CCIT Report #422 May 2003

A Competitive Minimax Approach to Robust Estimation in Linear Models

Yonina C. Eldar and Neri Merhav*

May 4, 2003

SP_EDICS category: 2-ESTM.

Abstract

We consider the problem of estimating, in the presence of model uncertainties, a random vector $\mathbf{x}$ that is observed through a linear transformation $\mathbf{H}$ and corrupted by additive noise. We first assume that both the covariance of $\mathbf{x}$ and the transformation $\mathbf{H}$ are not completely specified, and develop the linear estimator that minimizes the worst-case mean-squared error (MSE) across all possible covariance matrices and transformations $\mathbf{H}$ in the region of uncertainty. Although the minimax approach has enjoyed widespread use in the design of robust methods, we show that its performance is often unsatisfactory. To improve the performance over the minimax MSE estimator, we develop a competitive minimax approach, for the case where $\mathbf{H}$ is known but the covariance of $\mathbf{x}$ is subject to uncertainties, and seek the linear estimator that minimizes the worst-case *regret*, namely the worst-case difference between the MSE attainable using a linear estimator, ignorant of the signal covariance, and the optimal MSE attained using a linear estimator that knows the signal covariance. The linear minimax regret estimator is shown to be equal to a minimum MSE (MMSE) estimator corresponding to a certain choice of signal covariance, that depends explicitly on the uncertainty region. We demonstrate through an example that the minimax regret approach can improve the performance over both the minimax MSE approach and a "plug in" approach, in which the estimator is chosen to be equal to the MMSE estimator with an estimated covariance matrix replacing the true unknown covariance. We then show that although the optimal minimax regret estimator in the case in which the signal and noise are jointly Gaussian is nonlinear, we often do not loose much by restricting attention to linear estimators.

*Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: {yonina,merhav}@ee.technion.ac.il. Fax: +972-4-8323041.

1 Introduction

The theory of estimation in linear models has been studied extensively in the past century, following the classical works of Wiener [1] and Kolmogorov [2]. A fundamental problem considered by Wiener and Kolmogorov is that of estimating a stationary random signal in additive stationary noise, where the signal may be filtered by a linear time invariant (LTI) channel. The desired signal is estimated using a linear estimator which is obtained by filtering the received signal with an LTI estimation filter. When the signal and noise spectral densities as well as the channel are completely specified, the estimation filter minimizing the mean-squared error (MSE) is the well-known Wiener filter.

In practice, the actual spectral densities and the channel may not be known exactly. If the spectral densities and the channel deviate from the ones assumed, then the performance of the Wiener filter matched to the assumed spectral densities and channel can deteriorate considerably [3]. In such cases, it is desirable to design a robust filter whose performance is reasonably good across all possible spectral densities and channels, in the region of uncertainty.

The most common approach for designing robust estimation filters is the minimax MSE approach, initiated by Huber [4, 5], in which the estimation filter is chosen to maximize the worst-case MSE over an appropriately chosen class of spectral densities [6, 7, 8, 3, 9] where the channel is assumed to be known. A similar approach has also been used to develop a robust estimator for the case in which the spectral densities are known and the channel is subject to uncertainties [10]. The minimax approach, in which the goal is to optimize the worst-case performance, is one of the major techniques for designing robust systems with respect to modelling uncertainties, and has been applied to many problems in detection and estimation [11, 12, 13].

In this paper, we consider a finite-dimensional analogue of the classical Wiener filtering problem, so that we consider estimating a finite number of parameters from finitely many observations, where the motivation is to obtain non-asymptotic results. Specifically, we treat the problem of estimating a random vector $\mathbf{x}$ that is observed through a linear transformation $\mathbf{H}$ and corrupted by additive noise $\mathbf{w}$. If the signal and noise covariance matrices as well as the transformation $\mathbf{H}$ are completely specified, then the linear minimum MSE (MMSE) estimator of $\mathbf{x}$ for this problem is well known [14].

In many practical applications the covariance of the noise can be assumed known in the sense that it can be estimated within high accuracy. This is especially true if the noise components are uncorrelated and identically distributed, which is often the case in practice. The signal, on the other hand, will typically have a broader correlation function, so that estimating this correlation from the data with high accuracy often necessitates a larger sample size than is available. Therefore, in this paper, we develop methods for designing robust estimators in the case in which the covariance of the noise is known precisely, but the covariance of the desired signal $\mathbf{x}$ and the model matrix $\mathbf{H}$ are not completely specified.

Following the popular minimax approach, in Section 3 we consider the case in which $\mathbf{H}$ is known, and seek the linear estimator that minimizes the worst case MSE over all possible covariance matrices. As we show, the resulting estimator, referred to as the minimax MSE estimator, is an MMSE estimator matched to

the worst possible choice of covariance matrix. In Section 4, we develop a minimax estimator that minimizes the worst-case MSE when both the covariance matrix and the model matrix $\mathbf{H}$ are subject to uncertainties. In this case, we show that the optimal estimator can be found by solving a semidefinite programming (SDP) problem [15, 16, 17], which is a convex optimization problem that can be solved very efficiently, *e.g.,* using interior point methods [17, 18].

Although the minimax approach has enjoyed widespread use in the design of robust methods for signal processing and communication [11, 13], its performance is often unsatisfactory. The main limitation of this approach is that it tends to be overly conservative since it optimizes the performance for the worst possible choice of unknowns. As we show in the context of a concrete example in Section 6, this can often lead to degraded performance.

To improve the performance of the minimax MSE estimator, in Section 5, we propose a new competitive approach to robust estimation for the case where $\mathbf{H}$ is known, and seek a linear estimator whose performance is as close as possible to that of the optimal estimator for all possible values of the covariance matrix. Specifically, we seek the estimator that minimizes the worst-case *regret*, which is the difference between the MSE of the estimator, ignorant of the signal covariance, and the smallest attainable MSE with a linear estimator that knows the signal covariance. By considering the *difference* between the MSE and the optimal MSE rather than the MSE directly, we can counterbalance the conservative character of the minimax approach, as is evident in the example we consider in Section 6. It would also be desirable to develop the minimax estimator that minimizes the worst-case regret when both $\mathbf{H}$ and the covariance matrix are subject to uncertainties. However, since this problem is very difficult, for analytical tractability, we restrict our attention to the case in which $\mathbf{H}$ is known.

The minimax regret concept has recently been used to develop a linear estimator for the unknowns $\mathbf{x}$ in the same linear model considered in this paper, where it is assumed that $\mathbf{x}$ is *deterministic* but unknown [19]. Similar competitive approaches have been used in a variety of other contexts, for example, universal source coding [20], hypothesis testing [21, 22], and prediction (see [23] for a survey and references therein).

For analytical tractability, in our development we restrict attention to the class of linear estimators. In some cases, there is also theoretical justification for this restriction. As is well known [14], if $\mathbf{x}$ and $\mathbf{w}$ are jointly Gaussian vectors with known covariance matrices, then the estimator that minimizes the MSE, among all linear and nonlinear estimators, is the linear MMSE estimator. In Section 7, we show that this property does not hold when minimizing the worst-case regret with covariance uncertainties, even in the Gaussian case. Nevertheless, we demonstrate that in many cases we do not loose much by confining ourselves to linear estimators. In particular, we develop a lower bound on the smallest possible worst-case regret attainable with a third-order (cubic) nonlinear estimator, when estimating a Gaussian random variable contaminated by independent Gaussian noise, and show that the linear minimax regret estimator often nearly achieves this bound, particularly at high SNR. This provides additional justification for the restriction to linear estimators in the context of minimax regret estimation.

Before proceeding to the detailed development, in Section 2, we provide an overview of our problem.

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. $\mathbf{I}$ denotes the identity matrix of appropriate dimension, $(\cdot)^*$ denotes the Hermitian conjugate of the corresponding matrix, and $\hat{(\cdot)}$ denotes an estimated vector or matrix. The cross covariance between the random vectors $\mathbf{x}$ and $\mathbf{y}$ is denoted by $\mathbf{C}_{xy}$, and the covariance of $\mathbf{x}$ is denoted by $\mathbf{C}_x$. $\mathcal{N}(m,\sigma^2)$ denotes the Gaussian distribution with mean m and covariance σ^2.

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

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

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

If $\mathbf{H}$ and $\mathbf{C}_x$ are known and $\mathbf{C}_x$ is positive definite, 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

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

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

$$\begin{aligned} \hat{\mathbf{x}} &= \mathbf{C}_x\left(\mathbf{I} - \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{C}_x^{-1})^{-1}\right)\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} \\ &= (\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{C}_x^{-1})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \end{aligned} \tag{5}$$

If $\mathbf{C}_x$ or $\mathbf{H}$ are unknown, then we cannot implement the MMSE estimator (3). Instead, we may seek the estimator that minimizes the *worst-case* MSE over all possible choices of $\mathbf{C}_x$ and $\mathbf{H}$ that are consistent with our prior information on these unknowns. In Section 3 and Sections 5–6, we consider the case in which $\mathbf{H}$ is a known $n \times m$ matrix with rank m, and $\mathbf{C}_x$ is not completely specified. In Section 4, we consider the

case in which both $\mathbf{C}_x$ and $\mathbf{H}$ are subject to uncertainties.

To reflect the uncertainty in our knowledge of the true covariance matrix, we consider two different models of uncertainty which resemble the "band model" widely used in the continuous-time case [7, 25, 26, 3]. Although these models are similar in nature, depending on the optimality criteria, a particular model may be mathematically more convenient. In the first model, we assume that $\mathbf{C}_x$ commutes with $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$, and that each of the nonnegative eigenvalues $\delta_i \geq 0, 1 \leq i \leq m$ of $\mathbf{C}_x$ satisfies

$$l_i \leq \delta_i \leq u_i, \quad 1 \leq i \leq m, \tag{6}$$

where $l_i \geq 0$ and u_i are known. 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 commute [27].

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

$$\delta_i = \zeta_i + e_i, \quad e_i^2 \leq \epsilon_i^2, \quad 1 \leq i \leq m, \tag{7}$$

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.

In the second model,

$$\mathbf{C}_x = \widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x, \quad \|\delta\mathbf{C}_x\| \leq \epsilon, \tag{8}$$

where $\widetilde{\mathbf{C}}_x$ is known, $\|\cdot\|$ denotes the matrix spectral norm [24], *i.e.,* the largest singular value of the corresponding matrix, and ϵ is chosen such that $\widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x \geq 0$ for all $\|\delta\mathbf{C}_x\| \leq \epsilon$. In this model, $\mathbf{C}_x$ is not assumed to commute with $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. As a consequence, we can no longer constraint each of the eigenvalues of $\mathbf{C}_x$ as we did in the first model, but rather we can only restrict the largest eigenvalue, or equivalently, the spectral norm. If $\mathbf{C}_x$ is constrained to commute with $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ for all $\delta\mathbf{C}_x$, then the uncertainty model (8) is equivalent to the uncertainty model (7) with $\{\zeta_i, 1 \leq i \leq m\}$ equal to the eigenvalues of $\widetilde{\mathbf{C}}_x$ and $\epsilon_i = \epsilon, 1 \leq i \leq m$.

Given ζ_i in the first model or $\widetilde{\mathbf{C}}_x$ in the second model, a straightforward approach to estimating $\mathbf{x}$ is to use an MMSE estimate corresponding to the estimated covariance. However, as we demonstrate through an example in Section 6, 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.

In Section 3, we develop the minimax estimators that minimize the worst case MSE over all covariance matrices $\mathbf{C}_x$ that satisfy each of the two uncertainty models (6) and (8). As we show, the resulting estimators are MMSE estimators matched to the worst possible choice of eigenvalues *i.e.,* $\delta_i = u_i$ in the first model and $\mathbf{C}_x = \widetilde{\mathbf{C}}_x + \epsilon\mathbf{I}$ in the second model. Since these estimators are matched to the worst possible choice of

parameters, in general they tend to be overly conservative, which can often lead to degraded performance, as is evident in the example in Section 6. In this example, the minimax MSE estimator performs worse than the MMSE estimator matched to the estimated covariance matrix.

In Section 4, we consider the case in which the model matrix $\mathbf{H}$ is also subject to uncertainties, and develop a minimax MSE estimator that minimizes the worst case MSE over all possible covariance matrices $\mathbf{C}_x$ and model matrices $\mathbf{H}$. We assume that both $\mathbf{C}_x$ and $\mathbf{H}$ obey an uncertainty model of the form (8).

To improve the performance of the minimax estimators, in Section 5 we consider a competitive approach in which we seek the linear estimator that minimizes the worst-case regret, where the regret is the difference between the MSE of an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ and the best possible MSE attainable using a linear estimator where $\mathbf{C}_x$ is assumed to be known. In this case, we consider only the first model in which $\mathbf{C}_x$ commutes with $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and its eigenvalues are given by (6). As we show, the resulting estimator can also be interpreted as an MMSE estimator matched to a covariance matrix which depends on the nominal value ζ_i and the uncertainty interval ϵ_i, as well as on the eigenvalues of $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. In the example in Section 6, we demonstrate that the minimax regret estimator can improve the performance over both the minimax MSE estimator and the MMSE estimator matched to the estimated covariance matrix.

3 Minimax MSE For Known H

We first consider the case in which the model matrix $\mathbf{H}$ is known, and seek the linear estimator that minimizes the worst-case MSE over all possible values of $\mathbf{C}_x$ that commute with $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and with eigenvalues δ_i satisfying (6). Thus, let $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have an eigendecomposition

$$\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Lambda\mathbf{V}^*, \tag{9}$$

where $\mathbf{V}$ is a unitary matrix and Λ is a diagonal matrix with diagonal elements $\lambda_i > 0$. Then $\mathbf{C}_x$ has the form

$$\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*, \tag{10}$$

where Δ is a diagonal matrix with diagonal elements $l_i \le \delta_i \le u_i$.

We now consider the problem

$$\min_{\mathbf{G}} \max_{l_i \le \delta_i \le u_i} E(\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2) = \min_{\mathbf{G}} \left\{ \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \max_{l_i \le \delta_i \le u_i} \mathcal{Q}(\mathbf{C}_x) \right\}, \tag{11}$$

where from (2),

$$\mathcal{Q}(\mathbf{C}_x) = \mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\right). \tag{12}$$

To find the maximum value of $\mathcal{Q}(\mathbf{C}_x)$ we rely on the following lemma.

Lemma 1. *Let* $\mathbf{W}, \mathbf{T}$ *and* $\mathbf{M}$ *be nonnegative definite matrices with* $\mathbf{W} \le \mathbf{T}$. *Then* $Tr(\mathbf{MW}) \le Tr(\mathbf{MT})$.

Proof. Since $\mathbf{M} \ge 0$ and $\mathbf{T} - \mathbf{W} \ge 0$ we can define the nonnegative symmetric square-roots $\mathbf{M}^{1/2}$ and

$(\mathbf{T}-\mathbf{W})^{1/2}$. Denoting $\mathbf{A}=(\mathbf{T}-\mathbf{W})^{1/2}\mathbf{M}^{1/2}$ we have

$$\mathrm{Tr}(\mathbf{M}(\mathbf{T}-\mathbf{W}))=\mathrm{Tr}(\mathbf{M}^{1/2}(\mathbf{T}-\mathbf{W})^{1/2}(\mathbf{T}-\mathbf{W})^{1/2}\mathbf{M}^{1/2})=\mathrm{Tr}(\mathbf{A}^*\mathbf{A})\geq 0, \tag{13}$$

since $\mathrm{Tr}(\mathbf{Z})\geq 0$ for any $\mathbf{Z}\geq 0$. Thus, $\mathrm{Tr}(\mathbf{MT})\geq\mathrm{Tr}(\mathbf{MW})$. □

Let $\mathbf{C}_x$ be an arbitrary matrix of the form (10) with eigenvalues $l_i\leq\delta_i\leq u_i$. Then,

$$\mathbf{C}_x\leq\mathbf{V}\mathbf{Z}\mathbf{V}^*, \tag{14}$$

where $\mathbf{Z}$ is a diagonal matrix with diagonal elements u_i. This then implies from Lemma 1 that

$$\mathrm{Tr}\left(\mathbf{C}_x(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})\right)\leq\mathrm{Tr}\left(\mathbf{V}\mathbf{Z}\mathbf{V}^*(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})\right) \tag{15}$$

with equality if $\mathbf{C}_x=\mathbf{V}\mathbf{Z}\mathbf{V}^*$, so that $\mathcal{Q}(\mathbf{C}_x)$ is maximized for the worst possible choice of eigenvalues *i.e.,* $\delta_i=u_i$. The problem of (11) therefore reduces to minimizing the MSE of (2) where we substitute $\mathbf{C}_x=\mathbf{V}\mathbf{Z}\mathbf{V}^*$. The optimal estimator is then the linear MMSE estimator of (3) or (5) with $\mathbf{C}_x=\mathbf{V}\mathbf{Z}\mathbf{V}^*$.

Using the eigendecomposition of $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ given by (9), we can express $\hat{\mathbf{x}}$ of (5) as

$$\hat{\mathbf{x}}=(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}^*+\mathbf{C}_x^{-1})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}=\mathbf{V}(\Lambda+\mathbf{Z}^{-1})^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}=\mathbf{V}\mathbf{Q}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}, \tag{16}$$

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

$$q_i=\frac{u_i}{u_i\lambda_i+1}. \tag{17}$$

We now seek the linear estimator that minimizes the worst-case MSE over all covariance matrices $\mathbf{C}_x$ of the form (8). Thus, we consider the problem

$$\min_{\mathbf{G}}\max_{\|\delta\mathbf{C}_x\|\leq\epsilon}E(\|\mathbf{G}\mathbf{y}-\mathbf{x}\|^2)=\min_{\mathbf{G}}\left\{\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)+\max_{\|\delta\mathbf{C}_x\|\leq\epsilon}\mathcal{Q}(\delta\mathbf{C}_x)\right\}, \tag{18}$$

where

$$\mathcal{Q}(\delta\mathbf{C}_x)=\mathrm{Tr}\left((\widetilde{\mathbf{C}}_x+\delta\mathbf{C}_x)(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})\right). \tag{19}$$

Since the condition $\|\delta\mathbf{C}_x\|\leq\epsilon$ is equivalent to the condition $\delta\mathbf{C}_x\leq\epsilon\mathbf{I}$, we can use Lemma 1 to conclude that

$$\mathcal{Q}(\delta\mathbf{C}_x)=\mathrm{Tr}\left((\widetilde{\mathbf{C}}_x+\delta\mathbf{C}_x)(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})\right)\leq\mathrm{Tr}\left((\widetilde{\mathbf{C}}_x+\epsilon\mathbf{I})(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})\right), \tag{20}$$

with equality for $\delta\mathbf{C}_x=\epsilon\mathbf{I}$. Therefore, the problem (18) reduces to minimizing the MSE of (2) where we substitute $\delta\mathbf{C}_x=\epsilon\mathbf{I}$, and the optimal estimator is the linear MMSE estimator with $\mathbf{C}_x=\widetilde{\mathbf{C}}_x+\epsilon\mathbf{I}$.

We summarize our results on minimax MSE estimation with known $\mathbf{H}$ in the following theorem.

Theorem 1 (Minimax MSE estimators). *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* $\lambda_i > 0$, *let* $\mathcal{S}_1$ *denote the set of matrices* $\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*$ *where* Δ *is an* $m \times m$ *diagonal matrix with diagonal elements* $0 \leq l_i \leq \delta_i \leq u_i$, *and let* $\mathcal{S}_2$ *denote the set of matrices* $\mathbf{C}_x = \widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x$ *where* $\widetilde{\mathbf{C}}_x$ *is known and* $\|\delta\mathbf{C}_x\| \leq \epsilon$. *Here* ϵ *is chosen such that* $\widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x \geq 0$ *for all* $\|\delta\mathbf{C}_x\| \leq \epsilon$. *Then,*

1. *The solution to the problem* $\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\mathbf{C}_x \in \mathcal{S}_1} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$ *is an MMSE estimator matched to the covariance* $\mathbf{C}_x = \mathbf{V}\mathbf{Z}\mathbf{V}^*$ *where* $\mathbf{Z}$ *is an* $m \times m$ *diagonal matrix with diagonal elements* u_i, *and can be expressed as*

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

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

$$q_i = \frac{u_i}{u_i\lambda_i + 1}.$$

2. *The solution to the problem* $\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\mathbf{C}_x \in \mathcal{S}_2} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2)$ *is an MMSE estimator matched to the covariance* $\mathbf{C}_x = \widetilde{\mathbf{C}}_x + \epsilon\mathbf{I}$.

4 Minimax MSE For Unknown H

In the previous section, we developed the minimax MSE estimator under the assumption that the model matrix $\mathbf{H}$ is known exactly. In many engineering applications the model matrix $\mathbf{H}$ is also subject to uncertainties. For example, the matrix $\mathbf{H}$ may be estimated from noisy data in which case $\mathbf{H}$ is an approximation to some nominal underlying matrix. If the actual data matrix is $\widetilde{\mathbf{H}} + \delta\mathbf{H}$ for some unknown matrix $\delta\mathbf{H}$, then an estimator designed based on $\widetilde{\mathbf{H}}$ alone may perform poorly.

To explicitly take uncertainties in $\mathbf{H}$ into account, we now consider a robust estimator that minimizes the worst-case MSE over all possible covariance and model matrices. Specifically, suppose now that the model matrix $\mathbf{H}$ is not known exactly, but is rather given by

$$\mathbf{H} = \widetilde{\mathbf{H}} + \delta\mathbf{H}, \quad \|\delta\mathbf{H}\| \leq \rho, \tag{21}$$

where $\widetilde{\mathbf{H}}$ is known. Similarly, the covariance matrix $\mathbf{C}_x$ is given by (8). We then seek the linear estimator that is the solution to the problem

$$\begin{aligned} &\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\delta\mathbf{C}_x\| \leq \epsilon, \|\delta\mathbf{H}\| \leq \rho} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) = \\ &= \min_{\mathbf{G}} \max_{\|\delta\mathbf{C}_x\| \leq \epsilon, \|\delta\mathbf{H}\| \leq \rho} \left\{ \mathrm{Tr}\left((\widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x)(\mathbf{I} - \mathbf{G}(\widetilde{\mathbf{H}} + \delta\mathbf{H}))(\mathbf{I} - \mathbf{G}(\widetilde{\mathbf{H}} + \delta\mathbf{H}))^* \right) + \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) \right\}. \end{aligned} \tag{22}$$

In Theorem 2 below, we show that the problem (22) 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[1] [17, 18], which are guaranteed to converge to the global optimum, the optimal estimator can be computed very efficiently in polynomial time. The SDP formulation can also be used to derive necessary and sufficient conditions for optimality.

Theorem 2. *Let* $\mathbf{x}$ *denote the unknown parameters in the model* $\mathbf{y} = (\widetilde{\mathbf{H}} + \delta\mathbf{H})\mathbf{x} + \mathbf{w}$*, where* $\widetilde{\mathbf{H}}$ *is a known* $n \times m$ *matrix and* $\delta\mathbf{H}$ *is an unknown matrix satisfying* $\|\delta\mathbf{H}\| \leq \rho$*,* $\mathbf{x}$ *is a zero-mean random vector uncorrelated with* $\mathbf{w}$ *with covariance* $\widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x$ *where* $\widetilde{\mathbf{C}}_x$ *is a known* $m \times m$ *matrix and* $\delta\mathbf{C}_x$ *is an unknown matrix satisfying* $\|\delta\mathbf{C}_x\| \leq \epsilon$ *with* ϵ *chosen such that* $\widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x \geq 0$ *for all* $\|\delta\mathbf{C}_x\| \leq \epsilon$*, and* $\mathbf{w}$ *is a zero-mean random vector with covariance* $\mathbf{C}_w$*. Then the problem*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\delta\mathbf{C}_x\|\leq\epsilon, \|\delta\mathbf{H}\|\leq\rho} E\left(\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right)$$

is equivalent to the semidefinite programming problem

$$\min_{t,\mathbf{G},\lambda,\mathbf{X},\mathbf{Y}} t$$

subject to

$$Tr((\widetilde{\mathbf{C}}_x + \epsilon\mathbf{I})\mathbf{X}) + Tr(\mathbf{Y}) \leq t$$

$$\begin{bmatrix} \mathbf{Y} & \mathbf{G} \\ \mathbf{G}^* & \mathbf{C}_w^{-1} \end{bmatrix} \geq 0$$

$$\begin{bmatrix} \mathbf{X} - \lambda\mathbf{I} & (\mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}})^* & \mathbf{0} \\ \mathbf{I} - \mathbf{G}\widetilde{\mathbf{H}} & \mathbf{I} & -\rho\mathbf{G} \\ \mathbf{0} & -\rho\mathbf{G}^* & \lambda\mathbf{I} \end{bmatrix} \geq 0.$$

Proof. We begin by noting that

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{\|\delta\mathbf{C}_x\|\leq\epsilon, \|\delta\mathbf{H}\|\leq\rho} E\left(\|\hat{\mathbf{x}} - \mathbf{x}\|^2\right) = \min_{t,\mathbf{G},\tau} t \tag{23}$$

subject to

$$\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \tau \leq t \tag{24}$$

$$\mathrm{Tr}\left((\widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x)(\mathbf{I} - \mathbf{G}(\widetilde{\mathbf{H}} + \delta\mathbf{H}))(\mathbf{I} - \mathbf{G}(\widetilde{\mathbf{H}} + \delta\mathbf{H}))^*\right) \leq \tau, \quad \forall\delta\mathbf{C}_x : \|\delta\mathbf{C}_x\| \leq \epsilon, \forall\delta\mathbf{H} : \|\delta\mathbf{H}\| \leq \rho. \tag{25}$$

[1]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.

To simplify the constraint (25) we rely on the following proposition, the proof of which is provided in Appendix A.

Proposition 1. *Let $\mathcal{G}(\mathbf{A})$ and $\mathcal{Q}(\mathbf{B})$ be nonnegative matrices which are functions of the matrices $\mathbf{A}$ and $\mathbf{B}$ respectively. Then the problem*

$$\min \tau \tag{26}$$

subject to

$$Tr(\mathcal{G}(\mathbf{A})\mathcal{Q}(\mathbf{B})) \leq \tau, \quad \forall \mathbf{A} : \|\mathbf{A}\| \leq \alpha, \forall \mathbf{B} : \|\mathbf{B}\| \leq \beta \tag{27}$$

is equivalent to the problem of (26) subject to

$$Tr(\mathcal{G}(\mathbf{A})\mathbf{X}) \leq \tau, \quad \forall \mathbf{A} : \|\mathbf{A}\| \leq \alpha \tag{28}$$

$$\mathcal{Q}(\mathbf{B}) \leq \mathbf{X}, \quad \forall \mathbf{B} : \|\mathbf{B}\| \leq \beta. \tag{29}$$

Using Proposition 1, we can express the constraint (25) as

$$\mathrm{Tr}((\widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x)\mathbf{X}) \leq \tau, \quad \forall \delta\mathbf{C}_x : \|\delta\mathbf{C}_x\| \leq \epsilon; \tag{30}$$

$$(\mathbf{I} - \mathbf{G}(\widetilde{\mathbf{H}} + \delta\mathbf{H}))(\mathbf{I} - \mathbf{G}(\widetilde{\mathbf{H}} + \delta\mathbf{H}))^* \leq \mathbf{X}, \quad \forall \delta\mathbf{H} : \|\delta\mathbf{H}\| \leq \rho. \tag{31}$$

From Lemma 1 and the fact that (31) implies that $\mathbf{X} \geq 0$,

$$\max_{\|\delta\mathbf{C}_x\| \leq \epsilon} \mathrm{Tr}((\widetilde{\mathbf{C}}_x + \delta\mathbf{C}_x)\mathbf{X}) = \mathrm{Tr}((\widetilde{\mathbf{C}}_x + \epsilon\mathbf{I})\mathbf{X}), \tag{32}$$

so that (30) reduces to

$$\mathrm{Tr}((\widetilde{\mathbf{C}}_x + \epsilon\mathbf{I})\mathbf{X}) \leq \tau. \tag{33}$$

Since we would like to minimize τ, the optimal choice is $\tau = \mathrm{Tr}((\widetilde{\mathbf{C}}_x + \epsilon\mathbf{I})\mathbf{X})$.

To treat the constraint (31), 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 $\mathbf{M} \geq 0$ if and only if $\mathbf{C} > 0$ and $\Delta_{\mathbf{C}} \geq 0$ where $\Delta_{\mathbf{C}}$ is the Schur complement of $\mathbf{C}$ in $\mathbf{M}$ and is given by

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

Equivalently, $\mathbf{M} \geq 0$ if and only if $\mathbf{A} > 0$ and $\Delta_{\mathbf{A}} \geq 0$ where $\Delta_{\mathbf{A}}$ is the Schur complement of $\mathbf{A}$ in $\mathbf{M}$ and is given by

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

From Lemma 2 it follows that (31) is equivalent to the condition

$$\left[\begin{array}{cc} \mathbf{X} & (\mathbf{I}-\mathbf{G}(\widetilde{\mathbf{H}}+\delta\mathbf{H}))^* \\ \mathbf{I}-\mathbf{G}(\widetilde{\mathbf{H}}+\delta\mathbf{H}) & \mathbf{I} \end{array}\right] \geq 0, \quad \forall \delta\mathbf{H} : \|\delta\mathbf{H}\| \leq \rho, \tag{34}$$

which is equivalent to

$$\mathbf{A}(\mathbf{X},\mathbf{G}) \geq \mathbf{B}^*(\mathbf{G})\delta\mathbf{H}\mathbf{C} + \mathbf{C}^*\delta\mathbf{H}^*\mathbf{B}(\mathbf{G}), \quad \forall \delta\mathbf{H} : \|\delta\mathbf{H}\| \leq \rho, \tag{35}$$

where

$$\begin{aligned} \mathbf{A}(\mathbf{X},\mathbf{G}) &= \left[\begin{array}{cc} \mathbf{X} & (\mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}})^* \\ \mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}} & \mathbf{I} \end{array}\right]; \\ \mathbf{B}(\mathbf{G}) &= \left[\begin{array}{cc} \mathbf{0} & \mathbf{G}^* \end{array}\right]; \\ \mathbf{C} &= \left[\begin{array}{cc} \mathbf{I} & \mathbf{0} \end{array}\right]. \end{aligned} \tag{36}$$

We now exploit the following proposition, the proof of which can be found in [19], and relies on the Cauchy-Schwarz inequality and Lemma 3 below.

Proposition 2. *Given matrices* $\mathbf{P},\mathbf{Q},\mathbf{A}$ *with* $\mathbf{A}=\mathbf{A}^*$*,*

$$\mathbf{A} \geq \mathbf{P}^*\mathbf{Z}\mathbf{Q} + \mathbf{Q}^*\mathbf{Z}^*\mathbf{P}, \quad \forall \mathbf{Z} : \|\mathbf{Z}\| \leq \rho$$

if and only if there exists a $\lambda \geq 0$ *such that*

$$\left[\begin{array}{cc} \mathbf{A}-\lambda\mathbf{Q}^*\mathbf{Q} & -\rho\mathbf{P}^* \\ -\rho\mathbf{P} & \lambda\mathbf{I} \end{array}\right] \geq 0.$$

From Proposition 2 it follows that (35) is satisfied if and only if there exists a $\lambda \geq 0$ such that

$$\left[\begin{array}{ccc} \mathbf{X}-\lambda\mathbf{I} & (\mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}})^* & \mathbf{0} \\ \mathbf{I}-\mathbf{G}\widetilde{\mathbf{H}} & \mathbf{I} & -\rho\mathbf{G} \\ \mathbf{0} & -\rho\mathbf{G}^* & \lambda\mathbf{I} \end{array}\right] \geq 0, \tag{37}$$

so that (25) is equivalent to (33) and (37) which are both LMIs.

Finally, the constraint (24) can be expressed as

$$\mathrm{Tr}(\mathbf{Y}) \leq t-\tau, \tag{38}$$

$$\mathbf{G}\mathbf{C}_w\mathbf{G}^* \leq \mathbf{Y}, \tag{39}$$

which using Lemma 2 is equivalent to

$$\begin{bmatrix} \mathbf{Y} & \mathbf{G} \\ \mathbf{G}^* & \mathbf{C}_w^{-1} \end{bmatrix} \geq 0, \tag{40}$$

completing the proof of the theorem. □

5 Minimax Regret

To improve the performance over the minimax MSE approach, we now consider a competitive approach in which we seek a linear estimator whose performance is as close as possible to that of the optimal estimator for all possible values of $\mathbf{C}_x$ satisfying (6), where we assume, as in Section 3, that $\mathbf{H}$ is completely specified. Thus, instead of choosing a linear estimator to minimize the worst case MSE, we now seek the linear estimator $\hat{\mathbf{x}}$ that minimizes the worst-case regret, so that we partially compensate for the conservative character of the minimax approach.

The regret $\mathcal{R}(\mathbf{C}_x, \mathbf{G})$ is defined as the difference 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

$$\text{MSE}^o = \text{Tr}(\mathbf{C}_x - \mathbf{C}_{xy}\mathbf{C}_y^{-1}\mathbf{C}_{yx}) = \text{Tr}(\mathbf{C}_x) - \text{Tr}(\mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1}\mathbf{H}\mathbf{C}_x). \tag{41}$$

From (4) and (5) we have that $\mathbf{C}_x\mathbf{H}^*(\mathbf{H}\mathbf{C}_x\mathbf{H}^* + \mathbf{C}_w)^{-1} = (\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{C}_x^{-1})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$, so that (41) can be written in the equivalent form,

$$\text{MSE}^o = \text{Tr}\left(\left(\mathbf{I} - (\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{C}_x^{-1})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}\right)\mathbf{C}_x\right) = \text{Tr}\left((\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{C}_x^{-1})^{-1}\right), \tag{42}$$

which will be more convenient for our derivations.

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

$$\min_{\mathbf{G}} \max_{l_i \leq \delta_i \leq u_i} \mathcal{R}(\mathbf{C}_x, \mathbf{G}), \tag{43}$$

where $\mathbf{C}_x$ has an eigendecomposition of the form (10), and

$$\begin{aligned} \mathcal{R}(\mathbf{C}_x, \mathbf{G}) &= E(\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2) - \text{MSE}^o \\ &= \text{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \text{Tr}\left(\mathbf{C}_x(\mathbf{I} - \mathbf{G}\mathbf{H})^*(\mathbf{I} - \mathbf{G}\mathbf{H})\right) - \text{Tr}\left((\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{C}_x^{-1})^{-1}\right). \end{aligned} \tag{44}$$

The linear estimator that minimizes the worst-case regret is given by the following theorem.

Theorem 3 (Minimax regret estimator). *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*

$$\min_{\hat{\mathbf{x}}=\mathbf{G}\mathbf{y}} \max_{l_i \le \delta_i \le u_i} \left\{ E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) - \min_{\hat{\mathbf{x}}=\mathbf{G}(\mathbf{x})\mathbf{y}} E(\|\hat{\mathbf{x}} - \mathbf{x}\|^2) \right\} =$$
$$= \min_{\mathbf{G}} \max_{l_i \le \delta_i \le u_i} \left\{ Tr(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + Tr(\mathbf{C}_x(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H})) - Tr\left((\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{C}_x^{-1})^{-1}\right) \right\}$$

is

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

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

$$c_i = \frac{1}{\lambda_i}\left(1 - \frac{1}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}\right), \tag{45}$$

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

Proof. The proof of Theorem 3 is comprised of three parts. First, we show that the optimal $\mathbf{G}$ minimizing the worst-case regret has the form

$$\mathbf{G} = \mathbf{V}\mathbf{D}\Lambda^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}, \tag{46}$$

for some $m \times m$ matrix $\mathbf{D}$. We then show that $\mathbf{D}$ must be a diagonal matrix. Finally, we show that the diagonal elements c_i of $\mathbf{C} = \mathbf{D}\Lambda^{-1}$ are given by (45).

We begin by showing that the optimal $\mathbf{G}$ has the form given by (46). To this end, note that the regret $\mathcal{R}(\mathbf{C}_x, \mathbf{G})$ of (44) 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 $\mathbf{G}$,

$$\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) = \mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*) + \mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}(\mathbf{I}-\mathbf{P})\mathbf{C}_w^{1/2}\mathbf{G}^*) \ge \mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*) \tag{47}$$

where

$$\mathbf{P} = \mathbf{C}_w^{-1/2}\mathbf{H}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1/2} \tag{48}$$

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 $\mathbf{P}\mathbf{C}_w^{-1/2}\mathbf{H} = \mathbf{C}_w^{-1/2}\mathbf{H}$. Thus, to minimize $\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$ it is sufficient to consider matrices $\mathbf{G}$ that satisfy

$$\mathbf{G}\mathbf{C}_w^{1/2} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}. \tag{49}$$

Substituting (48) into (49), we have

$$\mathbf{G} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2} = \mathbf{G}\mathbf{H}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1} = \mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}, \tag{50}$$

for some $m \times m$ matrix $\mathbf{B}$. Denoting $\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^*$ and using the fact that $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Lambda\mathbf{V}^*$, (50) reduces to (46).

We now show that $\mathbf{D}$ must be a diagonal matrix. Substituting $\mathbf{C}_x = \mathbf{V}\Delta\mathbf{V}^*$ and $\mathbf{G}$ of (50) into (44), we can express $\mathcal{R}(\mathbf{C}_x, \mathbf{G})$ as

$$\begin{aligned}
&\mathcal{R}(\mathbf{C}_x, \mathbf{G}) = \\
&\quad = \mathrm{Tr}\left(\mathbf{D}^*\mathbf{D}\mathbf{V}^*(\mathbf{H}^*\mathbf{C}_w\mathbf{H})^{-1}\mathbf{V}\right) + \mathrm{Tr}\left(\mathbf{V}^*\mathbf{C}_x\mathbf{V}(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\right) - \mathrm{Tr}\left((\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{C}_x^{-1})^{-1}\right) \\
&\quad = \mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Lambda^{-1}) + \mathrm{Tr}\left(\Delta(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\right) - \mathrm{Tr}\left((\Lambda + \Delta^{-1})^{-1}\right). \qquad (51)
\end{aligned}$$

We conclude that the problem (43) reduces to finding $\mathbf{D}$ that minimizes

$$\mathcal{G}(\mathbf{D}) = \max_{l_i \le \delta_i \le u_i} \mathcal{L}(\mathbf{D}, \Delta) \qquad (52)$$

where

$$\mathcal{L}(\mathbf{D}, \Delta) = \mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Lambda^{-1}) + \mathrm{Tr}\left(\Delta(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\right) - \mathrm{Tr}\left((\Lambda + \Delta^{-1})^{-1}\right). \qquad (53)$$

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

$$\begin{aligned}
\mathcal{G}(\alpha\mathbf{D}_1 + (1-\alpha)\mathbf{D}_2) &= \max_{l_i \le \delta_i \le u_i} \mathcal{L}(\alpha\mathbf{D}_1 + (1-\alpha)\mathbf{D}_2, \Delta) \\
&< \max_{l_i \le \delta_i \le u_i} \{\alpha\mathcal{L}(\mathbf{D}_1, \Delta) + (1-\alpha)\mathcal{L}(\mathbf{D}_2, \Delta)\} \\
&\le \alpha \max_{l_i \le \delta_i \le u_i} \mathcal{L}(\mathbf{D}_1, \Delta) + (1-\alpha) \max_{l_i \le \delta_i \le u_i} \mathcal{L}(\mathbf{D}_2, \Delta) \\
&= \alpha\mathcal{G}(\mathbf{D}_1) + (1-\alpha)\mathcal{G}(\mathbf{D}_2), \qquad (54)
\end{aligned}$$

so that $\mathcal{G}(\mathbf{D})$ is also strictly convex in $\mathbf{D}$, and consequently has a unique global minimum. Let $\mathbf{J}$ be any diagonal matrix with diagonal elements equal to ± 1. Then

$$\begin{aligned}
\mathcal{G}(\mathbf{JDJ}) &= \max_{l_i \le \delta_i \le u_i} \left\{\mathrm{Tr}(\mathbf{JD}^*\mathbf{DJ}\Lambda^{-1}) + \mathrm{Tr}\left(\Delta(\mathbf{I}-\mathbf{JDJ})^*(\mathbf{I}-\mathbf{JDJ})\right) - \mathrm{Tr}\left((\Lambda + \Delta^{-1})^{-1}\right)\right\} \\
&= \max_{l_i \le \delta_i \le u_i} \left\{\mathrm{Tr}(\mathbf{D}^*\mathbf{DJ}\Lambda^{-1}\mathbf{J}) + \mathrm{Tr}\left(\mathbf{J}\Delta\mathbf{J}(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\right) - \mathrm{Tr}\left((\Lambda + \Delta^{-1})^{-1}\right)\right\} \\
&= \max_{l_i \le \delta_i \le u_i} \left\{\mathrm{Tr}(\mathbf{D}^*\mathbf{D}\Lambda^{-1}) + \mathrm{Tr}\left(\Delta(\mathbf{I}-\mathbf{D})^*(\mathbf{I}-\mathbf{D})\right) - \mathrm{Tr}\left((\Lambda + \Delta^{-1})^{-1}\right)\right\} \\
&= \mathcal{G}(\mathbf{D}), \qquad (55)
\end{aligned}$$

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

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

$$= \sum_{i=1}^{m} \max_{l_i \le \delta_i \le u_i} \left\{ \frac{(\lambda_i(d_i-1)\delta_i + d_i)^2}{\lambda_i(\lambda_i\delta_i+1)} \right\}. \tag{56}$$

The problem of minimizing $\mathcal{G}(\mathbf{D})$ can now be formulated as

$$\min_{t_i, d_i} \sum_{i=1}^{m} t_i \tag{57}$$

subject to

$$\max_{l_i \le \delta_i \le u_i} \left\{ \frac{(\lambda_i(d_i-1)\delta_i + d_i)^2}{\lambda_i(\lambda_i\delta_i+1)} \right\} \le t_i, \quad 1 \le i \le m, \tag{58}$$

which can be separated into m independent problems of the form

$$\min_{t,d} t \tag{59}$$

subject to

$$\max_{l_i \le \delta_i \le u_i} \left\{ \frac{(\lambda_i(d-1)\delta_i + d)^2}{\lambda_i(\lambda_i\delta_i+1)} \right\} \le t, \tag{60}$$

or, equivalently,

$$\frac{(\lambda_i(d-1)\delta_i + d)^2}{\lambda_i(\lambda_i\delta_i+1)} \le t, \quad \forall \delta_i : l_i \le \delta_i \le u_i. \tag{61}$$

To develop a solution to (57) and (58), we thus consider the problem of (59) and (61), 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 (61) can be written as

$$(\lambda(d-1)(\zeta+\epsilon)+d)^2 \le t\lambda(\lambda(\zeta+\epsilon)+1), \quad \forall e : e^2 \le \epsilon^2, \tag{62}$$

which in turn is equivalent to the following implication:

$$P(e) \stackrel{\triangle}{=} \epsilon^2 - e^2 \ge 0 \Rightarrow Q(e) \ge 0, \tag{63}$$

where

$$\begin{aligned} Q(e) &= t\lambda(\lambda(\zeta+e)+1) - (\lambda(d-1)(\zeta+e)+d)^2 \\ &= -e^2\lambda^2(d-1)^2 + 2e\left(\frac{t\lambda^2}{2} + \lambda(1-d)\left(d(\lambda\zeta+1)-\lambda\zeta\right)\right) + t\lambda(\lambda\zeta+1) - (d(\lambda\zeta+1)-\lambda\zeta)^2. \end{aligned} \tag{64}$$

We now rely on the following lemma [28, p. 23]:

Lemma 3. *[$\mathcal{S}$-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 $\mathbf{z}_0$ satisfying $P(\mathbf{z}_0) > 0$. Then the implication*

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

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 (63) with Lemma 3, it follows immediately that the condition (61) is equivalent to the condition

$$\begin{bmatrix} \alpha - \lambda^2(d-1)^2 & \frac{t\lambda^2}{2} + \lambda(1-d)\left(d(\lambda\zeta+1) - \lambda\zeta\right) \\ \frac{t\lambda^2}{2} + \lambda(1-d)\left(d(\lambda\zeta+1) - \lambda\zeta\right) & t\lambda(\lambda\zeta+1) - (d(\lambda\zeta+1) - \lambda\zeta)^2 - \alpha\epsilon^2 \end{bmatrix} \geq 0. \tag{65}$$

Note that if (65) is satisfied, then $\alpha - \lambda^2(d-1)^2 \geq 0$ which implies that $\alpha \geq 0$. Therefore, the problem of (59) and (61) is equivalent to minimizing t subject to (65).

To satisfy (65) we must have that

$$t\lambda(1+\lambda\zeta) \geq (d(\lambda\zeta+1) - \lambda\zeta)^2 + \alpha\epsilon^2, \tag{66}$$

from which it follows that $t \geq 0$. If $t = 0$ then from (66), $\alpha = 0$ and $d = \lambda\zeta/(\lambda\zeta+1)$. But then

$$\alpha - \lambda^2(d-1)^2 < 0 \tag{67}$$

which violates (65) so that $t > 0$. Defining $\beta = \alpha/t$, our problem reduces to

$$\min_{t,\beta,d} t \tag{68}$$

subject to

$$\begin{bmatrix} t\beta - \lambda^2(d-1)^2 & \frac{t\lambda^2}{2} + \lambda(1-d)\left(d(\lambda\zeta+1) - \lambda\zeta\right) \\ \frac{t\lambda^2}{2} + \lambda(1-d)\left(d(\lambda\zeta+1) - \lambda\zeta\right) & t\lambda(\lambda\zeta+1) - (d(\lambda\zeta+1) - \lambda\zeta)^2 - t\beta\epsilon^2 \end{bmatrix}, \tag{69}$$

which can be expressed as

$$t\mathbf{A} - \mathbf{b}\mathbf{b}^* \geq 0, \tag{70}$$

where

$$\mathbf{A} = \begin{bmatrix} \beta & \frac{\lambda^2}{2} \\ \frac{\lambda^2}{2} & \lambda(1+\lambda\zeta) - \beta\epsilon^2 \end{bmatrix}, \tag{71}$$

and

$$\mathbf{b} = \begin{bmatrix} \lambda(d-1) \\ d(1+\lambda\zeta) - \lambda\zeta \end{bmatrix}. \tag{72}$$

Suppose there exists a $\hat{t}$ such that

$$\hat{t}\mathbf{A} - \mathbf{b}\mathbf{b}^* = 0. \tag{73}$$

Then, as we now show, $\hat{t}$ minimizes (68) subject to (70). Since $\hat{t} \geq 0$ and from (73) $\hat{t}\mathbf{A} \geq 0$, if follows that

$\mathbf{A} \geq 0$. If $0 \leq t < \hat{t}$, then $\hat{t} - t > 0$ so that $(\hat{t} - t)\mathbf{A} \geq 0$, which implies that

$$t\mathbf{A} \leq \hat{t}\mathbf{A} = \mathbf{b}\mathbf{b}^*, \tag{74}$$

or

$$t\mathbf{A} - \mathbf{b}\mathbf{b}^* \leq 0. \tag{75}$$

Thus, either we have $t\mathbf{A} = \mathbf{b}\mathbf{b}^*$ or at least one of the eigenvalues of $t\mathbf{A} - \mathbf{b}\mathbf{b}^*$ is smaller than 0, which violates the constraint (70). Therefore, if there exists $\hat{t}, \hat{\beta}$ and $\hat{d}$ that solve (73), or equivalently the equations

$$\begin{aligned} \hat{t}\hat{\beta} &= \lambda^2(1-\hat{d})^2; \\ \hat{t}\left(\lambda(1+\lambda\zeta) - \hat{\beta}\epsilon^2\right) &= (\hat{d}(1+\lambda\zeta) - \lambda\zeta)^2; \\ \hat{t}\lambda &= 2(\hat{d}-1)\left(\hat{d}(1+\lambda\zeta) - \lambda\zeta\right), \end{aligned} \tag{76}$$

then the optimal value of t must also satisfy (76) for some β and d. It can be shown that (76) has a *unique* solution

$$\begin{aligned} \hat{d} &= 1 - \tfrac{1}{\sqrt{(1+\lambda\zeta)^2 - \lambda^2\epsilon^2}}; \\ \hat{\beta} &= \tfrac{\lambda^3}{2\left(\sqrt{(1+\lambda\zeta)^2 - \lambda^2\epsilon^2} - 1 - \lambda\zeta\right)}; \\ \hat{t} &= \tfrac{2\left(\sqrt{(1+\lambda\zeta)^2 - \lambda^2\epsilon^2} - 1 - \lambda\zeta\right)}{\lambda((1+\lambda\zeta)^2 - \lambda^2\epsilon^2)}, \end{aligned} \tag{77}$$

which is therefore the solution to the problem of (59) and (61).

The linear minimax estimator is therefore given by

$$\hat{\mathbf{x}} = \mathbf{V}\widehat{\mathbf{D}}\Lambda^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} = \mathbf{V}\mathbf{C}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}, \tag{78}$$

where $\mathbf{C} = \widehat{\mathbf{D}}\Lambda^{-1}$ is the diagonal matrix with diagonal elements $\hat{d}_i/\lambda_i$ and

$$\hat{d}_i = 1 - \frac{1}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}, \quad 1 \leq i \leq m, \tag{79}$$

which completes the proof of the theorem. □

As we now show, we can interpret the estimator of Theorem 3 as an MMSE estimator matched to a covariance matrix

$$\mathbf{C}_x = \mathbf{V}\mathbf{X}\mathbf{V}^*, \tag{80}$$

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

$$x_i = \frac{1}{\lambda_i}\left(\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2} - 1\right). \tag{81}$$

Note that if $\epsilon_i = 0$ so that the ith eigenvalue of the true covariance of $\mathbf{C}_x$ is equal to ζ_i then, as we expect, $x_i = \zeta_i$.

From (5) the MMSE estimate of $\mathbf{x}$ with covariance $\mathbf{C}_x$ given by (80) and $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} = \mathbf{V}\Lambda\mathbf{V}^*$ is

$$\hat{\mathbf{x}} = (\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{C}_x^{-1})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} = \mathbf{V}\left(\Lambda + \mathbf{X}^{-1}\right)^{-1}\mathbf{V}^*\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \tag{82}$$

Since

$$\frac{1}{\lambda_i + \frac{1}{x_i}} = \frac{x_i}{x_i\lambda_i + 1} = \frac{1}{\lambda_i}\left(1 - \frac{1}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}\right) = c_i, \tag{83}$$

the estimator $\hat{\mathbf{x}}$ of (82) is equivalent to the estimator given by Theorem 3. We thus have the following corollary to Theorem 3.

Corollary 4. *Let* $\mathbf{x}$ *denote the unknown parameters in the model* $\mathbf{y} = \mathbf{H}\mathbf{x}+\mathbf{w}$*. Then, under the assumptions of Theorem 3, the solution to the problem*

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

is an MMSE estimator matched to the covariance $\mathbf{C}_x = \mathbf{V}\mathbf{X}\mathbf{V}^*$*, where* $\mathbf{X}$ *is a diagonal matrix with diagonal elements*

$$x_i = \frac{1}{\lambda_i}\left(\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2} - 1\right)$$

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

Since the minimax regret estimator minimizes the regret 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 regret sense.

It is interesting to note that while the minimax MSE estimator of Theorem 1 for the model (6) is matched to a covariance matrix with eigenvalues $u_i \ge \zeta_i$, the minimax regret estimator of Theorem 3 is matched to a covariance matrix with eigenvalues $x_i \le \zeta_i$. Indeed, from (81) we have that

$$x_i \le \frac{\sqrt{(\lambda_i\zeta_i + 1)^2} - 1}{\lambda_i} = \zeta_i. \tag{84}$$

Expressing x_i as

$$x_i = \frac{1}{\lambda_i}\left((1+\lambda_i\zeta_i)\sqrt{1 - \frac{\lambda_i^2\epsilon_i^2}{(1+\lambda_i\zeta_i)^2}} - 1\right) = \frac{1}{\lambda_i}\left((1+\lambda_i\zeta_i)\sqrt{1-a_i} - 1\right), \tag{85}$$

where

$$a_i = \frac{\lambda_i^2\epsilon_i^2}{(1+\lambda_i\zeta_i)^2} < 1, \tag{86}$$

(since $\zeta_i \ge \epsilon_i$) and using the first order approximation $\sqrt{1-y} \approx 1 - (1/2)y$ for $0 \le y < 1$, we have that

$$x_i \approx \zeta_i - \frac{\lambda_i\epsilon_i^2}{2(1+\lambda_i\zeta_i)}. \tag{87}$$

Thus, the correction to the nominal covariance ζ_i is approximately $\lambda_i \epsilon_i^2/(2(1+\lambda_i\zeta_i))$, which is quadratic in the length of the uncertainty interval ϵ_i.

The minimax estimators for the uncertainty model (7) often lie in a different class than the estimator matched to the nominal covariance matrix. For example, suppose that $\lambda_i = \lambda, 1 \le i \le m$ and $\delta_i = \delta, 1 \le i \le m$ so that both $\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and the nominal covariance matrix of $\mathbf{x}$ are proportional to the identity. In this case, the MMSE estimator matched to the nominal covariance matrix is $\hat{\mathbf{x}} = a\mathbf{y}$ for some constant a so that $\hat{\mathbf{x}}$ is simply a scaled version of $\mathbf{y}$. This property also holds for the minimax MSE estimator of Theorem 1 with covariance uncertainty given by (8). However, for the minimax regret estimator and the minimax MSE estimator with covariance uncertainty given by (7), this property no longer holds in general. In particular, if $\epsilon_i \neq \epsilon_j$ for some i and j, then the optimal estimators will no longer be a scaled version of $\mathbf{y}$.

6 Example of the Minimax Regret Estimator

We now consider an example illustrating the minimax regret estimator of Theorem 3.

Consider the estimation problem in which

$$\mathbf{y} = \mathbf{x} + \mathbf{w}, \tag{88}$$

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

$$E(x_i x_j) = \rho^{|j-i|} \tag{89}$$

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 (88) 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

$$\widehat{\mathbf{C}}_x = [\widehat{\mathbf{C}}_y - \mathbf{C}_w]_+ = [\widehat{\mathbf{C}}_y - \sigma^2\mathbf{I}]_+, \tag{90}$$

where

$$\widehat{\mathbf{C}}_y(i,j) = \frac{1}{n}\sum_{k=1}^{n-|j-i|} y_k y_{k+|j-i|} \tag{91}$$

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 (90) can be regarded as the analogue for finite-length processes of the spectrum estimate based on the spectral subtraction method for infinite-length processes [29, 30].

Given $\widehat{\mathbf{C}}_x$, we may estimate $\mathbf{x}$ using an MMSE estimate matched to $\widehat{\mathbf{C}}_x$. However, as can be seen below in Fig. 1, we can further improve the estimation performance by using the minimax regret estimator of Theorem 3.

To compute the minimax regret 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

$$\tilde{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \sigma_w^2. \tag{92}$$

We further assume that $\mathbf{x}$ and $\mathbf{w}$ are uncorrelated Gaussian random vectors. The variance of $\tilde{\sigma}_x^2$ is given by

$$E\left\{\left(\tilde{\sigma}_x^2 - \sigma_x^2\right)^2\right\} = E\left\{\left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i^2 - \sigma_w^2 - \sigma_x^2\right)\right)^2\right\} = \frac{1}{n^2}\sum_{i,j=1}^{n} E(t_i t_j), \tag{93}$$

where $t_i = y_i^2 - \sigma_w^2 - \sigma_x^2$. Since $E(y_i^2) = \sigma_w^2 + \sigma_x^2$,

$$E(t_i t_j) = (y_i^2 - \sigma_w^2 - \sigma_x^2)(y_j^2 - \sigma_w^2 - \sigma_x^2) = E(y_i^2 y_j^2) - (\sigma_w^2 + \sigma_x^2)^2. \tag{94}$$

If $\mathbf{x}$ and $\mathbf{w}$ are Gaussian, then so is $\mathbf{y}$ so that

$$E(y_i^2 y_j^2) = E^2(y_i y_j) + E(y_i^2)E(y_j^2) = (\mathbf{C}_x(i,j) + \sigma_w^2\delta_{ij})^2 + (\sigma_w^2 + \sigma_x^2)^2, \tag{95}$$

where $\mathbf{C}_x(i,j)$ is the ijth element of $\mathbf{C}_x$. Combining (93), (94) and (95), we have

$$E\left\{\left(\tilde{\sigma}_x^2 - \sigma_x^2\right)^2\right\} = \frac{2}{n}\left((\sigma_x^2 + \sigma^2)^2 + \sum_{i=2}^{n}\mathbf{C}_x^2(1,i)\right). \tag{96}$$

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

$$\epsilon_i = \min\left(\zeta_i, A\sqrt{\frac{2}{n}\left((\widehat{\mathbf{C}}_x^2(1,1) + \sigma^2)^2 + \sum_{i=2}^{n}\widehat{\mathbf{C}}_x^2(1,i)\right)}\right), \tag{97}$$

where A is a proportionality factor.

In Fig. 1, we plot the MSE of the minimax regret 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 $A = 5$. The performance of the "plug in" MMSE estimator matched to the estimated covariance matrix $\widehat{\mathbf{C}}_x$ and the minimax MSE estimator are plotted for comparison. As can be seen from the figure, the minimax regret estimator can increase the estimation performance particularly at low to intermediate SNR values. It is also interesting to note that the popular minimax MSE approach is useless in this example, since it leads to an estimator whose performance is worse than the performance of an estimator based on the estimated covariance matrix.

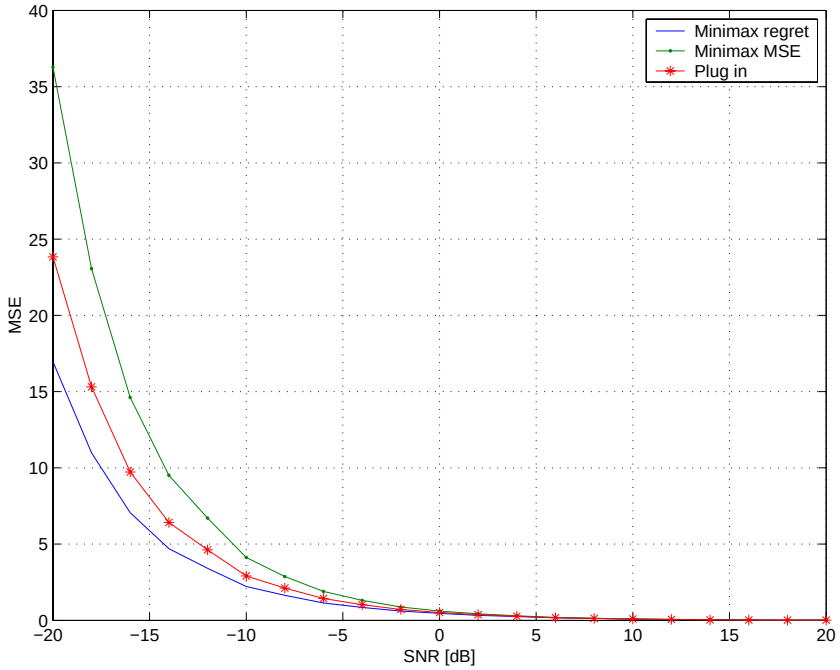

Figure 1: MSE in estimating $\mathbf{x}$ as a function of SNR using the minimax regret estimator, the minimax MSE estimator and the MMSE estimator matched to the estimated covariance matrix.

7 Nonlinear Minimax Regret Estimation

In the previous sections we developed *linear* estimators for estimating the unknown vector $\mathbf{x}$ in the linear model (1) when the covariance $\mathbf{C}_x$ is not known precisely. The restriction to linear estimators was made for analytical tractability since developing the optimal nonlinear estimator is a difficult problem. If $\mathbf{x}$ and $\mathbf{w}$ are jointly Gaussian vectors with known covariance matrices, then the estimator that minimizes the MSE among all linear and nonlinear estimators is the linear MMSE estimator, which provides theoretical justification for restricting attention to the class of linear estimators. As we now show, this property of the optimal estimator is no longer true when we consider minimizing the worst-case regret with covariance uncertainties, even if $\mathbf{x}$ and $\mathbf{w}$ are Gaussian. Nonetheless, we will demonstrate that when estimating a Gaussian random variable contaminated by independent Gaussian noise, the performance of the linear minimax regret estimator is close to that of the optimal nonlinear third-order estimator that minimizes the worst-case regret, so that at least in this case, we do not loose much by restricting our attention to linear estimators.

For the sake of simplicity, we now consider the problem of estimating the scalar x in the linear model

$$y = x + w, \tag{98}$$

where x and w are independent, zero-mean, Gaussian random variables with variances σ_x^2 and σ_w^2, respectively. We seek the possibly nonlinear estimator $\hat{x}$ of x that minimizes the worst-case regret over all variances σ_x^2 satisfying $l \leq \sigma_x^2 \leq u$ for some $0 \leq l \leq u$. In the case of model (98) the linear MMSE estimator is given

by

$$\hat{x} = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_w^2} y = \alpha y, \tag{99}$$

where

$$\alpha = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_w^2}, \tag{100}$$

and the optimal MSE is

$$\text{MSE}^o = E(\alpha y - x)^2 = \frac{\sigma_x^2 \sigma_w^2}{\sigma_x^2 + \sigma_w^2}. \tag{101}$$

Therefore, our problem reduces to finding

$$R = \min_{\hat{x}} \max_{l \le \sigma_x^2 \le u} \left\{ E(\hat{x} - x)^2 - \frac{\sigma_x^2 \sigma_w^2}{\sigma_x^2 + \sigma_w^2} \right\}. \tag{102}$$

Since x and w are jointly Gaussian, $E(x|y) = \alpha y$ with α given by (100), so that (102) can be expressed as

$$\begin{aligned} E(\hat{x} - x)^2 - \frac{\sigma_x^2 \sigma_w^2}{\sigma_x^2 + \sigma_w^2} &= \int_{-\infty}^{\infty} p_y(y) dy \int_{\infty}^{\infty} p_{x|y}(x|y) \left((\hat{x} - x)^2 - (\alpha y - x)^2 \right) dx \\ &= \int_{-\infty}^{\infty} p_y(y) \left(\hat{x}^2 - 2\hat{x} E(x|y) + 2\alpha y E(x|y) - \alpha^2 y^2 \right) dy \\ &= \int_{-\infty}^{\infty} p_y(y) \left(\hat{x} - \alpha y \right)^2 dy, \end{aligned} \tag{103}$$

where $p_x(x) = \mathcal{N}(0, \sigma_x^2)$ and $p_y(y) = \mathcal{N}(0, \sigma_x^2 + \sigma_w^2)$ denote the probability density functions (pdfs) of x and y respectively. Since there is a one-to-one correspondence between σ_x^2 and α, instead of maximizing (103) over $l \le \sigma_x^2 \le u$, we may maximize it over $l_\alpha \le \alpha \le u_\alpha$ where

$$\begin{aligned} l_\alpha &= \frac{l}{l + \sigma_w^2}; \\ u_\alpha &= \frac{u}{u + \sigma_w^2}, \end{aligned} \tag{104}$$

with $u_\alpha \le 1$. Thus,

$$R = \min_{\hat{x}} \max_{l_\alpha \le \alpha \le u_\alpha} \int_{-\infty}^{\infty} p_{y|\alpha}(y|\alpha) \left(\hat{x} - \alpha y \right)^2 dy. \tag{105}$$

Here

$$p_{y|\alpha}(y|\alpha) = \mathcal{N}\left(0, \frac{\sigma_w^2}{1 - \alpha} \right), \tag{106}$$

is the pdf of y given the value of α. We now note that instead of maximizing the objective in (105) over α, we can imagine that α is a random variable with pdf $p_\alpha(\alpha)$ which has support on the interval $\mathcal{I} = [l_\alpha, u_\alpha]$, and maximize the objective over all possible pdfs $p_\alpha(\alpha)$ with support on $\mathcal{I}$. This follows from the fact that the objective will be maximized for the pdf $p_\alpha(\alpha) = \delta(\alpha_0)$ where $\alpha_0 \in \mathcal{I}$ maximizes the objective over $l_\alpha \le \alpha \le u_\alpha$. We then have that

$$R = \min_{\hat{x}} \max_{p_\alpha(\cdot)} \int_{l_\alpha}^{u_\alpha} p_\alpha(\alpha) d\alpha \int_{-\infty}^{\infty} p_{y|\alpha}(y|\alpha) \left(\hat{x} - \alpha y \right)^2 dy. \tag{107}$$

Since the objective in (107) is convex in the minimization argument $\hat{x}$ and concave (linear) in the maximization argument $p_\alpha(\cdot)$, we can exchange the order of the minimization and maximization [31] so that

$$\begin{aligned} R &= \max_{p_\alpha(\cdot)} \min_{\hat{x}} \int_{l_\alpha}^{u_\alpha} p_\alpha(\alpha) d\alpha \int_{-\infty}^{\infty} p_{y|\alpha}(y|\alpha) \left(\hat{x} - \alpha y\right)^2 dy \\ &= \max_{p_\alpha(\cdot)} \left\{ \int_{-\infty}^{\infty} p_y(y) dy \min_{\hat{x}} \left(\int_{l_\alpha}^{u_\alpha} p_{\alpha|y}(\alpha|y) \left(\hat{x} - \alpha y\right)^2 d\alpha \right) \right\}, \end{aligned} \tag{108}$$

where $p_{\alpha|y}(\alpha|y)$ is the conditional probability of α given y induced by $p_\alpha(\alpha)$.

Differentiating the second integral with respect to $\hat{x}$ and equating to 0, the optimal $\hat{x}$ that minimizes (108) is

$$\hat{x} = yE(\alpha|y), \tag{109}$$

where $p_{y,\alpha}(y,\alpha) = p_{y|\alpha}(y|\alpha)p_\alpha(\alpha)$ with $p_{y|\alpha}(y|\alpha)$ given by (106) and

$$p_\alpha(\alpha) = \arg\max_{p_\alpha(\cdot)} \int_{-\infty}^{\infty} y^2 \operatorname{Var}\{\alpha|y\} p_y(y) dy, \tag{110}$$

with $\operatorname{Var}\{\alpha|y\}$ denoting the variance of α given y. Substituting $\hat{x}$ into (108), the minimax regret is

$$R = \max_{p_\alpha(\cdot)} \int_{-\infty}^{\infty} y^2 \operatorname{Var}\{\alpha|y\} p_y(y) dy. \tag{111}$$

As we now show, (109) implies that the minimax regret estimator is *nonlinear*, even though x and w are jointly Gaussian. Therefore, contrary to the MMSE estimator for the Gaussian case where σ_x^2 is known, the estimator minimizing the worst-case regret when σ_x^2 is unknown is nonlinear. Nonetheless, as we show below, in practice we do not loose much by restricting the estimator to be linear.

To show that (109) implies that $\hat{x}$ must be nonlinear in y, we note that since

$$p_{y|\alpha}(y|\alpha) \propto e^{-(1-\alpha)\frac{y^2}{2\sigma_w^2}}, \tag{112}$$

we can express $E(\alpha|y)$ as

$$E(\alpha|y) = \frac{\int_{\mathcal{I}} \alpha p_{y,\alpha}(y,\alpha) d\alpha}{\int_{\mathcal{I}} p_{y,\alpha}(y,\alpha) d\alpha} = \frac{\int_{\mathcal{I}} \alpha p_\alpha(\alpha) e^{\alpha y^2/(2\sigma_w^2)} d\alpha}{\int_{\mathcal{I}} p_\alpha(\alpha) e^{\alpha y^2/(2\sigma_w^2)} d\alpha} = \frac{d}{dz}\left\{ \ln \int_{\mathcal{I}} p_\alpha(\alpha) e^{\alpha z} d\alpha \right\} = \frac{d}{dz} \ln \phi(z), \tag{113}$$

where $z \stackrel{\triangle}{=} y^2/(2\sigma_w^2)$, $\phi(z)$ is the moment generating function of $p_\alpha(\alpha)$, and $\mathcal{I}$ denotes the support of $p_\alpha(\alpha)$.

It is immediate from (109) that $\hat{x}$ is linear if and only if $E(\alpha|y) = a$ for some constant a. This then implies from (113) that the derivative of $\ln \phi(z)$ must be equal to a constant, independent of z, which in turn implies that $\phi(z) = e^{az}$ (since $\phi(0) = 1$ for any moment generating function). Since $p_\alpha(\alpha)$ is the inverse Fourier transform of $\phi(-j\omega)$, in this case $p_\alpha(\alpha) = \delta(\alpha - a)$, and $\operatorname{Var}\{\alpha|y\} = 0$ so that from (111), the regret $R = 0$. Clearly, there are other choices of $p_\alpha(\alpha)$ for which $R > 0$ so that $p_\alpha(\alpha) = \delta(\alpha - a)$ does not maximize the regret, and $E(\alpha|y)$ cannot be equal to a constant.

In order to obtain an explicit expression for the minimax regret estimator of (109), we need to determine

the optimal pdf $p_\alpha(\alpha)$, which is a difficult problem. Since $E(\alpha|y)$ is the MMSE estimator of α given the random variable y, we may approximate $E(\alpha|y)$ by a linear estimator of α of the form $\hat{\alpha} = a + by$ for some a and $b \neq 0$ (we have seen already that $E(\alpha|y)$ cannot be equal to a constant). With this approximation,

$$\hat{x} = ay + by^2, \tag{114}$$

where a and b are the solution to

$$\min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} \int_{-\infty}^{\infty} p_y(y) \left(\hat{x} - \alpha y\right)^2 dy = \min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} E\left(\left(\hat{x} - \alpha y\right)^2\right). \tag{115}$$

Substituting (114) into (115), a and b are the solution to the problem

$$\min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} E\left(\left((a-\alpha)y + by^2\right)\right)^2 = \min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} \left\{(\alpha - a)^2 E(y^2) + b^2 E(y^4)\right\}, \tag{116}$$

where we used the fact that since y is Gaussian, $E(y^3) = 0$. Now, for any choice of a and b, $(\alpha - a)^2 E(y^2) + b^2 E(y^4) \geq (\alpha - a)^2 E(y^2)$ so that

$$\min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} \left\{(\alpha - a)^2 E(y^2) + b^2 E(y^4)\right\} \geq \min_{a} \max_{l_\alpha \leq \alpha \leq u_\alpha} (\alpha - a)^2 E(y^2), \tag{117}$$

with equality for $b = 0$. Thus the optimal estimator of the form (114) reduces to a linear estimator, which cannot be optimal.

Since the second-order approximation (114) results in a linear estimator, we next consider a third-order approximation of the form

$$\hat{x} = ay + by^3, \tag{118}$$

where now a and b are the solution to

$$\begin{aligned}
&\min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} E\left(\left((a-\alpha)y + by^3\right)^2\right) = \\
&\quad = \min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} \left\{(\alpha - a)^2 E(y^2) + b^2 E(y^6) + 2(a-\alpha) b E(y^4)\right\} \\
&\quad = \min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} \left\{(\alpha - a)^2 \frac{\sigma_w^2}{1-\alpha} + 15 b^2 \frac{\sigma_w^6}{(1-\alpha)^3} + 6(a-\alpha) b \frac{\sigma_w^4}{(1-\alpha)^2}\right\}.
\end{aligned} \tag{119}$$

Here we used the fact that y is a zero-mean Gaussian random variable so that [14]

$$E(y^n) = \begin{cases} 1 \cdot 3 \cdot \ldots \cdot (n-1)\sigma_y^n, & n \text{ even;} \\ 0, & n \text{ odd,} \end{cases} \tag{120}$$

where $\sigma_y^2 = \sigma_w^2/(1-\alpha)$ is the variance of y.

Finding the optimal values of a and b that are the solution to (119) is a difficult problem. Instead of solving this problem directly, we develop a lower bound on the minimax regret R achievable with a third-order nonlinear estimator of the from (118), and show that in many cases it is approximately achieved by

the linear minimax regret estimator of Theorem 3. In particular, we have the following theorem, the proof of which is provided in Appendix B.

Theorem 5. *Let x denote a zero-mean Gaussian random variable with variance σ_x^2 and let $y = x + w$ where w denotes a zero-mean Gaussian random variable with variance σ_w^2, independent of x. Let $\hat{x} = ay + by^3$ be a third-order estimator of x where a and b minimize the worst-case regret over all values of $\alpha = \sigma_x^2/(\sigma_x^2 + \sigma_w^2)$ satisfying $l_\alpha \leq \alpha \leq u_\alpha$ for some $0 \leq l_\alpha \leq u_\alpha$. Then the minimax regret given by*

$$R = \min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} E\left\{ \left[(a-\alpha)y + by^3 \right]^2 \right\},$$

satisfies $R \geq B$ where B is the solution to the convex optimization problem

$$B = \min_{\beta \geq 0, a, b, \tau, \gamma} \left\{ \frac{\tau^2}{1 - l_\alpha} + 6b^2 \frac{\sigma_w^6}{(1-l_\alpha)^3} \right\}$$

subject to

$$\begin{bmatrix} \gamma - \sigma_w & \frac{1}{2}\left((a - 2\zeta + 1)\sigma_w - \tau\right) \\ \frac{1}{2}\left((a - 2\zeta + 1)\sigma_w - \tau\right) & \tau(1-\zeta) - (1-\zeta)(a-\zeta)\sigma_w - 3b\sigma_w^3 - \gamma\epsilon^2 \end{bmatrix} \geq 0; \tag{121}$$

$$\begin{bmatrix} \beta + \sigma_w & -\frac{1}{2}\left((a - 2\zeta + 1)\sigma_w + \tau\right) \\ -\frac{1}{2}\left((a - 2\zeta + 1)\sigma_w + \tau\right) & \tau(1-\zeta) + (1-\zeta)(a-\zeta)\sigma_w + 3b\sigma_w^3 - \beta\epsilon^2 \end{bmatrix} \geq 0, \tag{122}$$

where $\zeta = (u_\alpha + l_\alpha)/2$ and $\epsilon = (u_\alpha - l_\alpha)/2$.

Note that a positive semidefinite constraint of the form

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \geq 0, \tag{123}$$

is equivalent to the three inequalities $a \geq 0, c \geq 0$ and $ac - b^2 \geq 0$.

In Fig. 2, we plot the bound B as a function of the SNR which is defined as $10\log(\sigma_x^2/\sigma_w^2)$ for $\sigma_x^2 = u = 50$ and $l = 30$. For comparison, we also plot the worst-case regret using the linear minimax regret estimator of Theorem 3. The value of B is computed using the `fmincon` function on Matlab. As can be seen from the figure, the worst-case regret using the linear minimax estimator is very close to the bound so that in this case we do not loose in performance by using a linear estimator instead of a nonlinear third-order estimator. In general, the performance of the linear minimax regret estimator is close to the bound for small values of $u - l$. If $u \gg l$, then the performance of the linear estimator approaches the bound only at high SNR.

8 Conclusion

We considered the problem of estimating a random vector $\mathbf{x}$ in the linear model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where the covariance matrix $\mathbf{C}_x$ of $\mathbf{x}$ and possibly also the model matrix $\mathbf{H}$ are subject to uncertainties. We developed

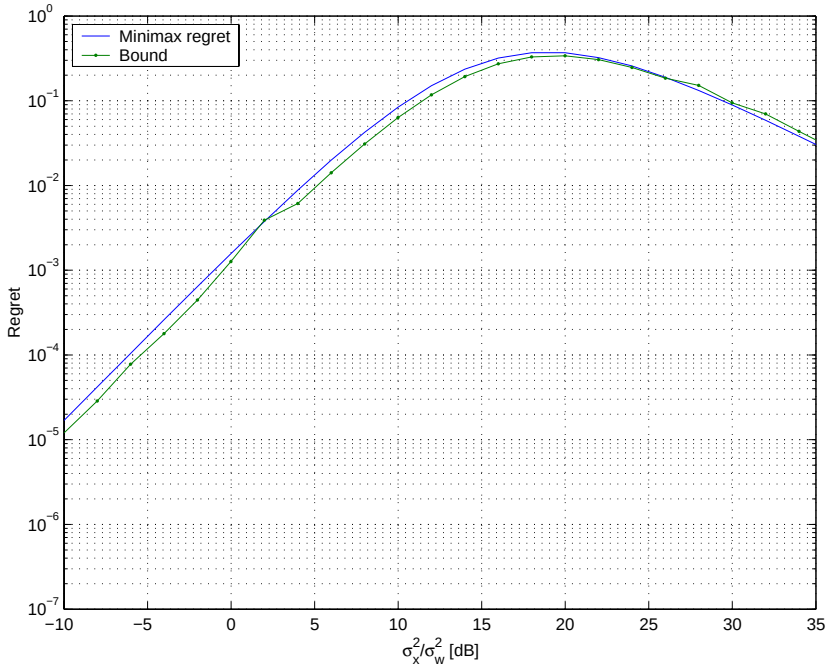

Figure 2: Worst-case regret in estimating x as a function of SNR using the linear minimax regret estimator, and the bound on the smallest worst-case regret attainable using a third-order estimator.

the minimax MSE estimators for the case in which $\mathbf{C}_x$ is subject to uncertainties and the model matrix is known, and for the case in which both $\mathbf{C}_x$ and $\mathbf{H}$ are not completely specified.

The main contribution of the paper is the development of a competitive minimax approach in which we seek the linear estimator that minimizes the worst-case regret, which is the difference between the MSE of the estimator and the best possible MSE attainable with a linear estimator that knows the covariance $\mathbf{C}_x$. As we demonstrated, the competitive minimax approach can increase the performance over the traditional minimax method, which in some cases turns out to be completely useless.

The minimax estimator has the interesting property that it often lies in a different class than the estimator matched to the nominal covariance matrix. We have seen an example of this property in Section 5 where the nominal estimator is proportional to the observations $\mathbf{y}$, while the linear minimax regret estimator is no longer equal to a constant times $\mathbf{y}$. Another example was considered in Section 7, where we showed that the optimal minimax regret estimator for the case in which $\mathbf{y}$ and $\mathbf{w}$ are jointly Gaussian is nonlinear, while the nominal estimator is linear.

In our development of the minimax regret, we assumed that $\mathbf{H}$ is completely specified and that $\mathbf{H}\mathbf{C}_w^{-1}\mathbf{H}$ commutes with $\mathbf{C}_x$ for all possible covariance matrices. An interesting direction for future research is to develop the minimax regret estimator for more general classes of $\mathbf{H}$ as well as in the presence of uncertainties in $\mathbf{H}$. It is also interesting to investigate the loss in performance with respect to an arbitrary nonlinear minimax regret estimator in the general linear model.

9 Acknowledgment

The first author wishes to thank Prof. A. Singer for first suggesting to her the use of the regret as a figure of merit in the context of estimation.

A Proof of Proposition 1

Let $\mathbf{X}$ be an arbitrary matrix satisfying (29). Then from Lemma 1,

$$\max_{\|\mathbf{A}\|\leq\alpha,\|\mathbf{B}\|\leq\beta} \mathrm{Tr}(\mathcal{G}(\mathbf{A})\mathcal{Q}(\mathbf{B})) \leq \max_{\|\mathbf{A}\|\leq\alpha,\|\mathbf{B}\|\leq\beta} \mathrm{Tr}(\mathcal{G}(\mathbf{A})\mathbf{X}). \tag{124}$$

Since

$$\mathrm{Tr}(\mathcal{G}(\mathbf{A})\mathbf{X}) = \mathrm{Tr}(\mathcal{G}(\mathbf{A})\mathcal{Q}(\mathbf{B})) + \mathrm{Tr}\left(\mathcal{G}(\mathbf{A})(\mathbf{X}-\mathcal{Q}(\mathbf{B}))\right), \tag{125}$$

we have that

$$\max_{\|\mathbf{A}\|\leq\alpha,\|\mathbf{B}\|\leq\beta} \mathrm{Tr}(\mathcal{G}(\mathbf{A})\mathbf{X}) \leq \max_{\|\mathbf{A}\|\leq\alpha,\|\mathbf{B}\|\leq\beta} \mathrm{Tr}(\mathcal{G}(\mathbf{A})\mathcal{Q}(\mathbf{B})) + \max_{\|\mathbf{A}\|\leq\alpha,\|\mathbf{B}\|\leq\beta} \mathrm{Tr}\left(\mathcal{G}(\mathbf{A})(\mathbf{X}-\mathcal{Q}(\mathbf{B}))\right). \tag{126}$$

Let $\tau_1 = \min \tau$ subject to (27), let $\tau_2 = \min \tau$ subject to (28) and (29), and let $\tau_3 = \min \tau$ subject to

$$\max_{\|\mathbf{A}\|\leq\alpha,\|\mathbf{B}\|\leq\beta} \mathrm{Tr}(\mathcal{G}(\mathbf{A})\mathcal{Q}(\mathbf{B})) + \max_{\|\mathbf{A}\|\leq\alpha,\|\mathbf{B}\|\leq\beta} \mathrm{Tr}\left(\mathcal{G}(\mathbf{A})(\mathbf{X}-\mathcal{Q}(\mathbf{B}))\right) \leq \tau, \tag{127}$$

and (29). It then follows from (124) and (126) that

$$\tau_1 \leq \tau_2 \leq \tau_3. \tag{128}$$

Since $\mathcal{G}(\mathbf{A}) \geq 0$ and $\mathbf{X} - \mathcal{Q}(\mathbf{B}) \geq 0$ for all $\|\mathbf{A}\| \leq \alpha, \|\mathbf{B}\| \leq \beta$ and $\mathbf{X}$ satisfying (29), it follows from Lemma 1 that $\max_{\|\mathbf{A}\|\leq\alpha,\|\mathbf{B}\|\leq\beta} \mathrm{Tr}\left(\mathcal{G}(\mathbf{A})(\mathbf{X}-\mathcal{Q}(\mathbf{B}))\right) \geq 0$. Therefore to minimize the value of τ in (127), $\mathbf{X}$ is chosen such that $\max_{\|\mathbf{A}\|\leq\alpha,\|\mathbf{B}\|\leq\beta} \mathrm{Tr}\left(\mathcal{G}(\mathbf{A})(\mathbf{X}-\mathcal{Q}(\mathbf{B}))\right) = 0$. But then $\tau_3 = \tau_1$ so that from (128) we conclude that $\tau_2 = \tau_1$, completing the proof of the proposition.

B Proof of Theorem 5

From (119) it follows that we can express R as

$$R = \min_{a,b} \max_{l_\alpha \leq \alpha \leq u_\alpha} \mathcal{G}(a,b,\alpha), \tag{129}$$

where

$$\begin{aligned} \mathcal{G}(a,b,\alpha) &= (\alpha-a)^2 \frac{\sigma_w^2}{1-\alpha} + 15b^2 \frac{\sigma_w^6}{(1-\alpha)^3} + 6(a-\alpha)b\frac{\sigma_w^4}{(1-\alpha)^2} \\ &= \frac{1}{1-\alpha}\left((a-\alpha)\sigma_w + \frac{3b\sigma_w^3}{1-\alpha}\right)^2 + 6b^2\frac{\sigma_w^6}{(1-\alpha)^3}. \end{aligned} \tag{130}$$

Since $l_\alpha \le \alpha \le u_\alpha \le 1$,

$$\mathcal{G}(a,b,\alpha) \ge \frac{1}{1-l_\alpha}\left((a-\alpha)\sigma_w + \frac{3b\sigma_w^3}{1-\alpha}\right)^2 + 6b^2\frac{\sigma_w^6}{(1-l_\alpha)^3}, \tag{131}$$

so that $R \ge B$ where

$$B = \min_{a,b}\max_{l_\alpha\le\alpha\le u_\alpha}\left\{\frac{1}{1-l_\alpha}\left((a-\alpha)\sigma_w + \frac{3b\sigma_w^3}{1-\alpha}\right)^2 + 6b^2\frac{\sigma_w^6}{(1-l_\alpha)^3}\right\}. \tag{132}$$

To compute B we note that (132) can be expressed as

$$B = \min_{a,b,t} t \tag{133}$$

subject to

$$\frac{1}{1-l_\alpha}\left((a-\alpha)\sigma_w + \frac{3b\sigma_w^3}{1-\alpha}\right)^2 + 6b^2\frac{\sigma_w^6}{(1-l_\alpha)^3} \le t, \quad \forall\alpha : l_\alpha \le \alpha \le u_\alpha, \tag{134}$$

which is equivalent to

$$\left((a-\alpha)\sigma_w + \frac{3b\sigma_w^3}{1-\alpha}\right)^2 \le (1-l_\alpha)t - 6b^2\frac{\sigma_w^6}{(1-l_\alpha)^2}, \quad \forall\alpha : l_\alpha \le \alpha \le u_\alpha. \tag{135}$$

Defining

$$\tau^2 = (1-l_\alpha)t - 6b^2\frac{\sigma_w^6}{(1-l_\alpha)^2} \ge 0, \tag{136}$$

we have that

$$B = \min_{a,b,\tau}\left\{\frac{\tau^2}{1-l_\alpha} + 6b^2\frac{\sigma_w^6}{(1-l_\alpha)^3}\right\} \tag{137}$$

subject to

$$\left((a-\alpha)\sigma_w + \frac{3b\sigma_w^3}{1-\alpha}\right)^2 \le \tau^2, \quad \forall\alpha : l_\alpha \le \alpha \le u_\alpha, \tag{138}$$

or, equivalently,

$$(a-\alpha)\sigma_w + \frac{3b\sigma_w^3}{1-\alpha} \le \tau, \quad \forall\alpha : l_\alpha \le \alpha \le u_\alpha; \tag{139}$$

$$(a-\alpha)\sigma_w + \frac{3b\sigma_w^3}{1-\alpha} \ge -\tau, \quad \forall\alpha : l_\alpha \le \alpha \le u_\alpha. \tag{140}$$

Consider first the constraint (139). Let $\alpha = \zeta + e$ where $\zeta = (u_\alpha + l_\alpha)/2$. Then $l_\alpha \le \alpha \le u_\alpha$ is equivalent to $e^2 \le \epsilon^2$ where $\epsilon = (u_\alpha - l_\alpha)/2$, so that (139) can be expressed as

$$(1-\zeta-e)(a-\zeta-e)\sigma_w + 3b\sigma_w^3 \le \tau(1-\zeta-e), \quad \forall e : e^2 \le \epsilon^2, \tag{141}$$

which in turn is equivalent to the implication

$$P(e) \stackrel{\triangle}{=} \epsilon^2 - e^2 \ge 0 \Rightarrow Q(e) \ge 0, \tag{142}$$

where

$$\begin{aligned} Q(e) &= \tau(1-\zeta-e) - (1-\zeta-e)(a-\zeta-e)\sigma_w - 3b\sigma_w^3 \\ &= -\sigma_w e^2 + e((a-2\zeta+1)\sigma_w - \tau) + \tau(1-\zeta) - (1-\zeta)(a-\zeta)\sigma_w - 3b\sigma_w^3. \end{aligned} \tag{143}$$

From (63) and Lemma 3 it follows that the condition (139) is equivalent to the condition (121), for some $\gamma \geq 0$. If (121) is satisfied, then $\gamma \geq \sigma_w \geq 0$ so that it is not necessary to impose the additional constraint $\gamma \geq 0$. Similarly, we can show that the condition (140) is equivalent to the condition (122) for some $\beta \geq 0$, completing the proof of the theorem.

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. Nemirovski, and A. Ben-Tal, "Robust mean-squared error estimation with bounded data uncertainties," in preparation.

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

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

[30] 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.

[31] M. Sion, "On general minimax theorems," *Pac. J. Math.*, vol. 8, pp. 171–176, 1958.