CCIT Report #449 October 2003

Robust Linear Estimation Under a Minimax MSE–Ratio Criterion

Yonina C. Eldar and Neri Merhav*

SP_EDICS category: 2-ESTM.

Abstract

In continuation to an earlier work, we further consider the problem of robust estimation of a random vector (or signal), with an uncertain covariance matrix, that is observed through a known linear transformation and corrupted by additive noise with a known covariance matrix. While in the earlier work, we developed and proposed a competitive minimax approach of minimizing the worst–case mean-squared error (MSE) *difference regret* criterion, here we study, in the same spirit, the minimum worst–case MSE *ratio regret* criterion, namely, the worst–case *ratio* (rather than difference) between the MSE attainable using a linear estimator, ignorant of the exact signal covariance, and the minimum MSE (MMSE) attainable by optimum linear estimation with a known signal covariance. We present the optimal linear estimator, under this criterion, in two ways: The first is as a solution to a certain semidefinite programming (SDP) problem, and the second is as an expression which is of closed form up to a single parameter whose value can be found by a simple line search procedure. We then show that the linear minimax ratio regret estimator can also be interpreted as the MMSE estimator that minimizes the MSE for a certain choice of signal covariance that depends on the uncertainty region. We demonstrate that in applications, the proposed minimax MSE ratio regret approach may outperform the well–known minimax MSE approach, the minimax MSE difference regret approach, and the "plug–in" approach, where in the latter, one uses the MMSE estimator with an estimated covariance matrix replacing the true unknown covariance.

*Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: {yonina,merhav}@ee.technion.ac.il. Fax: +972-4-8295757.

1 Introduction

The classical solutions to the problem of optimum linear estimation and filtering, dating back to Wiener [1] and Kolmogorov [2], are well-known to be sensitive to the exact knowledge of the second order statistics of the signal and noise. However, in a wide range of practical applications, these statistics may, unfortunately, be subjected to uncertainties. This fact has been the origin of the need for solid theories and methodologies of designing robust estimators and filters whose performance remains relatively insensitive and reasonably good across the region of uncertainty.

The most common approach to handle such uncertainties, has been the minimax approach, initiated by Huber [4, 5]. According to this approach, an estimator is sought to minimize the worst–case mean-squared error (MSE) over a given uncertainty class of spectral densities [6, 7, 8, 3, 9]. The idea of using minimax criteria for devising robust schemes that circumvent uncertainties, has been applied in quite a few additional problem areas in communications, signal processing, and statistics (see, e.g., [11, 3] and references therein).

In spite of the widespread use of the minimax approach, its performance sometimes turns out to be disappointingly poor. The reason seems to be rooted in the very definition of the minimax criterion, which is fundamentally pessimistic in nature: Optimizing performance for the worst case, might come at the expense of deteriorated performance in all other cases, since in this worst case, the conditions may be so poor, that they leave very little or no room for powerful solution strategies.

In light of this fact, to improve the performance of the ordinary minimax MSE approach, we have studied, in an earlier work [28], a modified minimax criterion, which is *competitive* in character. This competitive minimax criterion has been derived in [28] in the context of a simple linear model. Specifically, given an observation vector, resulting from a known linear transformation of the desired vector (or signal) to be estimated, and corrupted by an uncorrelated additive noise vector with a known covariance matrix, we sought in [28] a linear estimator, which is robust to covariance uncertainties of the desired signal, using the following approach: Rather than minimizing the worst–case (total) MSE, we derived the linear estimator that minimizes the worst–case *difference regret*, namely, the worst–case difference between the MSE of a linear estimator, ignorant of the exact signal covariance, and the MSE of the linear optimal estimator based on the exact signal covariance. The rationale was that such an estimator performs uniformly as close as possible to the linear optimal estimator across the uncertainty region, and since the minimax criterion is applied to the difference of MSE's, rather than the total MSE, it is not so pessimistic as the ordinary minimax approach. The same idea was also applied in [19] for the case where the unknown desired vector is *deterministic* rather than stochastic. As we have pointed out in [28], the competitive minimax approach is by no means new, as a general concept. It has been used extensively in a variety of other problem areas, such as, universal source coding [20], hypothesis testing [21, 22], and prediction (see [23] for a survey and references therein).

A possible drawback of the difference regret approach is that the value of the regret may not adequately reflect the estimator performance, since even a large regret should be considered insignificant if the value of the optimal MSE is relatively large. On the other hand, if the optimal MSE is small, then even a small regret should be considered significant. Therefore, instead of considering the worst-case difference regret we suggest developing a minimax ratio estimator that minimizes the worst-case *ratio* between the MSE of a linear estimator that does not know the signal covariance, and the best possible MSE. Generally speaking, the rationale is that, as before, such an estimator performs uniformly as close as possible to the linear optimal estimator across the uncertainty region, but where now the MSE is measured in dB. This makes sense as one might expect the relative loss in MSE performance to be scale–invariant.

In this paper, we study robust linear estimation for the same linear model under the criterion of minimax *ratio–regret*, rather than the difference–regret. We present the optimal linear estimator, under this new criterion, in two ways: The first representation is as a solution to a certain semidefinite programming (SDP) problem. This is practically meaningful since SDP programs are efficiently executable using standard software packages. The second representation is as an analytical expression, which is of closed form up to a single parameter whose value can be found by a simple line search procedure. We also show that the linear minimax ratio regret estimator, or for short, the minimax ratio estimator, can be interpreted as the MMSE estimator corresponding to a certain choice of signal covariance that depends on the uncertainty region. We demonstrate that in applications, the proposed minimax MSE ratio–regret approach may outperform the ordinary MSE approach, the minimax MSE difference–regret approach, and the "plug–in" approach, where in the latter, one uses the MMSE estimator with an estimated covariance matrix replacing the true unknown covariance.

The outline of this paper is as follows. In Section 2, we formulate the problem. In Section 3, we present the problem as an SDP. In Section 4, the alternative, closed–form solution is derived, and finally, in Section 5, performance is demonstrated through several examples.

2 Problem Formulation

In the sequel, we denote vectors in $\mathbb{C}^m$ by boldface lowercase letters and matrices in $\mathbb{C}^{n\times m}$ by boldface uppercase letters. The matrix $\mathbf{I}$ denotes the identity matrix of the appropriate dimension, $(\cdot)^*$ denotes the Hermitian conjugate of the corresponding matrix, and $\hat{(\cdot)}$ denotes an estimated vector. The cross-covariance matrix between the random vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\mathbf{C}_{xy}$, and the covariance matrix of $\mathbf{x}$ is denoted by $\mathbf{C}_x$.

Consider the problem of estimating the unknown parameters $\mathbf{x}$ in the linear model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}, \tag{1}$$

where $\mathbf{H}$ is a known $n \times m$ matrix with rank m, $\mathbf{x}$ is a zero-mean, length-m random vector with covariance matrix $\mathbf{C}_x$ and $\mathbf{w}$ is a zero-mean, length-n random vector with known positive definite covariance $\mathbf{C}_w$, uncorrelated with $\mathbf{x}$. We assume that we only have partial information about the covariance $\mathbf{C}_x$.

We seek to estimate $\mathbf{x}$ using a linear estimator so that $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ for some $m \times n$ matrix $\mathbf{G}$. We would like to design an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ to minimize the MSE, which is given by

$$\begin{aligned} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) &= \mathrm{Tr}(\mathbf{C}_x) + \mathrm{Tr}(\mathbf{C}_{\hat{x}}) - 2\mathrm{Tr}(\mathbf{C}_{x\hat{x}}) \\ &= \mathrm{Tr}(\mathbf{C}_x) + \mathrm{Tr}(\mathbf{G}(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)\mathbf{G}^*) - 2\mathrm{Tr}(\mathbf{C}_x\mathbf{H}^*\mathbf{G}^*) \\ &= \mathrm{Tr}\left(\mathbf{G}\mathbf{C}_w\mathbf{G}^*\right) + \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\right). \end{aligned} \tag{2}$$

If $\mathbf{C}_x$ is known, then the linear estimator minimizing (2) is the *MMSE estimator* [14]

$$\hat{\mathbf{x}} = \mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}\mathbf{y}. \tag{3}$$

An alternative form for $\hat{\mathbf{x}}$, that is sometimes more convenient, can be obtained by applying the matrix inversion lemma [24] to $(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}$ resulting in

$$(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1} = \mathbf{C}_w^{-1} - \mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}. \tag{4}$$

Substituting (4) into (3), the MMSE estimator $\hat{\mathbf{x}}$ can be expressed as

$$\hat{\mathbf{x}} = \mathbf{C}_x\left(\mathbf{I} - \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\right)\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}$$

$$= \mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \quad (5)$$

If $\mathbf{C}_x$ is unknown, then we cannot implement the MMSE estimator (3). In this case, we may choose an estimator to optimize a worst-case performance measure, over all covariance matrices in the region of uncertainty. To reflect the uncertainty in our knowledge of the true covariance matrix, we consider an uncertainty model that resembles the "band model" widely used in the continuous-time case [7, 25, 26, 3], and is the same as the model considered in [28]. Specifically, we assume that $\mathbf{C}_x$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix[1], and that each of the nonnegative eigenvalues $\delta_i \geq 0, 1 \leq i \leq m$ of $\mathbf{C}_x$ satisfies

$$l_i \leq \delta_i \leq u_i, \quad 1 \leq i \leq m, \quad (6)$$

where $l_i \geq 0$ and u_i are known.

The assumption that $\mathbf{C}_x$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have the same eigenvector matrix is made for analytical tractability. If $\mathbf{x}$ is a stationary random vector and $\mathbf{H}$ represents convolution of $\mathbf{x}$ with some filter, then both $\mathbf{C}_x$ and $\mathbf{H}$ will be Toeplitz matrices and are therefore approximately diagonalized by a Fourier transform matrix, so that in this general case $\mathbf{C}_x$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ approximately have the same eigenvector matrix [27].

In our development, we explicitly assume that the joint eigenvector matrix of $\mathbf{C}_x$ and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ is given. In practice, if the eigenvalues of $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have geometric multiplicity one, then we choose the eigenvector matrix of $\mathbf{C}_x$ to be equal to the eigenvector matrix of $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. In the case in which the eigenvector matrix of $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ is not uniquely specified, *e.g.,* in the case in which $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ is proportional to $\mathbf{I}$, as in one of the examples in Section 5, we may resolve this ambiguity by estimating the eigenvector matrix of $\mathbf{C}_x$ from the data.

The model (6) is reasonable when the covariance is estimated from the data. Specifically, denoting by $\zeta_i = (u_i + l_i)/2, \epsilon_i = (u_i - l_i)/2$ for $1 \leq i \leq m$, the conditions (6) can equivalently be expressed as

$$\delta_i = \zeta_i + e_i, \quad e_i^2 \leq \epsilon_i^2, \quad 1 \leq i \leq m, \quad (7)$$

so that each of the eigenvalues of $\mathbf{C}_x$ lies in an interval of length $2\epsilon_i$ around some nominal value ζ_i which we can think of as an estimate of the ith eigenvalue of $\mathbf{C}_x$ from the data vector $\mathbf{y}$. The interval specified by ϵ_i may be regarded as a confidence interval around our estimate ζ_i and can be chosen to be proportional to the standard deviation of the estimate ζ_i.

[1] If the eigenvalues of $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and $\mathbf{C}_x$ have geometric multiplicity 1, then $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and $\mathbf{C}_x$ have the same eigenvector matrix if and only if they commute [24].

Given $\{\zeta_i\}$, a straightforward approach to estimating $\mathbf{x}$ is to use an MMSE estimate corresponding to the estimated covariance. However, as demonstrated in [28] and in Section 5, by taking an uncertainty interval around ζ_i into account, and seeking a competitive minimax estimator in this interval, we can further improve the estimation performance.

To develop a competitive estimator, we consider a minimax ratio criterion, in which the estimator is obtained by minimizing the worst-case *ratio* between the MSE of a linear estimator that does not know the signal covariance, and the best possible MSE. In Section 3, we show that the minimax ratio estimator can be formulated as an SDP. In Section 4 we use the necessary and sufficient conditions for optimality of an SDP to develop more insight into the minimax ratio estimator. Specifically, we show that the minimax ratio estimator is an MMSE estimator matched to a covariance matrix which depends on a single parameter that can be found using a simple line search algorithm, for example, using the bisection method. In the examples in Section 5, we demonstrate that the minimax ratio estimator can improve the performance over the minimax MSE estimator and the minimax regret estimator of [28] for low SNR values.

3 The Minimax Ratio Estimator

We seek the linear estimator $\hat{\mathbf{x}}$ that minimizes the worst-case ratio $\mathcal{R}(\mathbf{C}_x, \mathbf{G})$, which is defined as the ratio between the MSE using an estimator $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$, and the smallest possible MSE attainable with an estimator of the form $\hat{\mathbf{x}} = \mathbf{G}(\mathbf{C}_x)\mathbf{y}$ when the covariance $\mathbf{C}_x$ is known, which we denote by MSE^o.

If $\mathbf{C}_x$ is known, then the MMSE estimator is given by (3) and the resulting optimal MSE is

$$\mathrm{MSE}^o = \mathrm{Tr}(\mathbf{C}_x) - \mathrm{Tr}(\mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}\mathbf{H}\mathbf{C}_x). \tag{8}$$

From (4) and (5), we have that $\mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1} = \mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$, so that (8) can be written in the equivalent form,

$$\mathrm{MSE}^o = \mathrm{Tr}\left(\mathbf{C}_x\left(\mathbf{I} - (\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x\right)\right) = \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\right), \tag{9}$$

which will be more convenient for our derivations.

Thus, we seek the matrix $\mathbf{G}$ that is the solution to the problem

$$\min_{\mathbf{G}} \max_{l_i \le \delta_i \le u_i} \mathcal{R}(\mathbf{C}_x, \mathbf{G}), \tag{10}$$

where

$$\mathcal{R}(\mathbf{C}_x, \mathbf{G}) = \frac{E(\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2)}{\mathrm{MSE}^o} = \frac{\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\right)}{\mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\right)}, \tag{11}$$

which can alternatively be expressed as

$$\min_{\gamma, \mathbf{G}} \gamma \tag{12}$$

subject to

$$\max_{l_i \le \delta_i \le u_i} \left\{ \frac{\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\right)}{\mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\right)} \right\} \le \gamma. \tag{13}$$

The constraint (13) is equivalent to

$$\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\right) - \gamma \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\right) \le 0, \quad \forall\ l_i \le \delta_i \le u_i, \tag{14}$$

or,

$$\max_{l_i \le \delta_i \le u_i} \left\{ \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\right) - \gamma \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\right) \right\} \le 0. \tag{15}$$

Thus, the problem of (10) is equivalent to (12) subject to (15).

We now show that the problem of (12) and (15) can be formulated as a convex *semidefinite programming (SDP) problem* [15, 16, 17], which is the problem of minimizing a linear functional subject to linear matrix inequalities (LMIs), *i.e.,* matrix inequalities in which the matrices depend *linearly* on the unknowns. (Note that even though the matrices are linear in the unknowns, the inequalities are nonlinear since a positive semidefinite constraint on a matrix reduces to nonlinear constraints on the matrix elements.) The main advantage of the SDP formulation is that it readily lends itself to efficient computational methods. Specifically, by exploiting the many well known algorithms for solving SDPs [16, 15], *e.g.,* interior point methods[2] [17, 18], which are guaranteed to converge to the global optimum, the optimal estimator can be computed very efficiently in polynomial time. Using principles of duality theory in vector space optimization, the SDP formulation can also be used to derive optimality conditions.

After a description of the general SDP problem in Section 3.1, in Section 3.2, we show that our minimax problem can be formulated as an SDP. In Section 4, we use the SDP formulation to develop more insight into the minimax ratio estimator.

[2]Interior point methods are iterative algorithms that terminate once a pre-specified accuracy has been reached. A worst case analysis of interior point methods shows that the effort required to solve an SDP to a given accuracy grows no faster than a polynomial of the problem size. In practice, the algorithms behave much better than predicted by the worst case analysis, and in fact in many cases the number of iterations is almost constant in the size of the problem.

3.1 Semidefinite Programming

A standard SDP is the problem of minimizing

$$P(\mathbf{x}) = \mathbf{c}^*\mathbf{x} \tag{16}$$

subject to

$$\mathbf{F}(\mathbf{x}) \geq 0, \tag{17}$$

where

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}_0 + \sum_{i=1}^{m} x_i \mathbf{F}_i. \tag{18}$$

Here $\mathbf{x} \in \mathcal{R}^m$ is the vector to be optimized, x_i denotes the ith component of $\mathbf{x}$, $\mathbf{c}$ is a given vector in $\mathcal{R}^m$, and $\mathbf{F}_i$ are given matrices in the space $\mathcal{B}_n$ of $n \times n$ Hermitian matrices[3].

The constraint (17) is an LMI, in which the unknowns x_i appear linearly. Indeed, any constraint of the form $\mathbf{A}(\mathbf{x}) \geq 0$ where the matrix $\mathbf{A}$ depends linearly on $\mathbf{x}$ can be put in the form of (17).

The problem of (16) and (17) is referred to as the *primal problem.* A vector $\mathbf{x}$ is said to be *primal feasible* if $\mathbf{F}(\mathbf{x}) \geq 0$, and is *strictly primal feasible* if $\mathbf{F}(\mathbf{x}) > 0$. If there exists a strictly feasible point, then the primal problem is said to be strictly feasible. We denote the optimal value of $P(\mathbf{x})$ by $\widehat{P}$.

An SDP is a convex optimization problem and can be solved very efficiently. Furthermore, iterative algorithms for solving SDPs are guaranteed to converge to the global minimum. The SDP formulation can also be used to derive necessary and sufficient conditions for optimality by exploiting principles of duality theory. The essential idea is to formulate a *dual problem* of the form $\max_{\mathbf{Z}} D(\mathbf{Z})$ for some linear functional D whose maximal value $\widehat{D}$ serves as a certificate for $\widehat{P}$. That is, for all feasible values of $\mathbf{Z} \in \mathcal{B}_n$, *i.e.,* values of $\mathbf{Z} \in \mathcal{B}_n$ that satisfy a certain set of constraints, and for all feasible values of $\mathbf{x}$, $D(\mathbf{Z}) \leq P(\mathbf{x})$, so that the dual problem provides a lower bound on the optimal value of the original (primal) problem. If in addition we can establish that $\widehat{P} = \widehat{D}$, then this equality can be used to develop conditions of optimality on $\mathbf{x}$.

The dual problem associated with the SDP of (16) and (17) is the problem of maximizing

$$D(\mathbf{Z}) = -\mathrm{Tr}(\mathbf{F}_0\mathbf{Z}) \tag{19}$$

[3]Although typically in the literature the matrices $\mathbf{F}_i$ are restricted to be real and symmetric, the SDP formulation can be easily extended to include Hermitian matrices $\mathbf{F}_i$; see *e.g.,* [29]. In addition, many of the standard software packages for efficiently solving SDPs, for example the Self-Dual-Minimization (SeDuMi) package [30, 31], allow for Hermitian matrices.

subject to

$$\begin{aligned} \mathrm{Tr}(\mathbf{F}_i\mathbf{Z}) &= c_i, \quad 1 \le i \le m; & (20) \\ \mathbf{Z} &\ge 0, & (21) \end{aligned}$$

where $\mathbf{Z} \in \mathcal{B}_n$. A matrix $\mathbf{Z} \in \mathcal{B}_n$ is said to be *dual feasible* if it satisfies (20) and (21) and is *strictly dual feasible* if it satisfies (20) and $\mathbf{Z} > 0$. If there exists a strictly feasible point, then the dual problem is said to be strictly feasible.

For any feasible $\mathbf{x}$ and $\mathbf{Z}$ we have that

$$P(\mathbf{x}) - D(\mathbf{Z}) = c^*\mathbf{x} + \mathrm{Tr}(\mathbf{F}_0\mathbf{Z}) = \sum_{i=1}^{m} x_i \mathrm{Tr}(\mathbf{F}_i\mathbf{Z}) + \mathrm{Tr}(\mathbf{F}_0\mathbf{Z}) = \mathrm{Tr}(\mathbf{F}(\mathbf{x})\mathbf{Z}) \ge 0, \tag{22}$$

so that as required, $D(\mathbf{Z}) \le P(\mathbf{x})$. Furthermore, it can be shown that if either the primal or the dual problem are strictly feasible, then $\widehat{P} = \widehat{D}$ and $\mathbf{x}$ is an optimal primal point if and only if $\mathbf{x}$ is primal feasible, and there exists a dual feasible $\mathbf{Z} \in \mathcal{B}_n$ such that

$$\mathbf{Z}\mathbf{F}(\mathbf{x}) = 0. \tag{23}$$

Equation (23) together with (20), (21) and (17) constitute a set of necessary and sufficient conditions for $\mathbf{x}$ to be an optimal solution to the problem of (16) and (17), when either the primal or the dual are strictly feasible.

3.2 Semidefinite Programming Formulation of the Estimation Problem

In Theorem 1 below, we show that the problem of (12) and (15) can be formulated as an SDP.

Theorem 1. *Let* $\mathbf{x}$ *denote the unknown parameters in the model* $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, *where* $\mathbf{H}$ *is a known* $n \times m$ *matrix with rank* m, $\mathbf{x}$ *is a zero-mean random vector uncorrelated with* $\mathbf{w}$ *with covariance* $\mathbf{C}_x$ *and* $\mathbf{w}$ *is a zero-mean random vector with covariance* $\mathbf{C}_w$. *Let* $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Lambda\mathbf{V}^*$ *where* $\mathbf{V}$ *is a unitary matrix and* Λ *is an* $m \times m$ *diagonal matrix with diagonal elements* $\lambda_i > 0$ *and let* $\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*$ *where* Δ *is an* $m \times m$ *diagonal matrix with diagonal elements* $0 \le l_i \le \delta_i \le u_i$. *Then the solution to the problem*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{l_i \le \delta_i \le u_i} \left\{ \frac{E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)}{\min_{\hat{\mathbf{x}}=\mathbf{G}(\mathbf{x})\mathbf{y}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)} \right\}$$

has the form

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\Lambda^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y},$$

where $\mathbf{D}$ *is an* $m \times m$ *diagonal matrix with diagonal elements* d_i *that are the solution to the semidefinite programming problem*

$$\min_{t_i, d_i, \alpha, \gamma} \gamma, \tag{24}$$

subject to

$$\begin{bmatrix} \alpha_i & \frac{t_i\lambda_i^2}{2} + \frac{\lambda_i(\gamma-1)}{2} & \lambda_i(d_i - 1) \\ \frac{t_i\lambda_i^2}{2} + \frac{\lambda_i(\gamma-1)}{2} & t_i\lambda_i(\lambda_i\zeta_i + 1) + \lambda_i\zeta_i(\gamma-1) - \alpha_i\epsilon_i^2 & d_i(1+\lambda_i\zeta_i) - \lambda_i\zeta_i \\ \lambda_i(d_i-1) & d_i(1+\lambda_i\zeta_i) - \lambda_i\zeta_i & 1 \end{bmatrix} \geq 0, \quad 1 \leq i \leq m;$$

$$\sum_{i=1}^m t_i \leq 0, \tag{25}$$

with $\zeta_i = (u_i + l_i)/2$ *and* $\epsilon_i = (u_i - l_i)/2$.

Proof. We prove Theorem 1 in three stages. First, we show that the optimal $\mathbf{G}$ has the form

$$\mathbf{G} = \mathbf{V}\mathbf{D}\Lambda^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}, \tag{26}$$

for some $m \times m$ matrix $\mathbf{D}$. We then show that $\mathbf{D}$ must be a diagonal matrix. Finally, we develop an expression for the diagonal elements of $\mathbf{D}$. The first two parts of the proof are similar to the proof of Theorem 3 in [28].

We begin by showing that the optimal $\mathbf{G}$ has the form given by (26). To this end, note that the ratio $\mathcal{R}(\mathbf{C}_x, \mathbf{G})$ of (11) depends on $\mathbf{G}$ only through $\mathbf{G}\mathbf{H}$ and $\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$. Now, for any choice of $\mathbf{G}$,

$$\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) = \mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*) + \mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}(\mathbf{I}-\mathbf{P})\mathbf{C}_w^{1/2}\mathbf{G}^*) \geq \mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*) \tag{27}$$

where

$$\mathbf{P} = \mathbf{C}_w^{-1/2}\mathbf{H}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1/2} \tag{28}$$

is the orthogonal projection onto the range space of $\mathbf{C}_w^{-1/2}\mathbf{H}$. In addition, $\mathbf{G}\mathbf{H} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2}\mathbf{H}$ since $\mathbf{P}\mathbf{C}_w^{-1/2}\mathbf{H} = \mathbf{C}_w^{-1/2}\mathbf{H}$. It follows then that any choice of $\mathbf{G}$ can be replaced by $\widetilde{\mathbf{G}} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2}$, with $\mathrm{Tr}(\widetilde{\mathbf{G}}\mathbf{C}_w\widetilde{\mathbf{G}}^*) \leq \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$ and $\widetilde{\mathbf{G}}\mathbf{H} = \mathbf{G}\mathbf{H}$, implying that $\widetilde{\mathbf{G}}$ is always at least as good as $\mathbf{G}$ in the sense of reducing the ratio regret. Since $\widetilde{\mathbf{G}} = \widetilde{\mathbf{G}}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2}$, it follows then that when seeking the optimal matrix

$\mathbf{G}$, it is sufficient to confine attention to matrices that satisfy

$$\mathbf{G}\mathbf{C}_w^{1/2} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}. \tag{29}$$

Substituting (28) into (29), we have

$$\mathbf{G} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2} = \mathbf{G}\mathbf{H}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1} = \mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}, \tag{30}$$

for some $m \times m$ matrix $\mathbf{B}$. Denoting $\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^*$ and using the fact that $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Lambda\mathbf{V}^*$, (30) reduces to (26).

We now show that $\mathbf{D}$ must be a diagonal matrix. Since $\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*$, $\mathbf{H}\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Lambda\mathbf{V}^*$ and $\mathbf{G}$ is given by (30), we have that

$$\begin{aligned}
&\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})\right) - \gamma\mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x+\mathbf{I})^{-1}\right) = \\
&= \mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Lambda^{-1}) + \mathrm{Tr}\left(\Delta(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\right) - \gamma\mathrm{Tr}\left(\Delta(\Lambda\Delta+\mathbf{I})^{-1}\right).
\end{aligned} \tag{31}$$

Therefore the problem of (12) and (15) reduces to finding $\mathbf{D}$ that minimizes γ subject to

$$\mathcal{G}(\mathbf{D}) = \max_{l_i \le \delta_i \le u_i} \mathcal{L}(\mathbf{D}, \Delta) \le 0, \tag{32}$$

where

$$\mathcal{L}(\mathbf{D}, \Delta) = \mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Lambda^{-1}) + \mathrm{Tr}\left(\Delta(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\right) - \gamma\mathrm{Tr}\left(\Delta(\Lambda\Delta+\mathbf{I})^{-1}\right). \tag{33}$$

Clearly, $\mathcal{L}(\mathbf{D})$ is strictly convex in $\mathbf{D}$. Therefore, for any $0 < \alpha < 1$,

$$\begin{aligned}
\mathcal{G}(\alpha\mathbf{D}_1 + (1-\alpha)\mathbf{D}_2) &= \max_{l_i \le \delta_i \le u_i} \mathcal{L}(\alpha\mathbf{D}_1 + (1-\alpha)\mathbf{D}_2, \Delta) \\
&< \max_{l_i \le \delta_i \le u_i} \{\alpha\mathcal{L}(\mathbf{D}_1, \Delta) + (1-\alpha)\mathcal{L}(\mathbf{D}_2, \Delta)\} \\
&\le \alpha \max_{l_i \le \delta_i \le u_i} \mathcal{L}(\mathbf{D}_1, \Delta) + (1-\alpha) \max_{l_i \le \delta_i \le u_i} \mathcal{L}(\mathbf{D}_2, \Delta) \\
&= \alpha\mathcal{G}(\mathbf{D}_1) + (1-\alpha)\mathcal{G}(\mathbf{D}_2),
\end{aligned} \tag{34}$$

so that $\mathcal{G}(\mathbf{D})$ is also strictly convex in $\mathbf{D}$, and consequently our problem has a unique global minimum. Let

$\mathbf{J}$ be any diagonal matrix with diagonal elements equal to ± 1. Then

$$\begin{aligned}
\mathcal{G}(\mathbf{JDJ}) &= \max_{l_i \le \delta_i \le u_i} \left\{ \mathrm{Tr}(\mathbf{JD}^*\mathbf{DJ}\Lambda^{-1}) + \mathrm{Tr}\left(\Delta(\mathbf{I} - \mathbf{JDJ})^*(\mathbf{I} - \mathbf{JDJ})\right) - \gamma \mathrm{Tr}\left(\Delta(\Lambda\Delta + \mathbf{I})^{-1}\right) \right\} \\
&= \max_{l_i \le \delta_i \le u_i} \left\{ \mathrm{Tr}(\mathbf{D}^*\mathbf{DJ}\Lambda^{-1}\mathbf{J}) + \mathrm{Tr}\left(\mathbf{J}\Delta\mathbf{J}(\mathbf{I} - \mathbf{D})^*(\mathbf{I} - \mathbf{D})\right) - \gamma \mathrm{Tr}\left(\Delta(\Lambda\Delta + \mathbf{I})^{-1}\right) \right\} \\
&= \max_{l_i \le \delta_i \le u_i} \left\{ \mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Lambda^{-1}) + \mathrm{Tr}\left(\Delta(\mathbf{I} - \mathbf{D})^*(\mathbf{I} - \mathbf{D})\right) - \gamma \mathrm{Tr}\left(\Delta(\Lambda\Delta + \mathbf{I})^{-1}\right) \right\} \\
&= \mathcal{G}(\mathbf{D}),
\end{aligned} \tag{35}$$

where we used the fact that $\mathbf{J}^2 = \mathbf{I}$ and for any diagonal matrix $\mathbf{M}$, $\mathbf{JMJ} = \mathbf{M}$. Since $\mathcal{G}(\mathbf{D})$ has a unique minimizer, we conclude that the matrix $\mathbf{D}$ that minimizes $\mathcal{G}(\mathbf{D})$ satisfies $\mathbf{D} = \mathbf{JDJ}$ for *any* diagonal matrix $\mathbf{J}$ with diagonal elements equal to ± 1, which in turn implies that $\mathbf{D}$ must be a diagonal matrix.

Denote by d_i, λ_i and δ_i the diagonal elements of $\mathbf{D}, \Lambda$ and Δ, respectively. Then we can express $\mathcal{G}(\mathbf{D})$ as

$$\begin{aligned}
\mathcal{G}(\mathbf{D}) &= \max_{l_i \le \delta_i \le u_i} \left\{ \sum_{i=1}^m \left(\frac{d_i^2}{\lambda_i} + \delta_i(1-d_i)^2 - \gamma \frac{\delta_i}{\lambda_i\delta_i + 1} \right) \right\} \\
&= \max_{l_i \le \delta_i \le u_i} \left\{ \sum_{i=1}^m \left(\frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i + 1)} \right) \right\} \\
&= \sum_{i=1}^m \max_{l_i \le \delta_i \le u_i} \left\{ \frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i + 1)} \right\}.
\end{aligned} \tag{36}$$

Our problem can now be formulated as

$$\min_{\gamma, t_i, d_i} \gamma \tag{37}$$

subject to

$$\max_{l_i \le \delta_i \le u_i} \left\{ \tfrac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i + 1)} \right\} \le t_i, \quad 1 \le i \le m; \tag{38}$$

$$\textstyle\sum_{i=1}^m t_i \le 0. \tag{39}$$

Expressing (38) as

$$\frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i + 1)} \le t_i, \quad \forall \delta_i : l_i \le \delta_i \le u_i, \quad 1 \le i \le m, \tag{40}$$

we develop a solution to our problem by first considering each of the constraints (40), where for brevity, we omit the index i.

Let $\delta = \zeta + e$ where $\zeta = (u + l)/2$. Then the condition $l \le \delta \le u$ is equivalent to the condition $e^2 \le \epsilon^2$

where $\epsilon = (u-l)/2$, so that (40) can be written as

$$(\lambda(d-1)(\zeta+e)+d)^2 + (1-\gamma)\lambda\delta \leq t\lambda(\lambda(\zeta+e)+1), \quad \forall e : e^2 \leq \epsilon^2, \tag{41}$$

which in turn is equivalent to the following implication:

$$P(e) \stackrel{\triangle}{=} \epsilon^2 - e^2 \geq 0 \Rightarrow Q(e) \geq 0, \tag{42}$$

where

$$\begin{aligned} Q(e) &= t\lambda(\lambda(\zeta+e)+1) - (\lambda(d-1)(\zeta+e)+d)^2 - (1-\gamma)\lambda\delta \\ &= -e^2\lambda^2(d-1)^2 + 2e\left(\frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} + \lambda(1-d)\left(d(\lambda\zeta+1)-\lambda\zeta\right)\right) + t\lambda(\lambda\zeta+1) - \\ &\quad -(d(\lambda\zeta+1)-\lambda\zeta)^2 + \lambda\zeta(\gamma-1). \end{aligned} \tag{43}$$

We now rely on the following lemma [32, p. 23]:

Lemma 1. *[$\mathcal{S}$-procedure] Let $P(\mathbf{z}) = \mathbf{z}^*\mathbf{A}\mathbf{z} + 2\mathbf{u}^*\mathbf{z} + v$ and $\mathcal{Q}(\mathbf{z}) = \mathbf{z}^*\mathbf{B}\mathbf{z} + 2\mathbf{x}^*\mathbf{z} + y$ be two quadratic functions of $\mathbf{z}$, where $\mathbf{A}$ and $\mathbf{B}$ are symmetric and there exists a $\mathbf{z}_0$ satisfying $P(\mathbf{z}_0) > 0$. Then the implication*

$$P(\mathbf{z}) \geq 0 \Rightarrow Q(\mathbf{z}) \geq 0$$

holds true if and only if there exists an $\alpha \geq 0$ such that

$$\begin{bmatrix} \mathbf{B} - \alpha\mathbf{A} & \mathbf{x} - \alpha\mathbf{u} \\ \mathbf{x}^* - \alpha\mathbf{u}^* & y - \alpha v \end{bmatrix} \geq 0.$$

Combining (42) with Lemma 1, it follows immediately that the condition (40) is equivalent to the condition

$$\begin{bmatrix} \alpha - \lambda^2(d-1)^2 & \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} + \lambda(1-d)\left(d(\lambda\zeta+1)-\lambda\zeta\right) \\ \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} + \lambda(1-d)\left(d(\lambda\zeta+1)-\lambda\zeta\right) & t\lambda(\lambda\zeta+1) - (d(\lambda\zeta+1)-\lambda\zeta)^2 + \lambda\zeta(\gamma-1) - \alpha\epsilon^2 \end{bmatrix} \geq 0. \tag{44}$$

Note that if (44) is satisfied, then $\alpha - \lambda^2(d-1)^2 \geq 0$, which implies that $\alpha \geq 0$.

We can express (44) as

$$\mathbf{A} - \mathbf{b}\mathbf{b}^* \geq 0, \tag{45}$$

where

$$\mathbf{A} = \begin{bmatrix} \alpha & \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} \\ \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & t\lambda(\lambda\zeta+1) + \lambda\zeta(\gamma-1) - \alpha\epsilon^2 \end{bmatrix}, \tag{46}$$

and

$$\mathbf{b} = \begin{bmatrix} \lambda(d-1) \\ d(1+\lambda\zeta) - \lambda\zeta \end{bmatrix}. \tag{47}$$

To treat the constraint (45), we rely on the following lemma [24, p. 472]:

Lemma 2. *Let*

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{B}^* \\ \mathbf{B} & \mathbf{C} \end{bmatrix}$$

be a Hermitian matrix. Then with $\mathbf{C} > 0$, $\mathbf{M} \geq 0$ *if and only if* $\Delta_{\mathbf{C}} \geq 0$, *where* $\Delta_{\mathbf{C}}$ *is the Schur complement of* $\mathbf{C}$ *in* $\mathbf{M}$ *and is given by*

$$\Delta_{\mathbf{C}} = \mathbf{A} - \mathbf{B}^*\mathbf{C}^{-1}\mathbf{B}.$$

From Lemma 2, it follows that (45) is equivalent to

$$\begin{bmatrix} \alpha & \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & \lambda(d-1) \\ \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & t\lambda(\lambda\zeta+1) + \lambda\zeta(\gamma-1) - \alpha\epsilon^2 & d(1+\lambda\zeta) - \lambda\zeta \\ \lambda(d-1) & d(1+\lambda\zeta) - \lambda\zeta & 1 \end{bmatrix} \geq 0, \tag{48}$$

completing the proof of the theorem. □

4 Alternative Derivation of the Minimax Ratio Estimator

In Theorem 1, we showed that the minimax ratio estimator can be formulated as an SDP. In this section, we develop further insight into the minimax ratio estimator, using the SDP optimality conditions. Specifically, we show that the minimax ratio estimator can be expressed in terms of a single parameter, which is a solution to a nonlinear equation, and be found using a simple line search algorithm.

To this end, we first show that the minimax ratio estimator, which is the solution to the problem

$$(\Gamma): \min_{\gamma,\mathbf{G}} \left\{ \gamma : \max_{l_i \leq \delta_i \leq u_i} \left\{ \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})\right) - \gamma \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\right) \right\} \leq 0 \right\}, \tag{49}$$

can be determined by first solving the problem

$$(\Phi):\ \min_{t,\mathbf{G}}\left\{t:\ \max_{l_i\le\delta_i\le u_i}\left\{\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)+\mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})\right)-\gamma\mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x+\mathbf{I})^{-1}\right)\right\}\le t\right\}, \tag{50}$$

where $\gamma\ge 1$ is fixed. Specifically, let $\hat{t}(\gamma)$ denote the optimal value of t in the problem Φ of (50), and let $\hat{\gamma}$ be the minimal value of $\gamma\ge 1$ such that $\hat{t}(\gamma)=0$ (as we show below in Proposition 2, such a γ always exists). Then, denoting by $\widehat{\mathbf{G}}$ the optimal $\mathbf{G}$ in the problem Φ with $\gamma=\hat{\gamma}$, we now show that $\widehat{\mathbf{G}}$ and $\hat{\gamma}$ are the optimal solutions to the problem Γ of (49): Since $\widehat{\mathbf{G}}$ and $\hat{\gamma}$ are feasible for Φ with $t=0$, they are also feasible for Γ. Now suppose, conversely, that there exists a feasible $\mathbf{G}$ and $\gamma<\hat{\gamma}$ for Γ. It then follows that $\hat{t}(\gamma)\le 0$. But since $\hat{t}(\gamma)$ is decreasing in γ and $\gamma<\hat{\gamma}$, we have that $\hat{t}(\gamma)\ge\hat{t}(\hat{\gamma})=0$, from which we conclude that $\hat{t}(\gamma)=0$, which is a contradiction since $\hat{\gamma}$ is the *minimal* value for which $\hat{t}(\gamma)=0$.

In Proposition 2 below, we show that $\hat{t}(\gamma)$ is continuous, and strictly decreasing in γ if $l_i\neq 0$ for at least one value of i, because in this case $\mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x+\mathbf{I})^{-1}\right)>0$ for all $l_i\le\delta_i\le u_i$. Since $\hat{t}(1)\ge 0$ and $\hat{t}(\gamma)\to-\infty$ for $\gamma\to\infty$ (again, because $\mathrm{Tr}\left(\mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x+\mathbf{I})^{-1}\right)>0$), we conclude that in this case, there is a unique γ such that $\hat{t}(\gamma)=0$. We also show that if $l_i=0, 1\le i\le m$, then $\hat{t}(\gamma)>0$ for $1\le\gamma<\gamma'$, $\hat{t}(\gamma)=0$ for $\gamma\ge\gamma'$, and $\hat{t}(\gamma)$ is strictly decreasing for $1\le\gamma\le\gamma'$, where γ' is the smallest value of γ such that $\hat{t}(\gamma)=0$. Therefore, we can find $\hat{\gamma}$ by a simple line search, as we discuss further in the paragraph following Proposition 2.

Thus, we now consider the problem Φ. Following the proof of Theorem 1, we have immediately that the optimal $\mathbf{G}$ has the form

$$\mathbf{G}=\mathbf{V}\mathbf{D}\Lambda^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}, \tag{51}$$

where $\mathbf{D}$ is a diagonal matrix with diagonal elements d_i that are the solution to

$$\min_{t,d_i} t \tag{52}$$

subject to

$$\max_{l_i\le\delta_i\le u_i}\left\{\sum_{i=1}^m\left(\frac{d_i^2}{\lambda_i}+\delta_i(1-d_i)^2-\gamma\frac{\delta_i}{\lambda_i\delta_i+1}\right)\right\}\le t. \tag{53}$$

Noting that

$$\max_{l_i\le\delta_i\le u_i}\left\{\sum_{i=1}^m\left(\frac{d_i^2}{\lambda_i}+\delta_i(1-d_i)^2-\gamma\frac{\delta_i}{\lambda_i\delta_i+1}\right)\right\}=$$

$$
\begin{aligned}
&= \max_{l_i \le \delta_i \le u_i} \left\{ \sum_{i=1}^{m} \left(\frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i+1)} \right) \right\} \\
&= \sum_{i=1}^{m} \max_{l_i \le \delta_i \le u_i} \left\{ \frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i+1)} \right\}, \qquad (54)
\end{aligned}
$$

our problem can be formulated as

$$
\min_{t_i, d_i} \sum_{i=1}^{m} t_i \qquad (55)
$$

subject to

$$
\max_{l_i \le \delta_i \le u_i} \left\{ \frac{(\lambda_i(d_i-1)\delta_i + d_i)^2 + (1-\gamma)\lambda_i\delta_i}{\lambda_i(\lambda_i\delta_i+1)} \right\} \le t_i, \quad 1 \le i \le m. \qquad (56)
$$

The constraints (56) are equal to the constraints (38), which, in turn, where shown in the proof of Theorem 1 to be equivalent to the LMIs

$$
\begin{bmatrix}
\alpha_i & \frac{t_i\lambda_i^2}{2} + \frac{\lambda_i(\gamma-1)}{2} & \lambda_i(d_i-1) \\
\frac{t_i\lambda_i^2}{2} + \frac{\lambda_i(\gamma-1)}{2} & t_i\lambda_i(\lambda_i\zeta_i+1) + \lambda_i\zeta_i(\gamma-1) - \alpha_i\epsilon_i^2 & d_i(1+\lambda_i\zeta_i) - \lambda_i\zeta_i \\
\lambda_i(d_i-1) & d_i(1+\lambda_i\zeta_i) - \lambda_i\zeta_i & 1
\end{bmatrix} \ge 0, \quad 1 \le i \le m. \qquad (57)
$$

Thus to solve Φ, we need to develop a solution to the problem

$$
\min_{t,d,\alpha} t \qquad (58)
$$

subject to

$$
\begin{bmatrix}
\alpha & \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & \lambda(d-1) \\
\frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & t\lambda(\lambda\zeta+1) + \lambda\zeta(\gamma-1) - \alpha\epsilon^2 & d(1+\lambda\zeta) - \lambda\zeta \\
\lambda(d-1) & d(1+\lambda\zeta) - \lambda\zeta & 1
\end{bmatrix} \ge 0, \qquad (59)
$$

which, using Lemma 2, can be equivalently expressed as

$$
\mathbf{A} - \mathbf{b}\mathbf{b}^* \ge 0, \qquad (60)
$$

with

$$
\mathbf{A} = \begin{bmatrix}
\alpha & \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} \\
\frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & t\lambda(1+\lambda\zeta) + \lambda\zeta(\gamma-1) - \alpha\epsilon^2
\end{bmatrix}, \qquad (61)
$$

and

$$
\mathbf{b} = \begin{bmatrix}
\lambda(d-1) \\
d(1+\lambda\zeta) - \lambda\zeta
\end{bmatrix}. \qquad (62)
$$

The problem of (58) and (59) for the special case in which $\gamma = 1$ was considered in [28], in which it was shown that the solution is given by the smallest value of t such that there exists a triplet (t, d, α) satisfying $\mathbf{A} = \mathbf{b}\mathbf{b}^*$. For $\gamma \geq 1$, the smallest value of t and the corresponding values of d and α such that there exists a triplet (t, d, α) satisfying $\mathbf{A} = \mathbf{b}\mathbf{b}^*$, are given by (63) below. As we show in Proposition 1, this solution is also optimal for the problem (58) and (59), as long as γ is smaller than a threshold.

Proposition 1. *The solution of*

$$\min_{t,d,\alpha} t$$

subject to

$$\begin{bmatrix} \alpha & \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & \lambda(d-1) \\ \frac{t\lambda^2}{2} + \frac{\lambda(\gamma-1)}{2} & t\lambda(\lambda\zeta+1) + \lambda\zeta(\gamma-1) - \alpha\epsilon^2 & d(1+\lambda\zeta) - \lambda\zeta \\ \lambda(d-1) & d(1+\lambda\zeta) - \lambda\zeta & 1 \end{bmatrix} \geq 0$$

is given by

$$\begin{aligned} \hat{d} &= 1 - \tfrac{\sqrt{\gamma}}{\sqrt{(1+\lambda\zeta)^2 - \lambda^2\epsilon^2}}; \\ \hat{\alpha} &= \tfrac{\gamma\lambda^2}{(1+\lambda\zeta)^2 - \lambda^2\epsilon^2}; \\ \hat{t} &= \tfrac{1}{\lambda} - \tfrac{2\sqrt{\gamma}}{\lambda\sqrt{(1+\lambda\zeta)^2 - \lambda^2\epsilon^2}} + \tfrac{\gamma\left(\lambda^2(\epsilon^2-\zeta^2)+1\right)}{\lambda((1+\lambda\zeta)^2 - \lambda^2\epsilon^2)}, \end{aligned} \tag{63}$$

for $1 \leq \gamma \leq \gamma^0$, where

$$\gamma^0 = \frac{1+\lambda(\zeta+\epsilon)}{1+\lambda(\zeta-\epsilon)}, \tag{64}$$

and by

$$\begin{aligned} \hat{d} &= \tfrac{\lambda(\zeta-\epsilon)}{1+\lambda(\zeta-\epsilon)}; \\ \hat{\alpha} &= \tfrac{\lambda(\gamma-1)}{2\epsilon(1+\lambda(\zeta-\epsilon))}; \\ \hat{t} &= \tfrac{(\gamma-1)(\epsilon-\zeta)}{1+\lambda(\zeta-\epsilon)}, \end{aligned} \tag{65}$$

for $\gamma \geq \gamma^0$.

Proof. We begin by showing that the values given by (63) are optimal for $1 \leq \gamma \leq \gamma^0$. To this end it is sufficient to show that $\mathbf{F}(\hat{d}, \hat{\alpha}, \hat{t}) \geq 0$, where $\mathbf{F}(d, \alpha, t)$ is the matrix in (59), that there exists a matrix $\mathbf{Z} \geq 0$ such that

$$\mathbf{F}(\hat{d}, \hat{\alpha}, \hat{t})\mathbf{Z} = 0;$$

$$\mathrm{Tr}\,(\mathbf{F}_d\mathbf{Z}) = 0;$$
$$\mathrm{Tr}\,(\mathbf{F}_\alpha\mathbf{Z}) = 0;$$
$$\mathrm{Tr}\,(\mathbf{F}_t\mathbf{Z}) = 1, \tag{66}$$

where

$$\mathbf{F}_d = \begin{bmatrix} 0 & 0 & \lambda \\ 0 & 0 & 1+\lambda\zeta \\ \lambda & 1+\lambda\zeta & 0 \end{bmatrix};$$
$$\mathbf{F}_\alpha = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -\epsilon^2 & 0 \\ 0 & 0 & 0 \end{bmatrix};$$
$$\mathbf{F}_t = \begin{bmatrix} 0 & \frac{\lambda^2}{2} & 0 \\ \frac{\lambda^2}{2} & \lambda(1+\lambda\zeta) & 0 \\ 0 & 0 & 0 \end{bmatrix}, \tag{67}$$

and that the dual problem is strictly feasible, *i.e.*, there exists a matrix $\mathbf{B}$ satisfying

$$\mathbf{B} > 0;$$
$$\mathrm{Tr}\,(\mathbf{F}_d\mathbf{B}) = 0;$$
$$\mathrm{Tr}\,(\mathbf{F}_\alpha\mathbf{B}) = 0;$$
$$\mathrm{Tr}\,(\mathbf{F}_t\mathbf{B}) = 1. \tag{68}$$

Let

$$\mathbf{B} = \begin{bmatrix} b\epsilon^2 & \frac{1}{\lambda^2}(1-b\lambda(\lambda\zeta+1)) & 0 \\ \frac{1}{\lambda^2}(1-b\lambda(\lambda\zeta+1)) & b & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{69}$$

where

$$b = \frac{1}{\lambda(\lambda\zeta+1)+\lambda^2(\epsilon-a)}, \tag{70}$$

and a is chosen such that $\epsilon^2 > (\epsilon - a)^2$. We can immediately verify that $\mathbf{B}$ satisfies (68). Next, we verify that $\mathbf{F}(\hat{d},\hat{\alpha},\hat{t})\geq 0$. Substituting (63) into the matrix in (59), we have that

$$\mathbf{F}(\hat{d},\hat{\alpha},\hat{t})=\begin{bmatrix}\gamma\lambda^2 & \lambda\left(\gamma(1+\lambda\zeta)-\beta\sqrt{\gamma}\right) & -\lambda\sqrt{\gamma}\beta\\ \lambda\left(\gamma(1+\lambda\zeta)-\beta\sqrt{\gamma}\right) & (\beta-\sqrt{\gamma}(1+\lambda\zeta))^2 & \beta^2-\beta\sqrt{\gamma}(1+\lambda\zeta)\\ -\lambda\sqrt{\gamma}\beta & \beta^2-\beta\sqrt{\gamma}(1+\lambda\zeta) & \beta^2\end{bmatrix}. \tag{71}$$

From Lemma 2, it follows that $\mathbf{F}(\hat{d},\hat{\alpha},\hat{t})\geq 0$ if and only if $\mathbf{W}\geq 0$, where

$$\mathbf{W}=\begin{bmatrix}\gamma\lambda^2 & \lambda\left(\gamma(1+\lambda\zeta)-\beta\sqrt{\gamma}\right)\\ \lambda\left(\gamma(1+\lambda\zeta)-\beta\sqrt{\gamma}\right) & (\beta-\sqrt{\gamma}(1+\lambda\zeta))^2\end{bmatrix}-$$
$$-\begin{bmatrix}\lambda^2\gamma & -\lambda\sqrt{\gamma}(\beta-\sqrt{\gamma}(1+\lambda\zeta))\\ -\lambda\sqrt{\gamma}(\beta-\sqrt{\gamma}(1+\lambda\zeta)) & (\beta-\sqrt{\gamma}(1+\lambda\zeta))^2\end{bmatrix}=0. \tag{72}$$

It remains to find a $\mathbf{Z}\geq 0$ satisfying (66). Solving (66) for $\mathbf{Z}$ with t,d,α given by (63) results in

$$\mathbf{Z}=\frac{1}{\lambda\beta^2\sqrt{\gamma}}\begin{bmatrix}\epsilon^2\kappa & \frac{1}{\lambda}\left(\beta^2\sqrt{\gamma}-(1+\lambda\zeta)\kappa\right) & \frac{1}{\lambda}(1+\lambda\zeta)(\kappa-\beta\gamma)\\ \frac{1}{\lambda}\left(\beta^2\sqrt{\gamma}-(1+\lambda\zeta)\kappa\right) & \kappa & \beta\gamma-\kappa\\ \frac{1}{\lambda}(1+\lambda\zeta)(\kappa-\beta\gamma) & \beta\gamma-\kappa & \kappa-\beta\gamma\end{bmatrix}, \tag{73}$$

where, for brevity, we defined

$$\beta\stackrel{\triangle}{=}\sqrt{(1+\lambda\zeta)^2-\lambda^2\epsilon^2}, \tag{74}$$

and

$$\kappa\stackrel{\triangle}{=}2\sqrt{\gamma}(1+\lambda\zeta)-\beta. \tag{75}$$

Since $\mathbf{Z}$ satisfies (66), the values given by (63) are optimal if, in addition, $\mathbf{Z}\geq 0$, which implies that

$$\kappa\geq\beta\gamma, \tag{76}$$

or, equivalently,

$$\beta\gamma-2\sqrt{\gamma}(1+\lambda\zeta)+\beta\leq 0. \tag{77}$$

It is straightforward to show that (77) is satisfied for

$$\frac{(1+\lambda(\zeta-\epsilon))^2}{\beta^2}\leq\gamma\leq\frac{(1+\lambda(\zeta+\epsilon))^2}{\beta^2}. \tag{78}$$

Since

$$\frac{(1+\lambda(\zeta-\epsilon))^2}{\beta^2} = \frac{(1+\lambda(\zeta-\epsilon))^2}{(1+\lambda(\zeta+\epsilon))\,(1+\lambda(\zeta-\epsilon))} = \frac{1+\lambda(\zeta-\epsilon)}{1+\lambda(\zeta+\epsilon)} \leq 1, \tag{79}$$

(76) is satisfied for $1 \leq \gamma \leq \gamma^0$, where γ^0 is given by (64).

We now show that for $1 \leq \gamma \leq \gamma^0$, $\mathbf{Z} \geq 0$. First suppose that $\kappa = \beta\gamma$, or equivalently, $\gamma = \gamma^0$. In this case, $\mathbf{Z} \geq 0$ if and only if

$$\begin{bmatrix} \epsilon^2\kappa & \frac{1}{\lambda}\left(\beta^2\sqrt{\gamma}-(1+\lambda\zeta)\kappa\right) \\ \frac{1}{\lambda}\left(\beta^2\sqrt{\gamma}-(1+\lambda\zeta)\kappa\right) & \kappa \end{bmatrix} \geq 0, \tag{80}$$

or,

$$\lambda^2\epsilon^2\kappa^2 - \left(\beta^2\sqrt{\gamma}-(1+\lambda\zeta)\kappa\right)^2 \geq 0. \tag{81}$$

Using the fact that $\kappa = \beta\gamma$, (81) can be expressed as

$$-\beta - \beta\gamma + 2\sqrt{\gamma}(1+\lambda\zeta) \geq 0, \tag{82}$$

which is equivalent to (77), and is therefore satisfied for $\gamma = \gamma^0$.

Next, suppose that $\kappa > \beta\gamma$. From Lemma 2, we then have that $\mathbf{Z} \geq 0$ if and only if

$$\begin{aligned} &\begin{bmatrix} \epsilon^2\kappa & \frac{1}{\lambda}\left(\beta^2\sqrt{\gamma}-(1+\lambda\zeta)\kappa\right) \\ \frac{1}{\lambda}\left(\beta^2\sqrt{\gamma}-(1+\lambda\zeta)\kappa\right) & \kappa \end{bmatrix} + (\beta\gamma-\kappa)\begin{bmatrix} \frac{1}{\lambda^2}(1+\lambda\zeta)^2 & -\frac{1}{\lambda}(1+\lambda\zeta) \\ -\frac{1}{\lambda}(1+\lambda\zeta) & 1 \end{bmatrix} = \\ &= \begin{bmatrix} \epsilon^2\kappa + \frac{1}{\lambda^2}(\beta\gamma-\kappa)(1+\lambda\zeta)^2 & \frac{1}{\lambda}\left(\beta^2\sqrt{\gamma}-\beta\gamma(1+\lambda\zeta)\right) \\ \frac{1}{\lambda}\left(\beta^2\sqrt{\gamma}-\beta\gamma(1+\lambda\zeta)\right) & \gamma\beta \end{bmatrix} \\ &\geq 0, \end{aligned} \tag{83}$$

which is equivalent to the conditions

$$\epsilon^2\kappa + \tfrac{1}{\lambda^2}(\beta\gamma-\kappa)(1+\lambda\zeta)^2 \geq 0; \tag{84}$$

$$\gamma\beta\left(\epsilon^2\kappa + \tfrac{1}{\lambda^2}(\beta\gamma-\kappa)(1+\lambda\zeta)^2\right) - \tfrac{1}{\lambda^2}\left(\beta^2\sqrt{\gamma}-\beta\gamma(1+\lambda\zeta)\right)^2 \geq 0. \tag{85}$$

Now,

$$\begin{aligned} \epsilon^2\kappa + \frac{1}{\lambda^2}(\beta\gamma-\kappa)(1+\lambda\zeta)^2 &= \frac{\kappa}{\lambda^2}\left(\lambda^2\epsilon^2-(1+\lambda\zeta)^2\right) + \frac{\beta\gamma}{\lambda^2}(1+\lambda\zeta)^2 \\ &= \frac{\beta}{\lambda^2}(\gamma(1+\lambda\zeta)^2-\beta\kappa) \end{aligned}$$

$$
\begin{aligned}
&= \frac{\beta}{\lambda^2}(\sqrt{\gamma}(1+\lambda\zeta)-\beta)^2 \\
&\geq 0, \qquad (86)
\end{aligned}
$$

so that (84) is satisfied. Finally,

$$
\begin{aligned}
&\gamma\beta\left(\epsilon^2\kappa + \frac{1}{\lambda^2}(\beta\gamma-\kappa)(1+\lambda\zeta)^2\right) - \frac{1}{\lambda^2}\left(\beta^2\sqrt{\gamma}-\beta\gamma(1+\lambda\zeta)\right)^2 = \\
&= \frac{\gamma\beta^2}{\lambda^2}(\sqrt{\gamma}(1+\lambda\zeta)-\beta)^2 - \frac{1}{\lambda^2}\left(\beta^2\sqrt{\gamma}-\beta\gamma(1+\lambda\zeta)\right)^2 \\
&= 0, \qquad (87)
\end{aligned}
$$

so that (85) is also satisfied, and $\mathbf{Z} \geq 0$.

We now show that for $\gamma \geq \gamma^0$, the values given by (65) are optimal. For these values, we have that

$$
\mathbf{F}(\hat{d},\hat{\alpha},\hat{t}) = \frac{1}{1+\lambda(\zeta-\epsilon)}\begin{bmatrix} \frac{\lambda(\gamma-1)}{2\epsilon} & \frac{\lambda(\gamma-1)}{2} & -\lambda \\ \frac{\lambda(\gamma-1)}{2} & \frac{\lambda\epsilon(\gamma-1)}{2} & -\lambda\epsilon \\ -\lambda & -\lambda\epsilon & 1+\lambda(\zeta-\epsilon) \end{bmatrix}. \qquad (88)
$$

From Lemma 2 it then follows that $\mathbf{F}(\hat{d},\hat{\alpha},\hat{t}) \geq 0$ if and only if

$$
\begin{bmatrix} \frac{\lambda(\gamma-1)}{2\epsilon} & \frac{\lambda(\gamma-1)}{2} \\ \frac{\lambda(\gamma-1)}{2} & \frac{\lambda\epsilon(\gamma-1)}{2} \end{bmatrix} - \frac{\lambda^2}{1+\lambda(\zeta-\epsilon)}\begin{bmatrix} 1 & \epsilon \\ \epsilon & \epsilon^2 \end{bmatrix} = \frac{\lambda(\gamma-1)(1+\lambda(\zeta-\epsilon))-2\epsilon\lambda^2}{2\epsilon(1+\lambda(\zeta-\epsilon))}\begin{bmatrix} 1 & \epsilon \\ \epsilon & \epsilon^2 \end{bmatrix} \geq 0. \qquad (89)
$$

Clearly, (89) is satisfied if

$$
\lambda(\gamma-1)(1+\lambda(\zeta-\epsilon)) - 2\epsilon\lambda^2 \geq 0, \qquad (90)
$$

or,

$$
\gamma \geq \frac{1+\lambda(\zeta+\epsilon)}{1+\lambda(\zeta-\epsilon)} = \gamma^0. \qquad (91)
$$

We now show that there exists a $\mathbf{Z} \geq 0$ satisfying (66). Solving (66) for $\mathbf{Z}$ with t, d, α given by (65) results in

$$
\mathbf{Z} = \frac{1}{\lambda^2(\zeta-\epsilon)+\lambda}\begin{bmatrix} \epsilon^2 & -\epsilon & 0 \\ -\epsilon & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad (92)
$$

and $\mathbf{Z} \geq 0$. Thus, the values given by (65) are optimal for $\gamma \geq \gamma^0$, completing the proof of the proposition. □

From Proposition 1 it follows that the solution to the problem Φ of (50) is given by

$$\mathbf{G} = \mathbf{V}\widehat{\mathbf{D}}\Lambda^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}, \tag{93}$$

where $\widehat{\mathbf{D}}$ is the diagonal matrix with diagonal elements

$$\hat{d}_i = \begin{cases} 1 - \frac{\sqrt{\gamma}}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}, & 1 \leq \gamma \leq \gamma_i^0; \\ \frac{\lambda_i(\zeta_i - \epsilon_i)}{1+\lambda_i(\zeta_i-\epsilon_i)}, & \gamma \geq \gamma_i^0, \end{cases} \tag{94}$$

with

$$\gamma_i^0 = \frac{1+\lambda_i(\zeta_i+\epsilon_i)}{1+\lambda_i(\zeta_i-\epsilon_i)}, \tag{95}$$

and $\hat{t}(\gamma) = \sum_{i=1}^m t_i(\gamma)$, where

$$t_i(\gamma) = \begin{cases} \frac{1}{\lambda_i} - \frac{2\sqrt{\gamma}}{\lambda_i\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}} + \frac{\gamma\left(\lambda_i^2(\epsilon_i^2-\zeta_i^2)+1\right)}{\lambda_i\left((1+\lambda_i\zeta_i)^2-\lambda_i^2\epsilon_i^2\right)}, & 1 \leq \gamma \leq \gamma_i^0; \\ \frac{(\gamma-1)(\epsilon_i-\zeta_i)}{1+\lambda_i(\zeta_i-\epsilon_i)}, & \gamma \geq \gamma_i^0. \end{cases} \tag{96}$$

To relate the solutions of problems Φ and Γ of (50) and (49) respectively, we now establish the required properties of $\hat{t}(\gamma)$, outlined at the beginning of the section.

Proposition 2. *Let $\hat{t}(\gamma) = \sum_{i=1}^m t_i(\gamma)$, where $t_i(\gamma)$ is given by (96). Then*

1. *$\hat{t}(\gamma)$ is continuous;*
2. *If $\epsilon_i < \zeta_i$ for some i, then $\hat{t}(\gamma)$ is strictly decreasing in $\gamma \geq 1$, and there is a unique γ such that $\hat{t}(\gamma) = 0$;*
3. *If $\epsilon_i = \zeta_i, 1 \leq i \leq m$, then $\hat{t}(\gamma) > 0$ and is strictly decreasing in $1 \leq \gamma \leq \max_i \gamma_i^0$, and $\hat{t}(\gamma) = 0$ for $\gamma \geq \max_i \gamma_i^0$, where γ_i^0 is given by (95).*

Proof. For $\gamma < \gamma_i^0$, $t_i(\gamma)$ is quadratic in γ, and for $\gamma > \gamma_i^0$, $t_i(\gamma)$ is linear in γ. Therefore, in both of these intervals, $t_i(\gamma)$ is continuous. We can also immediately verify that $t_i(\gamma)$ is continuous at $\gamma = \gamma_i^0$.

We now consider the monotonicity properties of $t_i(\gamma)$. Suppose first that $\lambda_i^2(\zeta_i^2 - \epsilon_i^2) = 1$. In this case, $t_i(\gamma)$ is linear in γ for $\gamma \leq \gamma_i^0$, and is strictly decreasing. Next suppose that $\lambda_i^2(\zeta_i^2 - \epsilon_i^2) \neq 1$. In this case, for $\gamma \leq \gamma_i^0$, $t_i(\gamma)$ is a quadratic function in $\sqrt{\gamma}$, which we denote by $Q(\gamma)$, with an extremum point at

$$\sqrt{\gamma} = \frac{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}{1-\lambda_i^2(\zeta_i^2-\epsilon_i^2)} \stackrel{\triangle}{=} \gamma_i^t. \tag{97}$$

If $\lambda_i^2(\zeta_i^2-\epsilon_i^2)<1$, then $Q(\gamma)$ has a minimum at γ_i^t. Consequently, $Q(\gamma)$ is decreasing for any $\gamma \leq (\gamma_i^t)^2$. Since, as we now show, $(\gamma_i^t)^2 \geq \gamma_i^0$, $t_i(\gamma)$ is decreasing for $\gamma \leq \gamma_i^0$. To prove that $(\gamma_i^t)^2 \geq \gamma_i^0$, we must have that

$$\frac{(1+\lambda_i\zeta_i)^2-\lambda_i^2\epsilon_i^2}{\left(1-\lambda_i^2(\zeta_i^2-\epsilon_i^2)\right)^2} \geq \frac{1+\lambda_i(\zeta_i+\epsilon_i)}{1+\lambda_i(\zeta_i-\epsilon_i)}, \tag{98}$$

or, equivalently,

$$1+\lambda(\zeta_i-\epsilon_i) \geq 1-\lambda_i^2(\zeta_i^2-\epsilon_i^2), \tag{99}$$

which is clearly satisfied. If, on the other hand, $\lambda_i^2(\zeta_i^2-\epsilon_i^2)>1$, then $Q(\gamma)$ has a maximum at $\gamma_i^t<0$, and $t_i(\gamma)$ is strictly decreasing for $1 \leq \gamma \leq \gamma_i^0$.

For $\gamma \geq \gamma_i^0$, $t_i(\gamma)$ is strictly decreasing as long as $\epsilon_i<\zeta_i$. If $\epsilon_i=\zeta_i$, then $t_i(\gamma)=0$ for all $\gamma \geq \gamma_0^i$.

We conclude that $\hat{t}(\gamma)$ is strictly decreasing in γ if for some i, $\epsilon_i<\zeta_i$. Since $\hat{t}(1)>0$ (unless $\epsilon_i=0, 1 \leq i \leq m$ in which case there is no uncertainty) and $\hat{t}(\gamma) \leq 0$ for $\gamma \rightarrow -\infty$, it follows that in this case there is a unique value γ such that $\hat{t}(\gamma)=0$.

If $\epsilon_i=\zeta_i, 1 \leq i \leq m$, then $\hat{t}(\gamma)$ is strictly decreasing for $\gamma \leq \max_i \gamma_i^0$, and $\hat{t}(\gamma)=0$ for $\gamma \geq \max_i \gamma_i^0$, which implies that $\hat{t}(\gamma)>0$ for $\gamma<\max_i \gamma_i^0$. $\square$

From Proposition 2 it follows that the solution to the original problem Γ is also given by (93) and (94), with γ chosen as the minimal value such that $\hat{t}(\gamma)=\sum_{i=1}^m t_i(\gamma)=0$. Since $\hat{t}(\gamma)$ is continuous and strictly decreasing in γ such that $\hat{t}(\gamma) \neq 0$, we can find the minimal value of γ satisfying $\hat{t}(\gamma)=0$ by a simple line search. Specifically, we may start by choosing $\gamma=1$. For each choice of γ we compute $\sum_{i=1}^m t_i(\gamma)$. If $\sum_{i=1}^m t_i(\gamma)>0$, then we increase γ, and if $\sum_{i=1}^m t_i(\gamma)<0$ we decrease γ, continuing until $\sum_{i=1}^m t_i(\gamma)=0$. If $\epsilon_i<\zeta_i$ for some i, then this is the optimal value of γ. If $\epsilon_i=\zeta_i$ for all i, then we continue decreasing the value of γ, looking for the smallest value such that $\sum_{i=1}^m t_i(\gamma)=0$. Due to the continuity and monotonicity properties of $\hat{t}(\gamma)$, the algorithm is guaranteed to converge.

We summarize our results in the following theorem.

Theorem 2. *Let* $\mathbf{x}$ *denote the unknown parameters in the model* $\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{w}$*, where* $\mathbf{H}$ *is a known* $n \times m$ *matrix with rank* m*,* $\mathbf{x}$ *is a zero-mean random vector with covariance* $\mathbf{C}_x$ *uncorrelated with* $\mathbf{w}$ *and* $\mathbf{w}$ *is a zero-mean random vector with covariance* $\mathbf{C}_w$*. Let* $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}=\mathbf{V}\Lambda\mathbf{V}^*$ *where* $\mathbf{V}$ *is a unitary matrix and* Λ *is an* $m \times m$ *diagonal matrix with diagonal elements* $\lambda_i>0$ *and let* $\mathbf{C}_x=\mathbf{V}\Delta\mathbf{V}^*$ *where* Δ *is an* $m \times m$ *diagonal matrix with diagonal elements* $0 \leq l_i \leq \delta_i \leq u_i$*. Then the solution to the problem*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{l_i \leq \delta_i \leq u_i} \left\{ \frac{E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)}{\min_{\hat{\mathbf{x}}=\mathbf{G}(\mathbf{x})\mathbf{y}} E(\|\hat{\mathbf{x}}-\mathbf{x}\|^2)} \right\}$$

is given by

$$\hat{\mathbf{x}} = \mathbf{V}\mathbf{D}\Lambda^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y},$$

where $\mathbf{D}$ *is an* $m \times m$ *diagonal matrix with diagonal elements*

$$\hat{d}_i = \begin{cases} 1 - \frac{\sqrt{\gamma}}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}, & 1 \le \gamma \le \gamma_i^0; \\ \frac{\lambda_i(\zeta_i - \epsilon_i)}{1+\lambda_i(\zeta_i-\epsilon_i)}, & \gamma \ge \gamma_i^0, \end{cases} \tag{100}$$

with

$$\gamma_i^0 = \frac{1+\lambda_i(\zeta_i+\epsilon_i)}{1+\lambda_i(\zeta_i-\epsilon_i)}, \tag{101}$$

and γ *is the smallest value such that* $\hat{t}(\gamma) = \sum_{i=1}^m t_i(\gamma) = 0$, *where*

$$t_i(\gamma) = \begin{cases} \frac{1}{\lambda_i} - \frac{2\sqrt{\gamma}}{\lambda_i\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}} + \frac{\gamma\left(\lambda_i^2(\epsilon_i^2-\zeta_i^2)+1\right)}{\lambda_i\left((1+\lambda_i\zeta_i)^2-\lambda_i^2\epsilon_i^2\right)}, & 1 \le \gamma \le \gamma_i^0; \\ \frac{(\gamma-1)(\epsilon_i-\zeta_i)}{1+\lambda_i(\zeta_i-\epsilon_i)}, & \gamma \ge \gamma_i^0. \end{cases} \tag{102}$$

Note, that the minimax regret estimator has the same form as the minimax ratio estimator, where d_i is given by (100) with $\gamma = 1$.

As we now show, we can interpret the estimator of Theorem 2 as an MMSE estimator matched to a covariance matrix

$$\mathbf{C}_x = \mathbf{V}\mathbf{X}\mathbf{V}^*, \tag{103}$$

where $\mathbf{X}$ is a diagonal matrix with diagonal elements

$$x_i = \begin{cases} \frac{1}{\lambda_i}\left(\frac{\sqrt{(1+\lambda_i\zeta_i)^2-\lambda_i^2\epsilon_i^2}}{\sqrt{\gamma}} - 1\right), & \gamma \le \gamma_i^0; \\ \zeta_i - \epsilon_i, & \gamma \ge \gamma_i^0. \end{cases} \tag{104}$$

It follows from (104), that if $\gamma \ge \gamma_i^0$, then x_i is equal to the lower bound on the uncertainty region of the ith eigenvalue of $\mathbf{C}_x$.

From (5), the MMSE estimate of $\mathbf{x}$ with covariance $\mathbf{C}_x$ given by (103) and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Lambda\mathbf{V}^*$ is

$$\hat{\mathbf{x}} = \mathbf{C}_x(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\mathbf{C}_x + \mathbf{I})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} = \mathbf{V}\mathbf{X}\left(\Lambda\mathbf{X}+\mathbf{I}\right)^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \tag{105}$$

Thus, the estimator $\hat{\mathbf{x}}$ of (105) is equivalent to the minimax ratio estimator given by Theorem 2, if

$\mathbf{X}(\Lambda\mathbf{X}+\mathbf{I})^{-1} = \mathbf{D}\Lambda^{-1}$. Now,

$$\frac{x_i}{\lambda_i x_i + 1} = \begin{cases} \frac{1}{\lambda_i}\left(1 - \frac{\sqrt{\gamma}}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}\right), & 1 \le \gamma \le \gamma_i^0; \\ \frac{\zeta_i - \epsilon_i}{1+\lambda_i(\zeta_i-\epsilon_i)}, & \gamma \ge \gamma_i^0, \end{cases} \tag{106}$$

so that indeed $\mathbf{X}(\Lambda\mathbf{X}+\mathbf{I})^{-1} = \mathbf{D}\Lambda^{-1}$.

Since the minimax ratio estimator minimizes the MSE for $\mathbf{C}_x = \mathbf{V}\mathbf{X}\mathbf{V}^*$, we may view the covariance $\mathbf{C}_x = \mathbf{V}\mathbf{X}\mathbf{V}^*$ as the "least-favorable" covariance in the ratio sense.

In [28] it was shown that the minimax regret estimator is an MMSE estimator matched to a covariance matrix with eigenvalues

$$c_i = \frac{1}{\lambda_i}\left(1 - \frac{1}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}\right). \tag{107}$$

Since the optimal value of γ is greater than 1 (unless there is no uncertainty), $x_i < c_i, 1 \le i \le m$, so that the minimax ratio estimator is matched to a covariance matrix with eigenvalues that are strictly smaller than the eigenvalues of the covariance matrix matched to the minimax regret estimator.

5 Example of the Minimax Ratio Estimator

We now consider examples illustrating the minimax ratio estimator. The examples we consider are the same as those considered in [28], for evaluating the minimax regret estimator.

Consider the estimation problem in which

$$\mathbf{y} = \mathbf{x} + \mathbf{w}, \tag{108}$$

where $\mathbf{x}$ is a length-n segment of a zero-mean stationary first order AR process with components x_i so that

$$E(x_i x_j) = \rho^{|j-i|} \tag{109}$$

for some parameter ρ, and $\mathbf{w}$ is a zero-mean random vector uncorrelated with $\mathbf{x}$ with known covariance $\mathbf{C}_w = \sigma^2\mathbf{I}$. We assume that we know the model (108) and that $\mathbf{x}$ is a segment of a stationary process, however, its covariance $\mathbf{C}_x$ is unknown.

To estimate $\mathbf{x}$, we may first estimate $\mathbf{C}_x$ from the observations $\mathbf{y}$. A natural estimate of $\mathbf{C}_x$ is given by

$$\widehat{\mathbf{C}}_x = [\widehat{\mathbf{C}}_y - \mathbf{C}_w]_+ = [\widehat{\mathbf{C}}_y - \sigma^2\mathbf{I}]_+, \tag{110}$$

where

$$\widehat{\mathbf{C}}_y(i,j) = \frac{1}{n}\sum_{k=1}^{n-|j-i|} y_k y_{k+|j-i|} \tag{111}$$

is an estimate of the covariance of $\mathbf{y}$ and $[\mathbf{A}]_+$ denotes the matrix in which the negative eigenvalues of $\mathbf{A}$ are replaced by 0. Thus, if $\mathbf{A}$ has an eigendecomposition $\mathbf{A} = \mathbf{U}\Sigma\mathbf{U}^{-1}$ where Σ is a diagonal matrix with diagonal elements σ_i, then $[\mathbf{A}]_+ = \mathbf{U}[\Sigma]_+\mathbf{U}^{-1}$ where $[\Sigma]_+$ is a diagonal matrix with ith diagonal element equal to $\max(0,\sigma_i)$. The estimate (110) can be regarded as the analogue for finite-length processes of the spectrum estimate based on the spectral subtraction method for infinite-length processes [33, 34].

Given $\widehat{\mathbf{C}}_x$, we may estimate $\mathbf{x}$ using an MMSE estimate matched to $\widehat{\mathbf{C}}_x$, which we refer to as a plug in estimator. However, as can be seen below in Figs. 1 and 2, we can further improve the estimation performance by using the minimax ratio estimator.

To compute the minimax ratio estimator, we choose $\mathbf{V}$ to be equal to the eigenvector matrix of the estimated covariance matrix $\widehat{\mathbf{C}}_x$, and $\zeta_i = \sigma_i$ where σ_i are the eigenvalues of $\widehat{\mathbf{C}}_x$. We would then like to choose ϵ_i to reflect the uncertainty in our estimate ζ_i. Since computing the standard deviation of ζ_i is difficult, we choose ϵ_i to be proportional to the standard deviation of an estimator $\tilde{\sigma}_x^2$ of the variance σ_x^2 of $\mathbf{x}$ where

$$\tilde{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \sigma_w^2. \tag{112}$$

We further assume that $\mathbf{x}$ and $\mathbf{w}$ are uncorrelated Gaussian random vectors. In this case, the variance of $\tilde{\sigma}_x^2$ is given by [28]

$$E\left\{\left(\tilde{\sigma}_x^2 - \sigma_x^2\right)^2\right\} = \frac{2}{n}\left((\sigma_x^2+\sigma^2)^2 + \sum_{i=2}^{n}\mathbf{C}_x^2(1,i)\right). \tag{113}$$

Since σ_x^2 and $\mathbf{C}_x(1,i)$ are unknown, we substitute their estimates $\widehat{\mathbf{C}}_x(1,i), 1 \le i \le m$. Finally, to ensure that $\epsilon_i \le \zeta_i$, we choose

$$\epsilon_i = \min\left(\zeta_i, A\sqrt{\frac{2}{n}\left((\widehat{\mathbf{C}}_x^2(1,1)+\sigma^2)^2 + \sum_{i=2}^{n}\widehat{\mathbf{C}}_x^2(1,i)\right)}\right), \tag{114}$$

where A is a proportionality factor.

In Fig. 1, we plot the MSE of the minimax ratio estimator averaged over 1000 noise realizations as a function of the SNR defined by $-10\log\sigma^2$ for $\rho = 0.8$, $n = 10$ and $A = 4$. The performance of the "plug in" MMSE estimator matched to the estimated covariance matrix $\widehat{\mathbf{C}}_x$, the minimax MSE estimator and the minimax regret estimator of [28] are plotted for comparison. We also plot the MSE resulting from a Wiener filter matched to the known covariance, which is the optimal MSE attainable when $\mathbf{C}_x$ is known. As can be seen from the figure, the minimax ratio estimator can significantly increase the estimation performance at low to intermediate SNR values, and in this range, the performance of the minimax ratio estimator is close to the optimal performance.

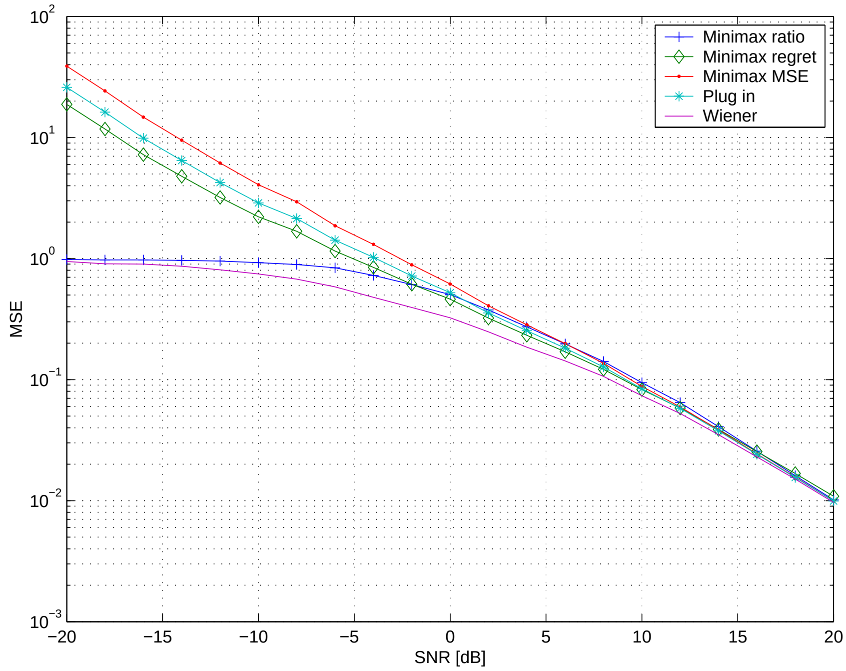

Figure 1: MSE in estimating $\mathbf{x}$ as a function of SNR using the minimax ratio estimator, the minimax regret estimator, the minimax MSE estimator and the plug in MMSE estimator matched to the estimated covariance matrix. The performance of the optimal Wiener estimator is plotted for comparison.

We next consider the case in which the vector $\mathbf{x}$ is filtered with an LTI filter with length-4 impulse response given by

$$h[0] = 1, \;\; h[1] = 0.4, \;\; h[2] = 0.2, \;\; h[3] = 0.1. \tag{115}$$

In Fig. 2, we plot the MSE of the minimax ratio, minimax regret, plug in and minimax MSE estimators averaged over 1000 noise realizations as a function of the SNR, for $\rho = 0.8$, $n = 10$ and $A = 4$. For comparison, we also plot the bound on the performance given by the MSE of the Wiener estimator. As can be seen, the trends in performance are similar to the previous example.

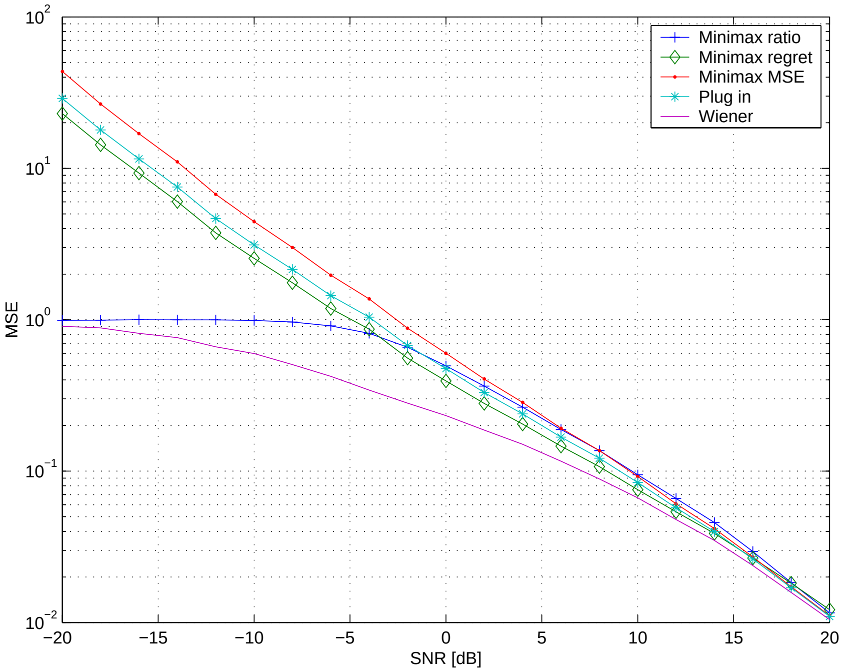

Figure 2: MSE in estimating $\mathbf{x}$ from a noisy filtered version as a function of SNR using the minimax ratio estimator, the minimax regret estimator, the minimax MSE estimator and the plug in MMSE estimator matched to the estimated covariance matrix. The performance of the optimal Wiener estimator is plotted for comparison.

References

[1] N. Wiener, *The Extrapolation, Interpolation and Smoothing of Stationary Time Series*, New York, NY: John Wiley & Sons, 1949.

[2] A. Kolmogorov, "Interpolation and extrapolation," *Bull. Acad. Sci., USSR, Ser. Math.*, vol. 5, pp. 3–14, 1941.

[3] K. S. Vastola and H. V. Poor, "Robust Wiener-Kolmogorov theory," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 316–327, Mar. 1984.

[4] P. J. Huber, "Robust estimation of a location parameter," *Ann. Math. Statist.*, vol. 35, pp. 73–101, 1964.

[5] P. J. Huber, *Robust Statistics*, New York: NY, John Wiley & Sons, Inc., 1981.

[6] L. Breiman, "A note on minimax filtering," *Annals of Probability*, vol. 1, pp. 175–179, 1973.

[7] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Inst.*, vol. 304, pp. 171–185, Oct./Nov. 1977.

[8] H. V. Poor, "On robust Wiener filtering," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 521–526, June 1980.

[9] J. Franke, "Minimax-robust prediction of discrete time series," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 68, pp. 337–364, 1985.

[10] G. Moustakides and S. A. Kassam, "Minimax robust equalization for random signals through uncertain channels," in *Proc. 20th Annual Allerton Conf. Communication, Control and Computing*, Oct. 1982, pp. 945–954.

[11] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *IEEE Proc.*, vol. 73, pp. 433–481, Mar. 1985.

[12] S. Verdú and H. V. Poor, "On minimax robustness: A general approach and applications," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 328–340, Mar. 1984.

[13] S. A. Kassam and H. V. Poor, "Robust signal processing for communication systems," *IEEE Commun. Mag.*, vol. 21, pp. 20–28, 1983.

[14] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, New York, NY: McGraw Hill, Inc., third edition, 1991.

[15] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, MPS-SIAM Series on Optimization, 2001.

[16] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 40–95, Mar. 1996.

[17] Y. Nesterov and A. Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming*, Philadelphia, PE: SIAM, 1994.

[18] F. Alizadeh, *Combinatorial Optimization With Interior Point Methods and Semi-Definite Matrices*, Ph.D. thesis, University of Minnesota, Minneapolis, MN, Oct. 1991.

[19] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation with bounded data uncertainties," Tech. Rep. CCIT 437, Technion—Israel Institute of Technology, Haifa, Israel, July 2003, submitted to *IEEE Trans. Signal Processing*, June 2003.

[20] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 783–795, Nov. 1973.

[21] M. Feder and N. Merhav, "Universal composite hypothesis testing: A competitive minimax approach," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1504–1517, June 2002.

[22] E. Levitan and N. Merhav, "A competitive Neyman-Pearson approach to universal hypothesis testing with applications," *IEEE Trans. Inform. Theory*, vol. 48, pp. 2215–2229, Aug. 2002.

[23] N. Merhav and M. Feder, "Universal prediction," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2124–2147, Oct. 1998.

[24] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge, UK: Cambridge Univ. Press, 1985.

[25] S. A. Kassam, "Robust hypothesis testing for bounded classes of probability densities," *IEEE Trans. Inform. Theory*, vol. IT-27, pp. 242–247, 1980.

[26] V. P. Kuznetsov, "Stable detection when the signal and spectrum of normal noise are inaccurately known," *Telecomm. Radio Eng. (English transl.)*, vol. 30/31, pp. 58–64, 1976.

[27] R. Gray, "Toeplitz and circulant matrices: A review," Tech. Rep. 6504-1, Inform. Sys. Lab., Stanford Univ., Stanford, CA, Apr. 1977.

[28] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation in linear models," Tech. Rep. CCIT 422, Technion—Israel Institute of Technology, Haifa, Israel, May 2003, *IEEE Trans. Signal Processing*, to appear.

[29] M. X. Goemans and D. P. Williamson, "Approximation algorithms for MAX-3-CUT and other problems via complex semidefinite programming," in *ACM Symposium on Theory of Computing*, 2001, pp. 443–452.

[30] J. F. Strum, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11–12, pp. 625–653, 1999.

[31] D. Peaucelle, D. Henrion, and Y. Labit, "Users guide for SeDuMi interface 1.03," available at `http://www.laas.fr/peaucell/SeDuMiInt.html`.

[32] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, Philadelphia, PA: SIAM, 1994.

[33] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Signal Proc.*, vol. 27, pp. 113–120, 1979.

[34] M. Berouti, R. Schwartz, and J. Makhoul, "Enhancement of speech corrupted by acoustic noise," in *Proc. Int. Conf. Acoust., Speech, Signal Processing (ICASSP-1979)*, 1979, pp. 208–211.