireless communications where smart (adaptive) antennas have emerged as one of the key technologies for the third and higher generations of mobile radio systems [23].

The traditional approach to the design of adaptive array algorithms assumes that there is no desired signal component in the beamformer training cell data [2, 4, 8].
Although this assumption may be relevant in several specific cases (for example, in certain radar and active sonar problems), in most applications the interference and noise observations are ‘contaminated’ by the signal component [36–38]. Such applications include, for example, passive sonar, wireless communications, microphone array processing, and radioastronomy. If signal-free beamformer training snapshots are available, adaptive array algorithms are known to be quite robust against errors in the steering vector of the desired signal and limited training sample size [2–8, 39, 40]. However, the situation is completely different in the case when the desired signal is present in the training data snapshots. It is well known that in the latter case, traditional adaptive beamforming methods suffer from the signal cancellation phenomenon, that is, they degrade severely in their performance and convergence rate. Such a degradation can take place even when the signal steering vector is precisely known at the beamformer but the sample size is limited [36, 38, 41].

In practical scenarios, the performance degradation of traditional adaptive beamforming techniques may become even more pronounced because most of these techniques are based on the assumption of an accurate knowledge of the array response to the desired signal. Moreover, these methods often use quite restrictive assumptions on the environment and interferences, for example, they assume that the received array data are stationary and/or that the interferers can be described using a low-rank model. As a result, such techniques can become severely degraded in scenarios when the exploited assumptions on the environment, antenna array and/or sources are wrong or inaccurate [36, 38].

One of the most typical reasons of performance degradation of adaptive beamformers is a mismatch between the presumed and the actual array responses to the desired signal. Such a mismatch can be caused by look direction/pointing errors [42–45], an imperfect array calibration (distorted antenna shape) [46], unknown wavefront distortions and signal fading [11, 47–49], near-far wavefront mismodeling [50], local scattering [51], as well as other effects [36, 52]. Traditional adaptive array algorithms are known to be extremely sensitive even to slight mismatches of such type because in the presence of them, an adaptive beamformer tends to mix up the signal and interference components, that is, it interprets the desired signal component in array observations as an additional interfering source and, consequently, suppresses the desired signal instead of maintaining distortionless response to it [36, 41]. This phenomenon is sometimes called self-nulling in the adaptive beamforming literature [38, 53].

Another cause of performance degradation of adaptive beamformers is a nonstationarity of the environment, antenna array, and/or sources. Such nonstationarity effects can be induced by rapid variations of the propagation channel, interferer and antenna motion and/or vibration, and are quite typical for radar, sonar, and wireless communications [54–58]. They may cause a substantial performance degradation of adaptive beamformers because they limit the training sample size and may lead to interference undernulling. When such nonstationarity effects are combined with the effect of the presence of the desired signal in the training cell, the aforementioned degradation can become much stronger than in the case of signal-free beamformer training data [56].
One typical example of negative effects of nonstationarity is the case when the interfering sources move rapidly. In such case, the array weights may not be able to adapt fast enough to compensate for this motion. That is, the interferers tend to be always located outside the narrow areas of the adapted beampattern nulls and to leak to the output of adaptive beamformer through the beampattern sidelobes [56]. The same situation may occur when moving or vibrating antenna arrays are employed, for example, towed arrays in sonar [14] or airborne antenna arrays [54].

In many practical sonar and wireless communications scenarios, the signal and interference wavefronts may suffer from a multiplicative noise and angular spreading. In sonar, this type of noise is caused by a long-distance propagation through a randomly inhomogeneous medium [10, 47, 48]. In wireless communications, the array signal response may suffer from fading and local scattering [49, 51].

In the presence of multiplicative noise, higher-rank signal source models have to be used instead of the point (rank-one) model because in this case, each source results into multiple rank-one components in the array covariance matrix [38]. It can be shown that, in such scenarios, the array response should be characterized by the signal covariance matrix rather than the signal steering vector [36, 38, 59]. As a result, the robustness of adaptive beamformers against mismatches between the presumed and actual signal covariance matrices (rather than the mismatches between the corresponding steering vectors) must be considered.

In this chapter, we provide an overview of traditional ad hoc robust adaptive beamforming techniques and give a detailed introduction to a recently emerged rigorous approach to robust minimum variance beamforming based on worst-case performance optimization [59–62]. This approach represents the current state of the art of robust adaptive beamforming. It is shown that it provides efficient solutions to the aforementioned robustness problems including the array response mismatch and data nonstationarity problems.

The remainder of this chapter is organized as follows. In the next section, some background on adaptive arrays is given and the traditional (robust and nonrobust) adaptive beamforming techniques are discussed. Then, in Section 2.3, the worst-case performance optimization-based adaptive beamformers are considered. In Section 2.4, simulation results are presented that demonstrate an improved robustness of these worst-case optimization-based beamformers as compared to the earlier robust and nonrobust techniques. Conclusions are given in Section 2.5.

2.2 BACKGROUND AND TRADITIONAL APPROACHES

The generic scheme of a narrowband beamformer is shown in Figure 2.1. The beamformer output signal can be written as

\[ y(k) = w^H x(k) \]

where \( k \) is the time index, \( x(k) = [x_1(k), \ldots, x_M(k)]^T \) is the \( M \times 1 \) complex vector of array observations, \( w = [w_1, \ldots, w_M]^T \) is the \( M \times 1 \) complex vector of beamformer
weights, $M$ is the number of array sensors, and $(\cdot)^T$ and $(\cdot)^H$ denote the transpose and Hermitian transpose, respectively. The training snapshot (array observation vector) is given by

$$x(t) = \beta s_s(t) + i(t) + n(t)$$  \hspace{0.5cm} (2.1)

where $s_s(t), i(t), \text{and } n(t)$ are the statistically independent components of the desired signal, interference, and sensor noise, respectively, and the binary parameter $\beta$ is equal to zero if the training cell snapshots are signal-free and is equal to one otherwise. In what follows, mostly the case $\beta = 1$ will be considered. If the desired signal is a point source and has a time-invariant wavefront, we obtain that

$$s_s(t) = s(t) a_s$$

where $s(t)$ is the complex signal waveform and $a_s$ is its $M \times 1$ steering vector. Then, taking into account that $\beta = 1$, (2.1) can be written as

$$x(t) = s(t) a_s + i(t) + n(t)$$

The optimal weight vector can be obtained by means of maximizing the signal-to-interference-plus-noise ratio (SINR) [4, 8]

$$\text{SINR} = \frac{w^H R_s w}{w^H R_{i+n} w}$$  \hspace{0.5cm} (2.2)

where

$$R_s \triangleq \mathbb{E}\{s_s(t)s_s^H(t)\}$$

$$R_{i+n} \triangleq \mathbb{E}\{(i(t) + n(t)) (i(t) + n(t))^H\}$$
are the $M \times M$ signal and interference-plus-noise covariance matrices, respectively, and $E\{\cdot\}$ denotes the statistical expectation. Note that the matrix $R_s$ can have an arbitrary rank, that is, $1 \leq \text{rank}\{R_s\} \leq M$.

In many practical situations, $\text{rank}\{R_s\} > 1$. Typical examples of such situations are scenarios with incoherently scattered sources or signals with randomly fluctuating wavefronts which frequently occur in sonar and wireless communications.

In the incoherently scattered source case, $R_s$ has the following form [63, 64]:

$$R_s = \sigma_s^2 \int_{-\pi/2}^{\pi/2} \rho(\theta) a(\theta)a^H(\theta) d\theta$$

(2.3)

where $\rho(\theta)$ is the normalized angular power density [$\int_{-\pi/2}^{\pi/2} \rho(\theta) d\theta = 1$], $\sigma_s^2$ is the signal power, and $a(\theta)$ is the array steering vector.

In the case of randomly fluctuating wavefronts, the signal covariance matrix takes another form [47, 48, 65]

$$R_s = \sigma_s^2 B \odot \{a, a_s^H\}$$

(2.4)

where $B$ is the $M \times M$ coherence loss matrix and $\odot$ is the Schur-Hadamard (elementwise) matrix product.

There are two commonly used models for the coherence loss matrix [47, 48, 63, 65]:

$$[B]_{m,n} = \exp\{-(m-n)^2 \xi\}$$

(2.5)

$$[B]_{m,n} = \exp\{-|m-n| \xi\}$$

(2.6)

where $\xi$ is the coherence loss parameter.

Obviously, the rank of $R_s$ in (2.3) and (2.4) can be higher than one. It is important to stress that in practice, both $\rho(\theta)$ and $B$ may be uncertain [11, 51]. Therefore, in the both cases of spatially spread and imperfectly coherent sources, we may expect a substantial mismatch between the presumed and actual signal covariance matrices [59].

In the special case of a point signal source, we have

$$R_s = \sigma_s^2 a_s a_s^H$$

In this case, $\text{rank}\{R_s\} = 1$ and (2.2) can be simplified to

$$\text{SINR} = \frac{\sigma_s^2 |w^H a_s|^2}{w^H R_{i+n} w}$$

(2.7)

To find the optimal solution for the weight vector, we should maximize the SINR in (2.2) or, alternatively, in (2.7). These optimization problems are equivalent to maintaining distortionless response to the desired signal while minimizing the
output interference-plus-noise power, that is,

\[
\begin{align*}
\min_w w^H R_{i+n} w & \quad \text{subject to } w^H R_s w = 1 \\
\min_w w^H R_{i+n} w & \quad \text{subject to } w^H a_s = 1
\end{align*}
\]

in the general-rank and rank-one signal cases, respectively. This approach is usually referred to as the minimum variance distortionless response (MVDR) beamforming \([4, 8]\).

The solution to (2.8) can be found by means of minimization of the function

\[
H(w, \lambda) = w^H R_{i+n} w + \lambda(1 - w^H R_s w)
\]

where \(\lambda\) is a Lagrange multiplier. Taking the gradient of (2.10) and equating it to zero, we obtain that the solution to (2.8) is given by the following generalized eigenvalue problem \([36, 59]\):

\[
R_{i+n} w = \lambda R_s w
\]

where the Lagrange multiplier \(\lambda\) can be interpreted as a corresponding generalized eigenvalue. It is easy to prove that all generalized eigenvalues in (2.11) are nonnegative real numbers. Indeed, using (2.11) we have that \(w^H R_{i+n} w = \lambda w^H R_s w\). Using the fact that the matrices \(R_{i+n}\) and \(R_s\) are positive semidefinite, we prove that \(\lambda\) is always real and non-negative.

The solution to the problem (2.8) is the generalized eigenvector that corresponds to the minimal generalized eigenvalue of the matrix pencil \(\{R_{i+n}, R_s\}\). Multiplying (2.11) by \(R_{i+n}^{-1}\), we can write this equation as

\[
R_{i+n}^{-1} R_s w = \frac{1}{\lambda} w
\]

which can be identified as the characteristic equation for the matrix \(R_{i+n}^{-1} R_s\). From the fact of non-negativeness of \(\lambda\), it follows that the minimal generalized eigenvalue \(\lambda_{\min}\) in (2.11) corresponds to the maximal eigenvalue \(1/\lambda_{\min}\) in (2.12). Using the latter fact, the optimal weight vector can be explicitly written as

\[
w_{\text{opt}} = P\{R_{i+n}^{-1} R_s\}
\]

where \(P\{\cdot\}\) is the operator which returns the principal eigenvector of a matrix, that is, the eigenvector that corresponds to its maximal eigenvalue. According to (2.8) and the fact that any eigenvector can be normalized arbitrarily, the resulting weight has to be normalized to satisfy the constraint \(w_{\text{opt}}^H R_s w_{\text{opt}} = 1\) in (2.8). However, it is clear that multiplying the weight vector by any nonzero
constant, we do not affect the output SINR (2.2). Hence, such normalization is immaterial [38].

The optimal solution (2.13) will not change if the interference-plus-noise covariance matrix $R_{i+n}$ would be replaced by the training data covariance matrix

$$R = E\{x(t)x^H(t)\} = R_{i+n} + R_s$$

(2.14)

Therefore, we have

$$w_{opt} = P\{R_{i+n}^{-1}R_s\} = P\{R^{-1}R_s\}$$

(2.15)

Note that (2.15) directly follows from (2.8) and (2.14).

In the rank-one signal source case, $R_s = \sigma^2_a a_s a_s^H$ and we have that equation (2.13) can be rewritten as

$$w_{opt} = P\{R_{i+n}^{-1}a_s a_s^H\} = \alpha R_{i+n}^{-1}a_s$$

(2.16)

where the constant $\alpha$ can be obtained from the MVDR constraint $w_{opt}^H a_s = 1$ in (2.9) and is equal to [4]

$$\alpha = \frac{1}{a_s^H R_{i+n}^{-1} a_s}$$

However, as has been noted before, this constant does not affect the output SINR and, therefore, is omitted in the sequel. Equation (2.16) is the classic Wiener solution for the weight vector of the optimal beamformer in the rank-one signal case [2, 4].

In practical applications, the true matrices $R_{i+n}$ and $R$ are unavailable but can be estimated from the received data or obtained from a priori information about the sources. Usually, the sample covariance matrix [2, 4]

$$\hat{R} = \frac{1}{N} \sum_{n=1}^{N} x(n)x^H(n)$$

(2.17)

is used in the optimization problems (2.8) and (2.9) instead of $R_{i+n}$, where $N$ is the training sample size. The solutions to these modified problems are usually referred to as the sample matrix inverse (SMI) beamformers [2]

$$w_{SMI} = P\{\hat{R}^{-1}R_s\}$$

(2.18)

$$w_{SMI} = \hat{R}^{-1}a_s$$

(2.19)

for the general-rank and rank-one cases, respectively.
The use of the sample covariance matrix \( \hat{R} \) instead of the exact array covariance matrix \( R \) in (2.19) is known to lead to a substantial performance degradation in the case when the signal component is present in the beamformer training data. It is well known that in the signal-free training data case the output SINR of the SMI beamformer (2.19) converges to the optimal SINR

\[
\text{SINR}_{\text{opt}} = \sigma_s^2 a_s^H R_{1+n}^{-1} a_s
\]  

(2.20)

so that the mean losses relative to (2.20) are less than 3 dB if the following condition is satisfied [2]:

\[ N \geq 2M \]  

(2.21)

However, this rule is no longer applicable when the desired signal contaminates the beamformer training data. In the latter case, the same performance loss can be achieved only when [41]

\[ N \geq \text{SINR}_{\text{opt}}(M - 1) \gg M \]  

(2.22)

where the SNR is assumed to be high. According to (2.22), in the presence of the desired signal in the beamformer training data, the SMI algorithm has much slower convergence and weaker robustness against finite sample effects than in the signal-free training data case.

In practice, the situation is further complicated by the fact that the signal covariance matrix is usually known imprecisely, that is, there is always a certain mismatch between the presumed signal covariance matrix \( R_s \) and its actual value which is hereafter denoted as \( \tilde{R}_s \). The main objective of the remainder of this section is to overview traditional ad hoc robust approaches to adaptive beamforming that aim to improve the beamformer performance in scenarios with arbitrary errors in the array response to the desired signal (i.e., the errors between the matrices \( R_s \) and \( \tilde{R}_s \)), small training sample size, and training data nonstationarity.

One of the most popular approaches to robust adaptive beamforming in the presence of such array response errors and small training sample size is the diagonal loading technique which was developed independently in [37, 66–68]. The central idea of this approach is to regularize the problem (2.8) by adding a quadratic penalty term to the objective function [68]. Then, in the finite sample case we obtain the following regularized problem [38]:

\[
\min_w w^H \tilde{R} w + \gamma w^H w \quad \text{subject to} \quad w^H R_s w = 1
\]  

(2.23)

where \( \gamma \) is the penalty weight (also called the diagonal loading factor). We will refer to the solution to (2.23) as the loaded SMI (LSMI) beamformer whose weight vector
has the following form [38, 59]:

$$w_{LSMI} = P\{ (\hat{R} + \gamma I)^{-1} R_s \} \quad (2.24)$$

where $I$ is the identity matrix. In the rank-one signal source case (rank($R_s$) = 1), (2.24) reduces to [37, 66, 68]

$$w_{LSMI} = (\hat{R} + \gamma I)^{-1} a_s \quad (2.25)$$

From (2.24) and (2.25), it is clear that adding the penalty term $\gamma w^H w$ to the objective function in (2.23) amounts to loading the diagonal of the sample covariance matrix $\hat{R}$ by the value of $\gamma$. This means that the diagonal loading operation can be interpreted in terms of injecting an artificial amount of white noise into the main diagonal of this matrix. An important property of diagonal loading is that it warrants invertibility of the diagonally loaded matrix $\hat{R} + \gamma I$ irrespectively whether $\hat{R}$ is singular or not.

Moreover, the diagonal loading approach is known to improve the performance of the SMI beamformer in scenarios with mismatched array response [36, 37, 45, 60]. However, the main shortcoming of traditional diagonal loading-based techniques is that there is no rigorous way of choosing the loading parameter $\gamma$. In [37], it was proposed to choose this parameter using the following white noise gain constraint:

$$|w^H a_s|^2 = \kappa \|w\|^2 \quad (2.26)$$

where hereafter $\|\|$ denotes the two-norm of a vector or a matrix, and the parameter $\kappa$ determines the required white noise gain. This constraint can be added to the MVDR beamformer as follows [37]:

$$\min_w w^H R_{i+n} w \quad \text{subject to } w^H a_s = 1, \quad |w^H a_s|^2 = \kappa w^H w \quad (2.27)$$

The solution to the problem (2.27) is given by [37]

$$w = \frac{(R_{i+n} + \gamma I)^{-1} a_s}{a_s^H (R_{i+n} + \gamma I)^{-1} a_s}$$

which, after replacing $R_{i+n}$ by $\hat{R}$ and ignoring the immaterial constant $(a_s^H (R_{i+n} + \gamma I)^{-1} a_s)^{-1}$, becomes equivalent to the LSMI beamformer (2.25) whose diagonal loading parameter should satisfy the white noise gain constraint (2.26).

Unfortunately, it is not quite clear how to choose the white noise gain parameter $\kappa$ and, as a rule, this parameter is chosen in a somewhat ad hoc way [37]. Also, there is no simple relationship between the parameters $\kappa$ and $\gamma$. Hence, an iterative procedure is required to obtain $\gamma$ for any given $\kappa$ [37].
A much simpler and more common ad hoc way of choosing the parameter $\gamma$ is based on estimating the noise power (e.g., using the noise-subspace eigenvalues or the minimal eigenvalue of the sample covariance matrix) and choosing $\gamma$ of the same or higher order of magnitude [8, 36–38, 45, 59, 66]. A typical choice of $\gamma$ is $10 \div 15$ dB higher than the noise power.

As the optimal choice of the diagonal loading factor is well known to be scenario-dependent [38], such a method of choosing fixed $\gamma$ is only suboptimal and may cause a substantial performance degradation of adaptive beamformers [59–62, 69].

Another popular robust adaptive beamforming technique in the rank-one signal case (i.e., in the presence of steering vector errors) and in situations with small sample size is the eigenspace-based beamformer [41, 70]. In contrast to the LSMI beamformer, this approach is only applicable to the rank-one signal case. The key idea of this technique is to reduce steering vector errors by projecting the signal steering vector onto the estimated signal-plus-interference subspace obtained via the eigendecomposition of the sample covariance matrix (2.17). This eigendecomposition can be written as

$$\hat{R} = \hat{E}\hat{A}\hat{E}^H + \hat{G}\hat{G}^H$$

where the $M \times (L+1)$ matrix $\hat{E}$ contains the $L+1$ signal-plus-interference subspace eigenvectors of $\hat{R}$, and the $(L+1) \times (L+1)$ diagonal matrix $\hat{A}$ contains the corresponding eigenvalues of this matrix. Similarly, the $M \times (M-L-1)$ matrix $\hat{G}$ contains the $M-L-1$ noise-subspace eigenvectors of $\hat{R}$, and the $(M-L-1) \times (M-L-1)$ diagonal matrix $\hat{\Gamma}$ contains the corresponding eigenvalues. The rank of the interference subspace, $L$, is assumed to be known. The weight vector of the eigenspace-based beamformer can be written as

$$w_{\text{eig}} = \hat{R}^{-1}P_E a_s$$

(2.28)

where

$$P_E = \hat{E}(\hat{E}^H\hat{E})^{-1}\hat{E}^H = \hat{E}\hat{E}^H$$

is the orthogonal projection matrix onto the estimated signal-plus-interference subspace. The weight vector (2.28) can be alternatively written as

$$w_{\text{eig}} = \hat{E}\hat{\Lambda}^{-1}\hat{E}^H a_s$$

(2.29)

If the rank of signal-plus-interference subspace is low and if the parameter $L$ is exactly known, the eigenspace-based beamformer is known to provide excellent robustness against arbitrary steering vector errors [70]. Unfortunately, this approach may degrade severely if the low-rank interference-plus-signal assumption is violated or if the subspace dimension $L$ is uncertain or known imprecisely. For example, in the presence of incoherently scattered (spatially dispersed) interfering sources,
interferers with randomly fluctuating wavefronts, and moving interferers, the low-
rank interference assumption may become violated and \( L \) can be uncertain. There-
fore, the eigenspace-based beamformer may be not a proper method of choice in
such cases [38]. Moreover, even if the low-rank model assumption remains relevant,
the eigenspace-based beamformer can be only used in scenarios where the signal-to-
noise ratio (SNR) is sufficiently high because, otherwise, subspace swap effects
become dominant and may cause a severe performance degradation of the eigen-
space-based beamformer [60]. All these shortcomings make it very difficult to use
this beamformer in practice where the dimension of the signal-plus-interference sub-
space may be uncertain and relatively high due to the source scattering and fading
effects as well as training data nonstationarity [10, 11, 14, 47, 49, 51, 54–58].

In the past decade, several advanced methods have been developed to mitigate
performance degradation of adaptive beamformers in the case of nonstationary
training data (e.g., in scenarios with moving interferers or rotating antenna)
[54–58]. For example, several authors independently used the idea of artificial
broadening the adaptive beampattern nulls to improve the robustness of adaptive
beamforming, see [55–58, 71, 72].

One approach to broaden the adaptive beampattern nulls has been proposed in
[55] and [56] using the data-dependent derivative constraints (DDCs). The essence
of this approach is to replace the sample covariance matrix \( \hat{R} \) in the SMI and LSMI
beamformers by the modified covariance matrix

\[
\hat{R} = \hat{R} + \sum_{k=1}^{K} \zeta_k B^k \hat{R} B^k
\]

(2.30)

where \( B \) is the known diagonal matrix whose entries are determined by the array
geometry, \( K \) is the highest order of the data-dependent constraints used, and the coef-
ficients \( \zeta_k \) determine the tradeoff between the constraints of different order. In prac-
tical applications, \( K = 1 \) is shown in [56] to be sufficient to provide satisfactory
robustness against interferer motion. Using \( K = 1 \), (2.30) can be simplified as

\[
\hat{R} = \hat{R} + \zeta_1 \hat{R} B \hat{R} B
\]

where \( \zeta_1 \) determines the tradeoff between the null depth and the null width. Under a
few mild conditions, the optimal value of \( \zeta_1 \) becomes independent of the source par-
parameters and can be easily computed from the known array parameters [56].

Another way to broaden the adaptive beampattern nulls is based on point con-
straints and is referred to as the so-called covariance matrix tapering (MT)
method [57, 58, 71–73]. The essence of this approach is to replace the sample
covariance matrix \( \hat{R} \) in the SMI or LSMI beamformer by the following tapered
covariance matrix:

\[
\hat{R}_T = \hat{R} \circ T
\]
where $T$ is the so-called $M \times M$ taper matrix and $\odot$ denotes the Schur–Hadamard matrix product. Using the taper matrix introduced in [71] and [72], we can express the elements of $T$ as

$$[T]_{jl} = \frac{\sin(i - l)\xi}{(i - l)\xi}$$

(2.31)

where the parameter $\xi$ determines the required beampattern null width. Another type of matrix taper is proposed in [57].

An interesting link between the MT and DDC approaches was discovered in [73]. In this work, it has been proven that the matrix (2.30) can be viewed as a tapered covariance matrix with particular choice of $T$. Hence, the DDC approach can be interpreted and implemented using the MT method. However, a serious shortcoming of the MT approach with respect to the DDC technique is that, in the general case, the former approach does not have computationally efficient on-line implementations [38].

The performance of both these methods has been studied thoroughly by means of computer simulations [56–58] and real sonar data processing [14]. The results of this study have shown that these two approaches provide an additional robustness relative to the SMI and LSMI beamformers in slowly moving interference cases, but their performance can become degraded in situations with rapidly moving interferers. Moreover, both these techniques exploit the assumptions of known array geometry and plane interferer wavefronts. Therefore, they may degrade in the case when the array is imperfectly calibrated (e.g., has a distorted shape or unknown sensor gains and phases) or when the wavefronts of the interferers deviate from the plane wavefront form because of multiplicative noise and signal fading/multipath effects or due to interferers located in the near field.

### 2.3 ROBUST MINIMUM VARIANCE BEAMFORMING BASED ON WORST-CASE PERFORMANCE OPTIMIZATION

In the previous section, main ad hoc approaches to robust adaptive beamforming have been discussed. In this section, we discuss a more powerful and theoretically rigorous worst-case performance optimization-based approach to robust adaptive beamforming that has been recently developed in [59–62].

#### 2.3.1 Rank-One Signal Case

First of all, let us consider the simplest case of a rank-one desired signal with mismatched steering vector. Let the vector of unknown mismatch between the actual steering vector $\tilde{a}_s$ and its presumed value $a_s$ be denoted as

$$\delta = \tilde{a}_s - a_s$$
Following the idea of [60], we assume that the unknown mismatch vector $\delta$ is norm-bounded by some known constant $\epsilon$, that is,

$$
\|\delta\| \leq \epsilon 
$$

(2.32)

To incorporate robustness into the MVDR beamforming problem, let us maximize the worst-case SINR by solving the following problem:

$$
\max_w \min_\delta \sigma^2_s \frac{|w^H(a_s + \delta)|^2}{w^H R_{++i} w} \quad \text{subject to } \|\delta\| \leq \epsilon
$$

This problem is equivalent to the following robust MVDR beamforming problem [60]:

$$
\min w^H R_{++i} w \quad \text{subject to } |w^H(a_s + \delta)| \geq 1 \quad \text{for all } \|\delta\| \leq \epsilon 
$$

(2.33)

The main modification in (2.33) with respect to the original problem (2.9) is that instead of requiring fixed distortionless response towards the single presumed steering vector $a_s$, such distortionless response is now maintained in (2.33) by means of inequality constraints for a continuum of all possible steering vectors that belong to the spherical uncertainty set

$$
A \triangleq \{c \mid c = a_s + \delta, \quad \|\delta\| \leq \epsilon\}
$$

The constraints in (2.33) guarantee that the distortionless response will be maintained in the worst case, that is, for the particular vector $\delta$ which corresponds to the smallest value of $|w^H(a_s + \delta)|$ provided that $\|\delta\| \leq \epsilon$.

In the finite sample case, $R_{++i}$ should be replaced by $\hat{R}$. Doing so and replacing the infinite number of constraints in (2.33) by the aforementioned single worst-case constraint, the problem (2.33) becomes

$$
\min w^H \hat{R} w \quad \text{subject to } \min_{\|\delta\| \leq \epsilon} |w^H a_s + w^H \delta| \geq 1 
$$

(2.34)

Note that the inequality constraint in (2.34) is equivalent to the equality constraint

$$
\min_{\|\delta\| \leq \epsilon} |w^H a_s + w^H \delta| = 1 
$$

(2.35)

The equivalence of the equality constraint (2.35) and the inequality constraint in (2.34) can be easily proved by contradiction as follows [60]. If they are not equivalent to each other then the minimum of the objective function in (2.34) is achieved when $\chi \triangleq \min_{\|\delta\| \leq \epsilon} |w^H a_s + w^H \delta| > 1$. However, replacing $w$ with $w/\sqrt{\chi}$, we can decrease the objective function $w^H \hat{R} w$ by the factor of $\chi > 1$ while the constraint in (2.34) will be still satisfied. This is an obvious contradiction to the original statement
that the objective function is minimized when $\chi > 1$. Therefore, the minimum of the objective function is achieved at $\chi = 1$ and this means that the inequality constraint in (2.34) is equivalent to the equality constraint (2.35).

If the sequel, we will use this constraint in both its inequality and equality equivalent forms. The following lemma [60] can be proved.

**Lemma 1.** If

\[ |w^H a_s| \geq \epsilon \|w\| \]  \hspace{1cm} (2.36)

then

\[ \min_{|\delta| \leq \epsilon} |w^H(a_s + \delta)| = |w^H a_s| - \epsilon \|w\| \]

**Proof.** See Appendix 2.A. \hfill \Box

Note that, according to (2.26), the condition (2.36) is used in Lemma 1 to guarantee a sufficient white noise gain [37].

Assuming that this condition is satisfied and using Lemma 1, we can rewrite problem (2.34) as the following quadratic minimization problem with a single nonlinear constraint:

\[ \min_w w^H \hat{R} w \quad \text{subject to} \quad |w^H a_s| - \epsilon \|w\| \geq 1 \]  \hspace{1cm} (2.37)

The nonlinear constraint in (2.37) is still nonconvex due to the absolute value operation on the left-hand side. To convert this problem to a convex one, we can use the fact that the cost function in (2.37) is unchanged when $w$ undergoes an arbitrary phase rotation [60]. Therefore, if $w_0$ is an optimal solution to (2.37), we can always rotate, without affecting the objective function value, the phase of $w_0$ so that $w^H a_s$ is real. Thus, without any loss of generality, $w$ can be chosen such that

\[ \text{Re} \{w^H a_s\} \geq 0 \]  \hspace{1cm} (2.38)

\[ \text{Im} \{w^H a_s\} = 0 \]  \hspace{1cm} (2.39)

Using this observation, the problem can be written as [60]

\[ \min_w w^H \hat{R} w \quad \text{subject to} \quad w^H a_s \geq \epsilon \|w\| + 1 \]  \hspace{1cm} (2.40)

where, according to the aforementioned fact that the constraint in (2.40) is satisfied with equality, (2.39) can be ignored because from $w^H a_s = \epsilon \|w\| + 1$ it follows that the value of $w^H a_s$ is real-valued and positive.

Comparing the white noise gain constraint (2.26) and the constraint in (2.40), we see that they have a high degree of similarity, although the latter constraint contains
an additional constant term in the right-hand side. This observation helps us to understand the relationship between the white noise gain constraint based beamformer (2.27) and the robust beamformer (2.40).

It is also important to stress that the original problem (2.33) appears to be computationally intractable (NP-hard), whereas the robust MVDR beamformer (2.40) of [60] belongs to the class of convex second-order cone (SOC) programming problems [74] which can be easily solved using standard and highly efficient interior point method software [75]. For example, using the primal-dual potential reduction method [74], the complexity of solving (2.40) is $O(M^3)$ per iteration, and the algorithm converges typically in less than 10 iterations (a well-known and widely accepted fact in the optimization community). Therefore, the overall computational complexity of the SOC programming based beamformer is $O(M^3)$ [60]. This complexity is comparable to that of the SMI and LSMI algorithms. An alternative way to solve problem (2.40) with the complexity $O(M^3)$ is to use the Newton-type algorithms developed in [62] and [76].

Let us overview the algorithm of [76]. As the constraint in (2.40) is satisfied with equality, we can rewrite this problem as

$$\min_w w^H \hat{R} w \quad \text{subject to } w^H a_s - \epsilon||w|| = 1$$

Using the Lagrange multiplier method, we can write the Lagrangian function as

$$L(w, \lambda) = w^H \hat{R} w - \lambda(w^H a_s - \epsilon||w|| - 1) \quad (2.41)$$

where $\lambda$ is the Lagrange multiplier. Differentiating (2.41) and equating the result to zero, we obtain the following equation:

$$\hat{R} w + \lambda \epsilon \frac{w}{||w||} = \lambda a_s \quad (2.42)$$

To solve (2.42), we need to know the Lagrange multiplier $\lambda$. However, using the fact that multiplying the weight vector by any arbitrary constant does not change the output SINR, we can transform this equation to [76]

$$\hat{R} w + \epsilon \frac{w}{||w||} = a_s \quad (2.43)$$

so that (2.43) does not contain the Lagrange multiplier anymore. For the sake of simplicity, the same notation $w$ is used in (2.43) for the rescaled weight vector as for the original one in (2.42).

Equation (2.43) can be rewritten as

$$\left( \hat{R} + \frac{\epsilon}{||w||} I \right) w = a_s \quad (2.44)$$
From (2.44), it can be seen that the robust MVDR beamformer (2.40) belongs to the class of diagonal loading techniques. Note that this beamformer uses adaptive diagonal loading because the diagonal loading factor $\epsilon/\|w\|$ depends on the norm of the weight vector and, therefore, is scenario-dependent. It should be stressed that, in contrast to the fixed diagonal loading approach used in the LSMI beamformer, such an adaptive diagonal loading technique optimally matches the diagonal loading factor to the known amount of uncertainty in the signal steering vector [60, 76].

A noteworthy observation following from (2.44) is that, if $\|w\|$ is available, then we can use (2.44) to calculate the weight vector of the robust MVDR beamformer. To determine $\|w\|$, the following simple method can be used [76]. Rewriting (2.44) as

$$w = \left(\hat{R} + \frac{\epsilon}{\|w\|}I\right)^{-1}a_s$$

and taking the norm squared of the both sides of (2.45), we have

$$\|w\|^2 = \left\|\left(\hat{R} + \frac{\epsilon}{\|w\|}I\right)^{-1}a_s\right\|^2$$

(2.46)

Introducing $\tau \triangleq \|w\| > 0$, we obtain that solving (2.46) is equivalent to finding a positive value of $\tau$ such that

$$\tau^2 = \left\|\left(\hat{R} + \frac{\epsilon}{\tau}I\right)^{-1}a_s\right\|^2$$

(2.47)

To simplify (2.47), let us use the eigendecomposition$^1$ of $\hat{R}$,

$$\hat{R} = U \Xi U^H$$

(2.48)

where $U$ is the $M \times M$ unitary matrix whose columns are the eigenvectors of $\hat{R}$ and $\Xi$ is the diagonal matrix of eigenvalues of $\hat{R}$ given by

$$\Xi = \text{diag}\{\xi_1, \ldots, \xi_M\}$$

Here, $\text{diag}\{\cdot\}$ denotes a diagonal matrix and $\{\xi_i\}_{i=1}^M$ are the real positive eigenvalues of $\hat{R}$. Without loss of generality, we assume that $\xi_1 \geq \xi_2 \geq \cdots \geq \xi_M > 0$.

Using (2.48), we can rewrite (2.47) as

$$\|U\Psi^{-1}(\tau)U^Ha_s\|^2 - \tau^2 = 0$$

(2.49)

$^1$Note that the eigendecomposition is also used in [69] in a similar way to derive a Newton-type algorithm.
where
\[ \Psi(\tau) = \Xi + \frac{\epsilon}{\tau} \]

Introducing the \( M \times 1 \) vector \( g \) as
\[ g = [g_1, \ldots, g_M]^T \triangleq U^H a_s \tag{2.50} \]

and taking into account that \( U \) is a unitary matrix, we can rewrite the left-hand side of (2.49) as
\[
\|U \Psi^{-1}(\tau) U^H a_s\|^2 - \tau^2 = \|\Psi^{-1}(\tau) g\|^2 - \tau^2
\]
\[
= \sum_{i=1}^{L} \left( \frac{|g_i|}{\xi_i + \frac{\epsilon}{\tau}} \right)^2 - \tau^2
\]
\[
= \sum_{i=1}^{L} \left( \frac{|g_i|}{\epsilon + \tau \xi_i} \right)^2 - 1 \tau^2 \tag{2.51}
\]

Using (2.51) and taking into account that \( \tau > 0 \), we obtain that solving (2.49) is equivalent to finding a positive value for \( \tau \) such that
\[
f(\tau) \triangleq \sum_{i=1}^{M} \left( \frac{|g_i|}{\epsilon + \tau \xi_i} \right)^2 - 1 = 0 \tag{2.52}
\]

Note that (2.52) may not always have a real and positive solution. The following lemma [76] states the necessary and sufficient conditions under which (2.52) has a unique positive solution.

**Lemma 2.** Equation (2.52) has a unique real-valued and positive solution if and only if
\[ \|a_i\| > \epsilon \tag{2.53} \]

**Proof.** See Appendix 2.B.

The condition similar to (2.53) has been also used in [69] and yields an intuitively appealing interpretation. As the parameter \( \epsilon \) characterizes the maximal norm of the mismatch between the presumed and the actual signal steering vectors, equation (2.53) simply states that the approach we are going to develop is applicable only if the maximum norm of such a mismatch does not exceed the norm of the presumed signal steering vector itself. In the sequel, we assume that (2.53) is always satisfied.
Using (2.52), we can upper-bound the function $f(\tau)$ as

$$f(\tau) < \frac{\sum_{i=1}^{M} |g_i|^2}{(\epsilon + \tau \Theta_M)^2} - 1$$

$$= \frac{\|g\|^2}{(\epsilon + \tau \Theta_M)^2} - 1$$

$$= \frac{\|a_s\|^2}{(\epsilon + \tau \Theta_M)^2} - 1 \triangleq f_{up}(\tau) \quad (2.54)$$

Noting that $f(\tau)$ and $f_{up}(\tau)$ are both decreasing functions for positive values of $\tau$ and that, according to Lemma 2, the root $\tau$ of $f(\tau)$ is positive, we obtain from (2.54) that this root is always smaller than the root

$$\tau_{up} = \frac{\|a_s\|}{\Theta_M} - \epsilon$$

of $f_{up}(\tau)$. Therefore, the value of $\tau$ lies in the interval $(0, \tau_{up})$. With this condition, the problem of computing $\tau$ becomes standard. For example, the algorithm of [77] can be used for this purpose [76]. The latter algorithm consists of a binary search followed by Newton–Raphson iterations. The binary search technique is used in this algorithm to obtain a proper initialization for the subsequent Newton–Raphson iterations. As shown in [77], this algorithm converges to a $v$-neighborhood of $\tau$ in $O(\log \log (\tau_{up}/v))$ iterations. The algorithm to compute $\|w\|$ can be summarized as follows [76]:

1. Use binary search to find $\tau_0 \in (0, \tau_{up})$ such that $f(\tau_0) > 0$ and $f\left(\frac{\tau_0 + \tau_0}{2}\right) < 0$ (see [77] for details).
2. Set $l = 1$ and select a small positive value of $\xi$ which will be used in the algorithm stopping criterion.
3. Obtain $\tau_l$ as

$$\tau_l = \tau_{l-1} - \frac{f(\tau_{l-1})}{f'(\tau_{l-1})}$$

where $f'(\tau_{l-1})$ is the derivative of $f(\tau)$ at $\tau = \tau_{l-1}$.
4. If $|f(\tau_l)| < \xi$, go to the next step. Otherwise, repeat steps 2 and 3.
5. Compute $\|w\|$ as $\tau = \tau_l$.

The value of $\|w\|$ which is computed by means of this procedure can be then substituted to (2.45) to obtain the resulting weight vector which solves the problem (2.40) [76].
The dominant computational complexity of this algorithm is determined by that of the eigendecomposition and inversion of the matrix $\hat{R}$ and is equal to $O(M^3)$ [76]. It is worth noting that this complexity is equivalent to that of the SMI and LSMI algorithms.

Several further extensions of the robust MVDR beamformer of [60] have been recently developed by different authors. In [62], this beamformer has been extended to the case of ellipsoidal (anisotropic) uncertainty. The authors of [62] considered the following problem:

$$ \min_w w^H \hat{R} w \quad \text{subject to} \quad \Re\{w^H a_s\} \geq 1, \quad \text{for all} \quad a_s \in \mathcal{E} \quad (2.55) $$

where $\mathcal{E}$ is an ellipsoid that covers the possible range of uncertainty of the steering vector $a_s$. In [62], some opportunities to estimate optimal parameters of $\mathcal{E}$ from the received array data are discussed.

In [69], a covariance fitting-based interpretation of the robust MVDR problems of [60] and [62] has been developed. Although the problem in [69] is formulated in a different form as compared to that of [60] and [62], the authors of [69] have shown that such reformulated problem (which is referred to as a robust Capon beamformer in [69]) leads to exactly the same beamforming solutions as those in [60] and [62]. An additional useful feature of the approach of [69] is its ability to estimate the mismatched signal steering vector. An alternative Newton-type algorithm is derived in [69] to compute the weight vectors of the robust MVDR beamformers of [60] and [62]. The problem formulation of [69] is further modified in [78] by adding an ad hoc quadratic constraint.

In [76], the approach of [60] has been extended to robust multiuser detection problems. In [79], an efficient Kalman filter-based on-line implementation of the robust MVDR beamformer of [60] with the complexity of $O(M^2)$ per step has been developed.

In [61], the approach of [60] is extended to a more general case where, apart from the steering vector mismatch, there is a nonstationarity of the training data (which, as mentioned before, may be caused by the nonstationarity of interference and propagation channel, as well as antenna motion or vibration). To explain the results of [61], let us define the data matrix as

$$ X = [x(1), x(2), \ldots, x(N)] \quad (2.56) $$

Using (2.56), the sample covariance matrix (2.17) can be expressed as

$$ \hat{R} = \frac{1}{N} XX^H $$

The approach of [61] suggests to model the uncertainty which is caused by nonstationarities of the training data by means of adding this uncertainty to the data matrix.
Towards this end, let us introduce the mismatch matrix

$$\Delta = \tilde{X} - X$$

where $\tilde{X}$ and $X$ are, respectively, the actual and presumed data matrices in the test cell (at the beamforming sample). The presumed data matrix corresponds to the measured training cell data. In real-time adaptive beamforming problems, such training cell data correspond to the measurements that are made prior to the test cell. Thus, because of possible data nonstationarity effects, such past data snapshots may inadequately model the current test cell, where the actual (but unknown) data matrix is $\tilde{X}$ rather than $X$. Hence, in the nonstationary case, the actual sample covariance matrix can be expressed as

$$\hat{R} = \frac{1}{N} \tilde{X} \tilde{X}^H = \frac{1}{N} (X + \Delta)(X + \Delta)^H$$

(2.57)

According to (2.57), the matrix $\hat{R}$ is Hermitian and non-negative definite. However, this matrix is unknown because the mismatch $\Delta$ is unknown.

The authors of [61] proposed to combine the robustness against interference nonstationarity and steering vector errors using the ideas similar to that originally proposed in [60]. They assume that the norms of both the steering vector mismatch $\delta$ and the data matrix mismatch $\Delta$ are bounded by some known constants, that is,

$$\|\delta\| \leq \epsilon, \quad \|\Delta\|_F \leq \eta$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. Then, the weight vector can be found from maximizing the worst-case SINR, that is, by solving the following problem:

$$\max_w \min_{\delta, \Delta} \frac{\sigma^2 |w^H(a_s + \delta)|^2}{w^H \hat{R} w} \quad \text{subject to} \quad \|\delta\| \leq \epsilon, \quad \|\Delta\|_F \leq \eta$$

(2.58)

Using (2.58) and (2.57), the robust formulation of the MVDR beamforming problem takes the following form [61]:

$$\min_w \max_{\Delta \|F \leq \eta} \| (X + \Delta)^H w \| \quad \text{subject to} \quad |w^H (a_s + \delta)| \geq 1 \quad \text{for all} \|\delta\| \leq \epsilon$$

(2.59)
Note that this problem represents a further extension of (2.33) with additional robustness against nonstationary training data. The key idea of (2.59) is to minimize the beamformer output power in the scenario with the worst-case nonstationarity mismatch of the data matrix subject to the constraint which maintains the distortionless response for the worst-case steering vector mismatch. Note that the latter constant is the same as in (2.33), while the objective function is further modified with respect to (2.33).

To simplify the problem (2.59), the authors of [61] replaced the infinite number of constraints by a single worst-case constraint

\[
\min_{\|\delta\|\leq \epsilon} |w^H a_s + w^H \delta| \geq 1
\] (2.60)

in the same way as it was done in (2.34) and made use of Lemma 1 and the following Lemma.

**Lemma 3.**

\[
\max_{\|\Delta\|_F \leq \eta} \| (X + \Delta)^H w \| = \| X^H w \| + \eta \| w \|
\]

**Proof.** See Appendix 2.C.

Using Lemmas 1 and 3 along with (2.60), and taking into account that the cost function in (2.59) remains unchanged when \( w \) undergoes an arbitrary phase rotation [61], the problem (2.59) can be converted to

\[
\min_w \| X^H w \| + \eta \| w \| \quad \text{subject to} \quad w^H a_s \geq \epsilon \| w \| + 1
\] (2.61)

where, similar to (2.40), the constraint is satisfied with equality. This guarantees that (2.38) and (2.39) are satisfied automatically and, hence, there is no need to add them as additional constraints to (2.61).

Problem (2.61) can be viewed as an extended version of (2.40). Note that (2.61) also belongs to the class of SOC programming problems and can be efficiently solved using standard interior point method software [75]. Clearly, the robust beamformer (2.40) is a particular case of (2.61), because if we set \( \eta = 0 \) in (2.61) then it transforms to (2.40).

To further improve the robustness against moving interferers, the beamformer (2.61) can be combined with the MT method [61]. For that purpose, one should replace the matrix \( X \) in (2.61) by \( \sqrt{N} R_t^{1/2} \).

### 2.3.2 General-Rank Signal Case

Now, let us consider the general-rank signal case and consider the robust MVDR beamformer that has been recently derived in [59]. Following the philosophy of this work, we take into account that in practical situations, both the signal and
interference-plus-noise covariance matrices are known with some errors. In other words, there is always a certain mismatch between the actual and presumed values of these matrices. This yields

\[
\tilde{R}_s = R_s + \Delta_1 \\
\tilde{R}_{i+n} = R_{i+n} + \Delta_2
\]

where the presumed signal and interference-plus-noise covariance matrices are denoted as \(R_s\) and \(R_{i+n}\), respectively, while their actual values are denoted as \(\tilde{R}_s\) and \(\tilde{R}_{i+n}\), respectively. Here, \(\Delta_1\) and \(\Delta_2\) are the unknown matrix mismatches. These mismatches may occur because of a limited number of data snapshots that are used to estimate the signal and interference-plus-noise covariance matrices, environmental nonstationarities (such as rapid motion of the desired signal and interferers), signal location errors, and, moreover, due to the fact that in many applications, signal- and interference-free samples are usually unavailable.

In the presence of the mismatches \(\Delta_1\) and \(\Delta_2\), equation (2.2) for the output SINR of an adaptive beamformer must be rewritten as

\[
\text{SINR} = \frac{\mathbf{w}^H \tilde{R}_s \mathbf{w}}{\mathbf{w}^H \tilde{R}_{i+n} \mathbf{w}}
\]

Let the unknown mismatch matrices \(\Delta_1\) and \(\Delta_2\) be bounded in their norm by some known constants as [59]

\[
\|\Delta_1\|_F \leq \varepsilon, \quad \|\Delta_2\|_F \leq \gamma
\]

To provide robustness against such norm-bounded mismatches, the authors of [59] used the idea similar to [60], that is, they obtained the beamformer weight vector via maximizing the worst-case output SINR. This corresponds to the following optimization problem [59]

\[
\max_{\mathbf{w}} \min_{\Delta_1, \Delta_2} \frac{\mathbf{w}^H (R_s + \Delta_1) \mathbf{w}}{\mathbf{w}^H (R_{i+n} + \Delta_2) \mathbf{w}} \quad \text{subject to} \quad \|\Delta_1\|_F \leq \varepsilon, \|\Delta_2\|_F \leq \gamma \quad (2.64)
\]

where \(\Delta_1\) and \(\Delta_2\) are Hermitian matrices.

This problem can be rewritten as

\[
\max_{\mathbf{w}} \min_{\|\Delta_1\|_F \leq \varepsilon} \frac{\mathbf{w}^H (R_s + \Delta_1) \mathbf{w}}{\mathbf{w}^H (R_{i+n} + \Delta_2) \mathbf{w}}
\]

To solve (2.65), the following result can be used [59].
Lemma 4

\[
\begin{align*}
\min_{\|\Delta_1\| \leq \varepsilon} \mathbf{w}^H (\mathbf{R}_s + \Delta_1) \mathbf{w} &= \mathbf{w}^H (\mathbf{R}_s - \varepsilon \mathbf{I}) \mathbf{w} \\
\max_{\|\Delta_2\| \leq \gamma} \mathbf{w}^H (\mathbf{R}_{i+n} + \Delta_2) \mathbf{w} &= \mathbf{w}^H (\mathbf{R}_{i+n} + \gamma \mathbf{I}) \mathbf{w}
\end{align*}
\]

where the worst-case mismatch matrices \(\Delta_1\) and \(\Delta_2\) are given by

\[
\Delta_1 = -\varepsilon \frac{\mathbf{w} \mathbf{w}^H}{\|\mathbf{w}\|^2}, \quad \Delta_2 = \gamma \frac{\mathbf{w} \mathbf{w}^H}{\|\mathbf{w}\|^2}
\]

respectively.

Proof. See Appendix 2.D. \(\square\)

Using Lemma 4, the problem (2.65) can be converted to

\[
\max_{\mathbf{w}} \frac{\mathbf{w}^H (\mathbf{R}_s - \varepsilon \mathbf{I}) \mathbf{w}}{\mathbf{w}^H (\mathbf{R}_{i+n} + \gamma \mathbf{I}) \mathbf{w}}
\]

which, in turn, is equivalent to the following modified MVDR problem:

\[
\min_{\mathbf{w}} \mathbf{w}^H (\mathbf{R}_{i+n} + \gamma \mathbf{I}) \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H (\mathbf{R}_s - \varepsilon \mathbf{I}) \mathbf{w} = 1 \tag{2.67}
\]

Note that the problems (2.64) and (2.67) are equivalent if \(\varepsilon\) is smaller than the maximal eigenvalue of \(\mathbf{R}_s\). In the opposite case (when \(\varepsilon\) is larger than the maximal eigenvalue of \(\mathbf{R}_s\)), the matrix \(\mathbf{R}_s - \varepsilon \mathbf{I}\) is negative definite and (2.67) does not have any solution because the constraint in (2.67) cannot be satisfied. Therefore, the parameter \(\varepsilon\) which is smaller than the maximal eigenvalue of \(\mathbf{R}_s\) has to be chosen. A simple interpretation of this condition is that the allowed uncertainty in the signal covariance matrix should be sufficiently small.

Clearly, the structure of the problem (2.67) is similar to that of the problems (2.8) and (2.23). Using this fact, the solution to (2.67) can be expressed in the following form [59]:

\[
\mathbf{w}_{\text{rob}} = \mathcal{P}\{(\mathbf{R}_{i+n} + \gamma \mathbf{I})^{-1}(\mathbf{R}_s - \varepsilon \mathbf{I})\} \tag{2.68}
\]

In practical situations, the matrix \(\mathbf{R}_{i+n}\) is not available and the sample covariance matrix \(\hat{\mathbf{R}}\) should be used in lieu of \(\mathbf{R}_{i+n}\) in (2.67). The solution to such a modified problem yields the following sample version of the robust beamformer (2.68):

\[
\mathbf{w}_{\text{rob}} = \mathcal{P}\{\hat{\mathbf{R}} + \gamma \mathbf{I})^{-1}(\mathbf{R}_s - \varepsilon \mathbf{I}))} \tag{2.69}
\]
In the rank-one signal case, assuming without loss of generality that $\sigma_s^2 = 1$ (i.e.,
absorbing the constant $1/\sigma_s^2$ in $\varepsilon$), we obtain that the robust MVDR beamformer
(2.69) can be rewritten as

$$w_{rob} = \mathcal{P}\{(\tilde{R} + \gamma I)^{-1}(a_s a_s^H - eI)\} \quad (2.70)$$

From (2.69) it follows that the worst-case performance optimization approach of
[59] leads to a new diagonal loading-based beamformer which naturally combines
both the negative and positive types of diagonal loading, where the negative loading
is applied to the presumed covariance matrix of the desired signal $R_s$, while the
positive loading is applied to the sample covariance matrix $\tilde{R}$.

Setting $\varepsilon = 0$, we obtain that in this case (2.69) converts to the conventional
LSMI beamformer (2.24). Hence, this beamformer can be interpreted as a solution
to the worst-case performance optimization problem involving errors in the sample
covariance matrix. This explains a commonly known fact that diagonal loading can
be efficiently applied to a substantially broader class of problems than the small
sample size problem (which, however, was originally one of the main arguments
why to use diagonal loading).

Interestingly, the robust beamformer (2.69) offers a simpler and somewhat more
motivated way of choosing the parameters $\varepsilon$ and $\gamma$ as compared to the way of choos-
ing $\gamma$ in the diagonal loading method based on the white noise gain constraint.
Indeed, the choice of $\varepsilon$ and $\gamma$ in (2.69) is dictated by the physical parameters
of the environment (upper bounds on the covariance matrix mismatches). It
appears that in many practical situations it is relatively easy to obtain the parameters
$\gamma$ and $\varepsilon$ based on some preliminary knowledge of the type of environment
considered [38].

An important difference between the general-rank robust MVDR beamformer
(2.69) and rank-one robust MVDR beamformers (2.40) and (2.61) is that (2.69) is
not able to take into account the constraint that the actual signal covariance
matrix $R_s$ must be non-negative definite, while the techniques (2.61) and (2.69)
take into account this constraint. To clarify this point, note that the matrix $R_s$ in
(2.69) is not necessarily positive semidefinite. From the form of (2.70) it also
becomes clear that in the rank-one signal case, this matrix always has negative
eigenvalues if $\varepsilon > 0$. As a result, the aforementioned non-negative definiteness con-
straint is not satisfied in the problem (2.67). Ignoring this constraint may, in fact,
lead to an overly conservative approach (when more robustness than necessary is
provided) [38], although from the simulation results of [59] it follows that this
does not affect seriously the performance of (2.69).

An interesting interpretation of the robust beamformer (2.69) in terms of positive-
only diagonal loading has been obtained in [59]. According to (2.69), the weight
vector $w_{rob}$ satisfies the following characteristic equation

$$(\tilde{R} + \gamma I)^{-1}(R_s - eI)w_{rob} = \mu w_{rob} \quad (2.71)$$
where $\mu$ is the maximal eigenvalue of the matrix $(\hat{\mathbf{R}} + \gamma \mathbf{I})^{-1}(\mathbf{R}_s - \varepsilon \mathbf{I})$ and $\mathbf{w}_{\text{rob}}$ plays the role of the principal eigenvector of this matrix. Equation (2.71) can be rewritten as

$$(\mu \hat{\mathbf{R}} + (\mu \gamma + \varepsilon) \mathbf{I}) \mathbf{w}_{\text{rob}} = \mathbf{R}_s \mathbf{w}_{\text{rob}}$$

The latter equation is equivalent to

$$\left[ \hat{\mathbf{R}} + \left( \frac{\gamma + \varepsilon}{\mu} \right) \mathbf{I} \right]^{-1} \mathbf{R}_s \mathbf{w}_{\text{rob}} = \mu \mathbf{w}_{\text{rob}}$$

which implies that the robust beamformer (2.69) can be reinterpreted in terms of traditional (positive-only) diagonal loading with the *adaptive* loading factor $\gamma + \varepsilon / \mu$. However, it should be stressed that (2.73) is not a characteristic equation for the matrix $(\hat{\mathbf{R}} + (\gamma + \varepsilon / \mu) \mathbf{I})^{-1} \mathbf{R}_s$ because $\mu$ is involved in both left- and right-hand sides of (2.73). This fact poses major obstacles to find the weight vector $\mathbf{w}_{\text{rob}}$ directly from equation (2.73) and clarifies that (2.69) yields an easy way to solve equation (2.73) indirectly and in a closed form. However, equation (2.73) shows that the robust beamformer (2.69) that uses both the negative and positive types of diagonal loading is equivalent to the traditional diagonal loading method (with positive diagonal loading only) whose loading factor is selected adaptively, to optimally match to the given amount of uncertainty in the signal and data covariance matrices.

An efficient on-line implementation of the robust MVDR beamformer (2.69) has been developed in [59] where the following lemma has been proved.

**Lemma 5.** For arbitrary $M \times M$ Hermitian matrix $\mathbf{Y}$ and arbitrary $M \times M$ full-rank Hermitian matrix $\mathbf{Z}$ the following relationship holds

$$\mathcal{P}\{\mathbf{Y}\mathbf{Z}\} = \mathbf{Z}^{-1/2}\mathcal{P}\{\mathbf{Z}^{1/2}\mathbf{Y}\mathbf{Z}^{1/2}\}$$

(2.74)

**Proof.** See Appendix 2.E. \hfill \Box

Applying this lemma to the beamformer (2.69), we rewrite it as

$$\mathbf{w}_{\text{rob}} = (\mathbf{R}_s - \varepsilon \mathbf{I})^{-1/2}\mathcal{P}\{(\mathbf{R}_s - \varepsilon \mathbf{I})^{1/2}(\hat{\mathbf{R}} + \gamma \mathbf{I})^{-1}(\mathbf{R}_s - \varepsilon \mathbf{I})^{1/2}\}$$

$$= (\mathbf{R}_s - \varepsilon \mathbf{I})^{-1/2}\mathcal{P}\{\mathbf{G}^{-1}\}$$

(2.75)

where the matrix $\mathbf{G}$ is defined as

$$\mathbf{G} \triangleq (\mathbf{R}_s - \varepsilon \mathbf{I})^{-1/2}(\hat{\mathbf{R}} + \gamma \mathbf{I})(\mathbf{R}_s - \varepsilon \mathbf{I})^{-1/2}$$

(2.76)

It is noteworthy that even if the matrix $\mathbf{R}_s$ is singular or ill-conditioned, the matrix $\mathbf{R}_s - \varepsilon \mathbf{I}$ can be made full-rank (well-conditioned) by a proper choice of the parameter $\varepsilon$. Furthermore, for any nonzero $\varepsilon$, $\text{rank}\{\mathbf{R}_s - \varepsilon \mathbf{I}\} = M$ almost surely.
To develop an on-line implementation of the beamformer (2.69), let us consider the case of rectangular sliding window of the length \(N\) where the update of the matrix \(\hat{\mathbf{R}}_{dl} = \hat{\mathbf{R}} + \gamma \mathbf{I}\) in the \(n\)th step can be computed as [80]

\[
\hat{\mathbf{R}}_{dl}(n) = \hat{\mathbf{R}}_{dl}(n-1) + \frac{1}{N} \mathbf{x}(n)\mathbf{x}^H(n) - \frac{1}{N} \mathbf{x}(n-N)\mathbf{x}^H(n-N) \tag{2.77}
\]

Note that (2.77) represents the so-called rank-two update [80]. The diagonal load should be added to the initialization step of (2.77), that is, \(\gamma \mathbf{I}\) should be chosen to initialize the matrix \(\hat{\mathbf{R}}_{dl}\). Using (2.77), we can rewrite the corresponding update of the matrix (2.76) as

\[
\mathbf{G}(n) = \mathbf{G}(n-1) + \tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^H(n) - \tilde{\mathbf{x}}(n-N)\tilde{\mathbf{x}}^H(n-N) \tag{2.78}
\]

where the transformed training snapshots are defined as

\[
\tilde{\mathbf{x}}(i) = \frac{1}{\sqrt{N}} (\mathbf{R}_s - \varepsilon \mathbf{I})^{-1/2} \mathbf{x}(i)
\]

and, according to (2.76), \(\gamma (\mathbf{R}_s - \varepsilon \mathbf{I})^{-1}\) should be chosen to initialize the matrix \(\mathbf{G}\).

According to equations (2.75) and (2.78), on-line algorithms for updating the weight vector \(\mathbf{w}_{\text{rob}}\) should be based on combining the matrix inversion lemma and some subspace tracking algorithm to track the principal eigenvector of the matrix \(\mathbf{G}^{-1}\). Any of subspace tracking algorithms available in the literature can be used for this purpose [80, 81]. As the complexities of the existing subspace tracking techniques lie between \(O(M)\) and \(O(M^2)\) per step, the total complexity of this on-line implementation of the robust MVDR beamformer (2.69) is \(O(M^2)\) per step [59]. This conclusion can be made because, regardless of the complexity of the subspace tracking algorithm used, \(O(M^2)\) operations per step are required to update the weight vector (2.75).

Further extensions of the worst-case approach of [59] to the robust blind multiuser detection problem can be found in [82].

### 2.4 NUMERICAL EXAMPLES

In all numerical examples, we assume a uniform linear array (ULA) of \(M = 20\) omnidirectional sensors spaced half-wavelength apart. All the results are averaged over 100 simulation runs. Throughout all examples, we assume that there is one desired and one interfering source. The desired signal is assumed to be always present in the training data cell and the interference-to-noise ratio (INR) is equal to 20 dB. We compare the performances of the benchmark SMI beamformer, conventional SMI beamformer, LSMI beamformer with fixed diagonal loading, and our robust MVDR beamformers (2.40) and (2.69) with adaptive diagonal loading (these techniques are referred to as the rank-one and general-rank robust beamformers, respectively). Note that the benchmark SMI beamformer corresponds to the ideal case when the matrix \(\mathbf{R}_s\) in (2.18) is known exactly. This algorithm does not
correspond to any real situation and is included in our simulations for the sake of comparison only. All other beamformers tested use a mismatched covariance matrix (or steering vector) of the desired signal. Following [59], the diagonal loading parameter $\gamma = 30$ is chosen for the LSMI algorithm (2.24) and our robust algorithm (2.69) in all examples.

Additionally, the optimal SINR curve is displayed in each figure.

In our first example, we consider a point source scenario where $\text{rank}\{\hat{R}_s\} = \text{rank}\{R_s\} = 1$. Both the desired signal and interferer are assumed to be plane waves impinging on the array from the directions $20^\circ$ and $-20^\circ$, respectively, while the presumed signal direction is equal to $22^\circ$. That is, there is the $2^\circ$ signal look direction mismatch in this scenario.

Figure 2.2 displays the output SINRs of the beamformers tested versus $N$ for $\text{SNR} = 0$ dB. The SINRs of the same beamformers are shown in Figure 2.3 versus SNR for $N = 100$. The parameters $\epsilon = 4$ and $\epsilon = 16$ are chosen for the robust beamformers (2.40) and (2.69), respectively.\(^2\)

In the second example, again a point source scenario is considered where the steering vector of the desired signal and interferer are plane wavefronts impinging on the array from $30^\circ$ and $-30^\circ$, respectively, and are additionally distorted in phase. For both wavefronts and in each run, these phase distortions have been independently and randomly drawn from a Gaussian random generator with zero mean and the variance of 0.2. Note that the distortions change from run to run but remain fixed from snapshot to snapshot. The presumed signal steering vector does not take into account any distortions, that is, it corresponds to a plane wave with the DOA of $30^\circ$. This example models the case of coherent scattering, imperfectly calibrated array, or wavefront perturbation in an inhomogeneous medium [60]. In wireless communications, such scenario may be used to model the case of spatial signature estimation errors caused by a limited amount of pilot symbols.

Figure 2.4 displays the output SINRs of the beamformers tested versus $N$ for the fixed $\text{SNR} = 0$ dB in the second example. The performance of the same methods versus the SNR for the fixed training data size $N = 100$ is shown in Figure 2.5.

In the third example, a scenario with non-point full-rank sources is considered. In this example, we assume locally incoherently scattered desired signal and interferer with Gaussian and uniform angular power densities characterized by the central angles of $30^\circ$ and $-30^\circ$, respectively. Each of these sources is assumed to have the same angular spread equal to $4^\circ$. The presumed signal covariance matrix, however, ignores local scattering effects and corresponds to the case of a point (rank-one) plane wavefront source with the DOA of $32^\circ$. The parameters $\epsilon = 3$ and $\epsilon = 9$ are chosen in this example.

Figure 2.6 shows the performances of the methods tested versus $N$ for the fixed $\text{SNR} = 0$ dB. The performance of the same methods versus the SNR for the fixed training data size $N = 100$ is displayed in Figure 2.7.

\(^2\)Note that the choice of $\epsilon = 4$ is consistent to the choice of $\epsilon = 16$ because $\epsilon$ is related to the Euclidean norm of the signal steering vector mismatch, whereas $\epsilon$ is related to the Frobenius norm of the signal covariance matrix mismatch.
Figure 2.2  Output SINRs versus $N$, first example.

Figure 2.3  Output SINRs versus SNR; first example.
2.4 NUMERICAL EXAMPLES

Figure 2.4 Output SINRs versus $N$; second example.

Figure 2.5 Output SINRs versus SNR; second example.
Figure 2.6  Output SINRs versus $N$; third example.

Figure 2.7  Output SINRs versus SNR; third example.
Similar to the third example, in our last example we assume a scenario with non-point full-rank sources. We model incoherently scattered desired signal and interferer with the Gaussian and uniform angular power densities and the central angles of $20^\circ$ and $-20^\circ$, respectively. Each of these sources is assumed to have the same angular spread equal to $4^\circ$. In contrast to the previous example, the presumed covariance matrix is also full rank and corresponds to a Gaussian incoherently distributed source with the central angle of $22^\circ$ and angular spread of $6^\circ$. That is, there is a signal mismatch both in the central angle and angular spread. In this example, $\varepsilon = 9$ is taken (note that the rank-one robust MVDR beamformer (2.40) is not applicable to this example and its performance is not shown).

Figure 2.8 depicts the performance of the methods tested versus $N$ for the fixed SNR $= 0$ dB. The performance of these methods versus the SNR for the fixed training data size $N = 100$ is shown in Figure 2.9.

### 2.4.1 Discussion

Figures 2.2–2.9 clearly demonstrate that in all our simulation examples, the robust MVDR beamformers (2.40) and (2.69) consistently outperform the other beamformers tested and achieve the SINR that is close to the optimal one for all tested values of SNR and $N$. This conclusion holds true for both the rank-one and full-rank signal scenarios considered in our examples and shows that the performance losses remain small compared to the ideal (nonmismatched) case.
In all examples where both the beamformers (2.40) and (2.69) are tested, their performance can be observed to be nearly identical. However, in all examples these robust MVDR techniques outperform the SMI and LSMI beamformers. These performance improvements are especially pronounced at high SNRs.

Interestingly, the robust MVDR beamformers (2.40) and (2.69) not only substantially outperform the SMI and LSMI beamformers, but also perform better than the benchmark SMI beamformer. This can be explained by the fact that, although the benchmark SMI beamformer perfectly knows the signal covariance matrix $R_s$, it exploits the sample estimate $\hat{R}$ of the interference-plus-noise covariance matrix and, because of this, it suffers from severe signal self-nulling.

### 2.5 CONCLUSIONS

This chapter has provided an overview of the main advances in the area of robust adaptive beamforming. After reviewing the required types of robustness and known ad hoc solutions, a recently emerged rigorous approach to robust adaptive beamforming based on the worst-case performance optimization has been addressed in detail. This approach greatly improves the robustness of traditional minimum variance beamformers in the presence of various types of unknown mismatches and nonidealities. Both the rank-one and general-rank signal cases have been investigated in detail. Several state-of-the-art robust MVDR beamformers that are able
to achieve different robustness tradeoffs have been introduced and studied in these cases. These algorithms include both closed-form solutions and convex optimization-based techniques which can be efficiently implemented using modern convex optimization algorithms and software and whose order of computational complexity is similar to that of the traditional SMI and LSMI adaptive beamformers.

**APPENDIX 2.A: Proof of Lemma 1**

Using the triangle and Cauchy-Schwarz inequalities along with the inequality (2.32) yields

\[
|w^H a_s + w^H \delta| \geq |w^H a_s| - |w^H \delta| \geq |w^H a_s| - \epsilon \|w\| \tag{A.1}
\]

Also, it can be readily verified that

\[
|w^H a_s + w^H \delta| = |w^H a_s| - \epsilon \|w\| \tag{A.2}
\]

if \( |w^H a_s| > \epsilon \|w\| \) and if

\[
\delta = -\frac{w}{\|w\|} \epsilon e^{j\phi}
\]

where

\[
\phi = \text{angle}\{w^H a_s\}
\]

Combining (A.1) and (A.2), we prove the lemma.

**APPENDIX 2.B: Proof of Lemma 2**

We first show that if \( \epsilon < \|a_s\| \) then the solution of \( f(\tau) = 0 \) is a positive value. To show this, we note that

\[
f(0) = \frac{\sum_{i=1}^{M} |g_i|^2}{e^2} - 1
\]

\[
= \frac{\|g\|^2}{e^2} - 1
\]

\[
= \frac{\|a_s\|^2}{e^2} - 1 \tag{B.1}
\]

where in the last row of (B.1) we have used the equation \( \|g\| = \|a_s\| \) which follows from (2.50) and the fact that the matrix \( U \) is unitary. If \( \epsilon < \|a_s\| \), then from (B.1) it is clear that \( f(0) > 0 \). On the other hand, according to (2.52), \( f(+\infty) = -1 \) and,
since \( f(\tau) \) is continuous for positive values of \( \tau \), it has a root in the interval \((0, +\infty)\). This completes the proof of the sufficiency part of Lemma 2.

The necessity of the condition \( \varepsilon < \|a_1\| \) for \( f(\tau) = 0 \) to have a positive solution can be proved by contradiction. Assume that the equation \( f(\tau) = 0 \) has a positive solution while \( \varepsilon \geq \|a_1\| \). Since \( \tau \) and \( \{\xi_i\}_{i=1}^M \) are all positive, using the definition of \( f(\tau) \) in (2.52), we conclude that for any positive \( \tau \)

\[
\begin{align*}
f(\tau) &< \frac{\sum_{i=1}^M |g_i|^2}{\varepsilon^2} - 1 \\
&= \frac{\|g\|^2}{\varepsilon^2} - 1 \\
&= \frac{\|a_1\|^2}{\varepsilon^2} - 1 \quad (B.2)
\end{align*}
\]

If \( \varepsilon \geq \|a_1\| \), it follows from (B.2) that \( f(\tau) < 0 \) for all positive values of \( \tau \). This is an obvious contradiction to the assumption that \( f(\tau) \) is zero for some positive \( \tau \). The necessity part of Lemma 2 is proven.

The proof of uniqueness is as follows. Assume that \( \tau_1 \) and \( \tau_2 \) are two positive values of \( \tau \) such that \( f(\tau_1) = f(\tau_2) \). Then, using (2.52), we can write

\[
\sum_{i=1}^M \left( \frac{|g_i|}{\varepsilon + \tau_1 \xi_i} \right)^2 - \sum_{i=1}^M \left( \frac{|g_i|}{\varepsilon + \tau_2 \xi_i} \right)^2 = 0
\]

which means that

\[
(\tau_2 - \tau_1) \sum_{i=1}^M |g_i|^2 \xi_i [2\varepsilon + \xi_i (\tau_2 + \tau_1)] \left[ \varepsilon + \tau_1 \xi_i \right]^2 \left[ \varepsilon + \tau_2 \xi_i \right]^2 = 0
\]

where, because of the positiveness of \( \tau_1, \tau_2 \) and \( \xi_i \) \((i = 1, \ldots, M)\),

\[
\sum_{i=1}^M |g_i|^2 \xi_i [2\varepsilon + \xi_i (\tau_2 + \tau_1)] \left[ \varepsilon + \tau_1 \xi_i \right]^2 \left[ \varepsilon + \tau_2 \xi_i \right]^2 > 0
\]

This means that \( \tau_1 = \tau_2 \) and, therefore, the solution to \( f(\tau) = 0 \) is unique. With this statement, the proof of Lemma 2 is complete.

**APPENDIX 2.C: Proof of Lemma 3**

Let us introduce

\[
f(w) \triangleq \max_{\|\Delta\| \leq \eta} \| (X + \Delta)^H w \|
\]
First of all, we will show that

\[ f(w) \leq \|X^H w\| + \eta \|w\| \]  \hspace{1cm} (C.1)

For any matrix \( \Delta \), we have that \( \|\Delta\| \leq \|\Delta\|_F \) (recall here that \( \|\cdot\| \) denotes the matrix 2-norm). Therefore, for any \( \Delta \), we obtain

\[
\|X^H w + \Delta^H w\| \leq \|X^H w\| + \|\Delta^H w\|
\]

\[
\leq \|X^H w\| + \|\Delta\||w|
\]

\[
\leq \|X^H w\| + \|\Delta\|_F \|w\|
\]

\[
\leq \|X^H w\| + \eta \|w\|
\]

and (C.1) is proved.

Next, we show that

\[ f(w) \geq \|X^H w\| + \eta \|w\| \]  \hspace{1cm} (C.2)

Introducing

\[ \Delta_* \triangleq \frac{\eta w w^H X}{\|w\| \|X^H w\|} \]

and using the property

\[ \|\Delta_*\|_F^2 = \text{trace}\{\Delta_*^H \Delta_*\} \]

it is easy to verify that \( \|\Delta_*\|_F = \eta \). Therefore,

\[
f(w) = \max_{\|\Delta\|_F \leq \eta} \|X + \Delta^H w\|
\]

\[
\geq \|X + \Delta_*^H w\|
\]

\[
= \|X^H w + \frac{\eta X^H w w^H}{\|w\| \|X^H w\|} w\|
\]

\[
= \|X^H w + \frac{\eta \|w\|}{\|X^H w\|} X^H w\|
\]

\[
= \|X^H w\| + \eta \|w\| \]  \hspace{1cm} (C.3)

With (C.3), equation (C.2) is proved. Comparing (C.1) and (C.2), we finally prove Lemma 3.
APPENDIX 2.D: Proof of Lemma 4

Let us solve the following constrained optimization problems

\[
\begin{align*}
\min_{\Delta_1} & \quad w^H (R_s + \Delta_1) w \\
\text{subject to} & \quad \|\Delta_1\|_F \leq \varepsilon
\end{align*}
\]

\[
\begin{align*}
\max_{\Delta_2} & \quad w^H (R_{i+n} + \Delta_2) w \\
\text{subject to} & \quad \|\Delta_2\|_F \leq \gamma
\end{align*}
\]

We observe that the objective functions in (D.1) and (D.2) are linear because they
are minimized (or maximized) with respect to \(\Delta_1\) (or \(\Delta_2\)) rather than \(w\). From the
linearity of these objective functions, it follows that the inequality constraints in
(D.1) and (D.2) are satisfied with equality. Therefore, the solutions to (D.1) and
(D.2) can be obtained using Lagrange multipliers method, by means of minimiz-
ing/maximizing the functions

\[
L(\Delta_1, \lambda) = w^H (R_s + \Delta_1) w + \lambda (\|\Delta_1\|_F - \varepsilon)
\]

\[
L(\Delta_2, \tilde{\lambda}) = w^H (R_{i+n} + \Delta_2) w + \tilde{\lambda} (\|\Delta_2\|_F - \gamma)
\]

respectively, where \(\lambda\) and \(\tilde{\lambda}\) are the corresponding Lagrange multipliers. Equating
the gradients \(\partial L(\Delta_1, \lambda)/\partial \Delta_1\) and \(\partial L(\Delta_2, \tilde{\lambda})/\partial \Delta_2\) to zero yields

\[
\Delta_1 = -\frac{1}{2\lambda} ww^H, \quad \Delta_2 = -\frac{1}{2\tilde{\lambda}} ww^H
\]

(D.3)

Using \(\|\Delta_1\|_F = \varepsilon\) and \(\|\Delta_1\|_F = \gamma\) along with (D.3), we obtain

\[
\Delta_1 = -\varepsilon \frac{ww^H}{\|w\|^2}, \quad \Delta_2 = \gamma \frac{ww^H}{\|w\|^2}
\]

(D.4)

where the signs in (D.4) are determined by the fact that (D.1) and (D.2) are the
minimization and maximization problems, respectively. Using (D.4) yields

\[
\min_{\|\Delta_1\| \leq \varepsilon} w^H (R_s + \Delta_1) w = w^H (R_s - \varepsilon \frac{ww^H}{\|w\|^2}) w
\]

\[
= w^H (R_s - \varepsilon I) w
\]

\[
\min_{\|\Delta_2\| \leq \gamma} w^H (R_{i+n} + \Delta_2) w = w^H (R_{i+n} + \gamma \frac{ww^H}{\|w\|^2}) w
\]

\[
= w^H (R_{i+n} + \gamma I) w
\]

respectively, and the proof of Lemma 4 is complete.
APPENDIX 2.E: Proof of Lemma 5

Let us write the characteristic equation for the matrix $YZ$ as

$$YZu_i = \mu_i u_i$$  \hspace{1cm} (E.1)

where $\{\mu_i\}_{i=1}^M$ and $\{u_i\}_{i=1}^M$ are the eigenvalues and corresponding eigenvectors of the matrix $YZ$. Multiplying this equation by $Z^{1/2}$ yields

$$Z^{1/2}YZ^{1/2}u_i = \mu_i Z^{1/2}u_i$$  \hspace{1cm} (E.2)

which is also the characteristic equation for the matrix $Z^{1/2}YZ^{1/2}$, that is

$$Z^{1/2}YZ^{1/2}v_i = \mu_i v_i$$  \hspace{1cm} (E.3)

where the eigenvectors of the matrices $YZ$ and $Z^{1/2}YZ^{1/2}$ are related as

$$v_i = Z^{1/2}u_i$$  \hspace{1cm} (E.4)

for all $i = 1, 2, \ldots, M$. Applying this result to the principal eigenvectors of the matrix $YZ$ and $Z^{1/2}YZ^{1/2}$, we obtain (2.74) and Lemma 5 is proved.

REFERENCES


