CCIT Report #547
August 2005

Comparing Between Estimation Approaches: Admissible and Dominating Linear Estimators

Yonina C. Eldar

Abstract

We treat the problem of evaluating the performance of linear estimators for estimating a deterministic parameter vector $\mathbf{x}$ in a linear regression model, with the mean-squared error (MSE) as the performance measure. Since the MSE depends on the unknown vector $\mathbf{x}$, direct comparison between estimators is a difficult problem. Here we consider a framework for examining the MSE of different linear estimation approaches based on the concepts of admissible and dominating estimators. We develop a general procedure for determining whether or not a linear estimator is MSE admissible, and for constructing an estimator strictly dominating a given inadmissible method, so that its MSE is smaller for all $\mathbf{x}$. In particular we show that both problems can be addressed in a unified manner for arbitrary constraint sets on $\mathbf{x}$ by considering a certain convex optimization problem. We then demonstrate the details of our method for the case in which $\mathbf{x}$ is constrained to an ellipsoidal set, and for unrestricted choices of $\mathbf{x}$.

As a by product of our results, we derive a closed form solution for the minimax MSE estimator on an ellipsoid, which is valid for arbitrary model parameters, as long as the signal-to-noise-ratio exceeds a certain threshold.

Key Words—Linear estimation, regression, admissible estimators, dominating estimators, mean-squared error (MSE) estimation, minimax MSE estimation.

I. Introduction

An important estimation problem that has been treated extensively in the literature is the problem of estimating a deterministic parameter vector $\mathbf{x}$ in the regression model $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w}$, where $\mathbf{y}$ are the observations, $\mathbf{H}$ is a given model matrix, and $\mathbf{w}$ is a random noise vector with positive definite covariance matrix. In an estimation context, the goal is to construct an estimator $\hat{\mathbf{x}}$ of $\mathbf{x}$ from the observations $\mathbf{y}$ that is close to $\mathbf{x}$ in some sense. In this paper, we focus on *linear* estimation, in which the estimator $\hat{\mathbf{x}}$ is linear in the observations $\mathbf{y}$.

A popular measure of estimator performance is the mean-squared error (MSE), which is the average squared-norm of the estimation error $\hat{\mathbf{x}} - \mathbf{x}$. Due to the fact that the parameter vector $\mathbf{x}$ is assumed to be fixed, the averaging is only over the noise, and not over $\mathbf{x}$, typically resulting in a parameter-dependent MSE. Since the MSE generally depends on $\mathbf{x}$, it cannot be minimized directly. Thus, alternative criteria for constructing estimators must be sought.

Beginning with the celebrated least-squares estimator proposed by Gauss, a myriad of linear and nonlinear estimators have been developed for the regression model with the common goal of leading to"good" MSE performance. The first estimators considered for this problem where restricted to be linear and unbiased [1], [2], [3]. In the past 30 years attempts have been made to develop linear estimators that may be biased but

Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel. E-mail: yonina@ee.technion.ac.il. This work was supported in part by the EU 6th framework programme, via the NEWCOM network of excellence, and by the Israel Science Foundation under Grant No. 536/04.

closer to the true parameter $\mathbf{x}$ in an MSE sense. These include the Tikhonov regularizer [4], [5], the shrunken estimator [6], and the covariance shaping least-squares estimator [7]. Another recent approach is to assume that $\mathbf{x}$ is constrained to a subset $\mathcal{U}$, and then seek linear minimax estimators that minimize a worst-case measure of MSE on $\mathcal{U}$ [8], [9], [10], [11], [12], [13], [14].

The difficulty encountered in this estimation problem is that the MSE performance of a linear estimator depends generally on $\mathbf{x}$, rendering comparison between the different estimators a difficult (and often impossible) task. This explains the multitude of approaches suggested for this problem. Since ranking the estimators in terms of MSE is not obvious, an important practical question is how to decide which estimator to use. Although in general this question is hard to answer, some estimators may be uniformly better than others in terms of MSE. An estimator $\hat{\mathbf{x}}$ is said to *dominate* a given estimator $\hat{\mathbf{x}}_0$ on a set $\mathcal{U}$ if its MSE is never larger than that of $\hat{\mathbf{x}}_0$ for all values of $\mathbf{x}$ in $\mathcal{U}$, and is strictly smaller for some $\mathbf{x}$ in $\mathcal{U}$ [15]; $\hat{\mathbf{x}}$ *strictly dominates* $\hat{\mathbf{x}}_0$ if its MSE is smaller than that of $\hat{\mathbf{x}}_0$ for all $\mathbf{x}$ in $\mathcal{U}$. An estimator that is not dominated by any other estimator is said to be *admissible* on $\mathcal{U}$. Clearly, a desirable property of an estimator is that it is admissible: otherwise it is dominated by some other linear estimator which will have better MSE performance for all choices of $\mathbf{x}$.

Constructing an admissible estimator is a trivial task; for example, if $\mathcal{U}$ includes the zero vector, then $\hat{\mathbf{x}} = 0$ is admissible since it is the only linear estimator with zero MSE for $\mathbf{x} = 0$. A more interesting problem is to determine whether a given linear estimator is admissible, or more generally, to characterize all linear admissible estimators on a subset $\mathcal{U}$. Such a description can lead to practical methods for selecting between estimators. If our estimator of choice is inadmissible, then we are guaranteed theoretically that a better estimator exists in an MSE sense. However, the theory does not provide a concrete method for finding such an estimator. Therefore, given an inadmissible estimator, it is of interest to develop a general procedure for constructing a dominating admissible estimator, *i.e.,* an estimator that is uniformly better in terms of MSE.

The problem of characterizing all linear admissible estimators has received considerable attention in the statistical literature. Admissible linear estimators for arbitrary values of $\mathbf{x}$ are discussed in [16], [17], [18], [19], ellipsoidal constraints are treated in [20], [21], [22], and linear inequality constraints on $\mathbf{x}$ are considered in [23]. However, these references do not address the important practical problem of constructing an admissible estimator dominating a given inadmissible approach.

A specific class of dominating estimators that has been studied extensively in the literature are estimators dominating the least-squares method when the noise is Gaussian. Following the seminal work of Stein [24], many nonlinear approaches dominating the least-squares estimator have been developed, including the James-Stein estimator [25], among others [26], [27], [28], [29]. In [14] it was shown that the linear minimax MSE estimator that minimizes the worst-case MSE on a bounded set $\mathcal{U}$ dominates the least-squares estimator on $\mathcal{U}$.

In this paper we focus on linear estimation, and develop a general procedure for determining whether or not an estimator is admissible on an arbitrary constraint set $\mathcal{U}$. We also propose a method for constructing admissible estimators that strictly dominate any given inadmissible strategy. Our approach unifies many of the previous admissibility results and provides explicit methods for constructing admissible dominating estimators. Thus, in applications, these results can be used to choose between the multitude of linear estimators available in the literature, and can also suggest new methods of estimation when the conventional techniques are inadmissible.

To reduce the abstract concepts of admissible and dominating estimators to a concrete mathematical problem, we show that both issues of admissibility and domination can be addressed by considering a certain convex optimization problem. The advantage of this formulation is that we can use duality theory to develop explicit necessary and sufficient admissibility and domination conditions. To demonstrate the details of our approach, we consider the case in which $\mathcal{U}$ consists of vectors $\mathbf{x}$ that satisfy the weighted norm constraint $\mathbf{x}^*\mathbf{T}\mathbf{x} \leq U^2$ for some positive definite matrix $\mathbf{T}$ and scalar U. For this setting we present explicit necessary and sufficient conditions on a linear estimator $\hat{\mathbf{x}}$ to be admissible which are easy to verify. Given an inadmissible estimator, we also develop closed from expressions for linear dominating estimators in many special cases. In the general case, we show that a dominating estimator can be found by solving a semidefinite programming problem (SDP) [30], [31], which is a convex optimization problem that can be solved efficiently. As a by product of our results, we provide a closed form solution for the minimax MSE estimator of [10] that minimizes the worst-case MSE for signal-to-noise-ratios (SNRs) exceeding a certain threshold. Although this problem has been treated previously by Pilz in [9], we show that the solution obtained in [9] is different that ours, and is incorrect in general. We also investigate the admissibility of the Tikhonov and shrunken estimators on ellipsoidal constraint sets. We then extend the admissibility and domination results to the case in which $\mathbf{x}$ is not restricted.

The paper is organized as follows. In Section II we introduce our problem and discuss the main ideas and results. In Section III we show that both problems of determining the admissibility of an estimator on an arbitrary constraint set $\mathcal{U}$ as well as constructing estimators dominating inadmissible estimators on $\mathcal{U}$ can be treated by considering a certain convex optimization problem. Dominating estimators and necessary and sufficient admissibility conditions on an ellipsoidal constraint set are developed in Sections IV and V, respectively. The minimax MSE estimator and a comparison with the estimator developed in [9] are also treated in Section IV. Section VI presents some examples of inadmissible and dominating estimators. Finally, the results are extended to unrestricted choices of $\mathbf{x}$ in Section VII.

II. Problem Formulation and Main Results

We denote vectors in $\mathbb{C}^m$ by boldface lowercase letters and matrices in $\mathbb{C}^{n\times m}$ by boldface uppercase letters. The identity matrix is denoted by $\mathbf{I}$, $(\cdot)^*$ and $(\cdot)^\dagger$ denote the Hermitian conjugate and the pseudoinverse of the corresponding matrix respectively, and $\operatorname{diag}(\delta_1, \ldots, \delta_m)$ denotes an $m \times m$ diagonal matrix with diagonal elements δ_i. The notation $\mathbf{A} \succ 0$ ($\mathbf{A} \succeq 0$) means that $\mathbf{A}$ is Hermitian and positive (nonnegative) definite, and $\mathbf{Y} \succeq \mathbf{Z}$ means that $\mathbf{Y} - \mathbf{Z} \succeq 0$. Given a Hermitian matrix $\mathbf{A}$, we denote by $(\mathbf{A})^{1/2}$ the Hermitian square-root, and by $\lambda_{\max}(\mathbf{A})$ and $\lambda_{\min}(\mathbf{A})$ the largest and smallest eigenvalues, respectively. The range space and null space of a matrix $\mathbf{A}$ are denoted respectively by $\mathcal{R}(\mathbf{A})$ and $\mathcal{N}(\mathbf{A})$.

We consider the problem of estimating the deterministic parameter vector $\mathbf{x}$ in the linear regression model

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

where $\mathbf{H}$ is a known $n \times m$ matrix and $\mathbf{w}$ is a zero-mean random vector with covariance matrix $\mathbf{C}_w$. We assume for simplicity that $\mathbf{H}$ has full column rank and that $\mathbf{C}_w \succ 0$. The vector $\mathbf{x}$ is known to lie in a subset

$\mathcal{U} \subseteq \mathbb{C}^m$, where $\mathcal{U}$ can be arbitrary; in particular, we may choose $\mathcal{U} = \mathbb{C}^m$.

Our goal is to design a *linear* estimator of the form $\hat{\mathbf{x}} = \mathbf{G}\mathbf{y}$ that minimizes the MSE, which is given by

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

Evidently, the MSE depends in general on the unknown parameter vector $\mathbf{x}$ and therefore cannot be minimized directly. An alternative strategy to designing linear estimators is to restrict the estimator to be unbiased, so that $\mathbf{G}\mathbf{H} = \mathbf{I}$, and then seek the estimator that minimizes the variance. This leads to the celebrated least-squares estimator. However, this method does not necessarily result in a small MSE. A myriad of approaches have been suggested to further reduce the MSE, for example, [5], [4], [6], [7], [12], [10], [11], [8], [9] and references therein.

Two important concepts in comparing the MSE of different estimators are those of domination and admissibility. An estimator $\hat{\mathbf{x}}_1$ dominates an estimator $\hat{\mathbf{x}}_2$ on $\mathcal{U}$ if

$$\begin{aligned} &E\left\{\|\hat{\mathbf{x}}_1-\mathbf{x}\|^2\right\} \leq E\left\{\|\hat{\mathbf{x}}_2-\mathbf{x}\|^2\right\}, \text{ for all } \mathbf{x}\in\mathcal{U}; \\ &E\left\{\|\hat{\mathbf{x}}_1-\mathbf{x}\|^2\right\} < E\left\{\|\hat{\mathbf{x}}_2-\mathbf{x}\|^2\right\}, \text{ for some } \mathbf{x}\in\mathcal{U}. \end{aligned} \tag{3}$$

The estimator $\hat{\mathbf{x}}_1$ strictly dominates $\hat{\mathbf{x}}_2$ on $\mathcal{U}$ if

$$E\left\{\|\hat{\mathbf{x}}_1-\mathbf{x}\|^2\right\} < E\left\{\|\hat{\mathbf{x}}_2-\mathbf{x}\|^2\right\}, \text{ for all } \mathbf{x}\in\mathcal{U}. \tag{4}$$

If $\hat{\mathbf{x}}_1$ dominates $\hat{\mathbf{x}}_2$ then clearly it is a better estimator in terms of MSE. An estimator $\hat{\mathbf{x}}$ is *admissible* if it is not dominated by any other linear estimator. If an estimator is inadmissible, then there exists another linear approach which leads to lower MSE on $\mathcal{U}$. Thus, although we cannot directly compare the performance of different strategies, we would like our estimator to at least be admissible.

The previous discussion raises three important questions which we would like to address:

1. Given a linear estimator $\hat{\mathbf{x}}_0$ of $\mathbf{x}$, can we verify whether or not it is admissible on $\mathcal{U}$?
2. If $\hat{\mathbf{x}}_0$ is inadmissible, then
   (a) can we develop a systematic method to construct an admissible linear estimator that dominates $\hat{\mathbf{x}}_0$ on $\mathcal{U}$?
   (b) can we verify whether or not a given estimator dominates $\hat{\mathbf{x}}_0$ on $\mathcal{U}$?

In the next section we provide solutions to the questions above for arbitrary constraint sets $\mathcal{U}$. We then specialize the results in Sections IV–VII to two choices of $\mathcal{U}$.

III. Admissible and Dominating Estimators on Arbitrary Parameter Sets

We begin with the problem of determining whether a given linear estimator $\hat{\mathbf{x}}$ dominates the linear estimator $\hat{\mathbf{x}}_0$. An explicit condition can be obtained by noting that (3) is equivalent to

$$\begin{aligned} &\sup_{\mathbf{x}\in\mathcal{U}}\left\{\epsilon(\mathbf{G},\mathbf{x})-\epsilon(\mathbf{G}_0,\mathbf{x})\right\} \leq 0; \\ &\inf_{\mathbf{x}\in\mathcal{U}}\left\{\epsilon(\mathbf{G},\mathbf{x})-\epsilon(\mathbf{G}_0,\mathbf{x})\right\} < 0. \end{aligned} \tag{5}$$

For particular choices of $\mathcal{U}$, (5) can be reduced to easily verifiable conditions. This is illustrated for ellipsoidal parameter sets in Section IV.

Thus, given an estimator $\hat{\mathbf{x}}_0$, we can determine whether it is dominated by another estimator $\hat{\mathbf{x}}$ by checking if it satisfies (5). However, we still need a constructive method for finding a dominating estimator $\hat{\mathbf{x}}$. In addition, it is desirable to construct $\hat{\mathbf{x}}$ not only to dominate $\hat{\mathbf{x}}_0$ but to also be admissible. The following theorem shows that both problems of admissibility and of constructing a strictly dominating estimator can be addressed by examining the solution to a convex optimization problem.

Theorem 1: Let $\mathbf{x}$ denote the deterministic 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, and $\mathbf{w}$ is a zero-mean random vector with covariance $\mathbf{C}_w \succ 0$. Let $\hat{\mathbf{x}}_0 = \mathbf{G}_0\mathbf{y}$ be a given linear estimate of $\mathbf{x}$, let $\epsilon(\mathbf{G}, \mathbf{x}) = E\{\|\mathbf{G}\mathbf{y} - \mathbf{x}\|^2\}$, and let $\mathcal{U} \subseteq \mathbb{C}^m$. Define the matrix $\widehat{\mathbf{G}}$ as

$$\widehat{\mathbf{G}} = \arg\min_{\mathbf{G}} \mathcal{M}(\mathbf{G}, \mathbf{G}_0) \tag{6}$$

where

$$\mathcal{M}(\mathbf{G}, \mathbf{G}_0) = \sup_{\mathbf{x} \in \mathcal{U}} \{\epsilon(\mathbf{G}, \mathbf{x}) - \epsilon(\mathbf{G}_0, \mathbf{x})\}, \tag{7}$$

let $\mathcal{V}(\mathbf{G}_0) = \min_{\mathbf{G}} \mathcal{M}(\mathbf{G}, \mathbf{G}_0)$, and define $\hat{\mathbf{x}} = \widehat{\mathbf{G}}\mathbf{y}$. Then

1. $\widehat{\mathbf{G}}$ is unique;
2. $\hat{\mathbf{x}}_0$ is admissible on $\mathcal{U}$ if and only if $\widehat{\mathbf{G}} = \mathbf{G}_0$, or equivalently, if and only if $\mathcal{V}(\mathbf{G}_0) = 0$;
3. If $\mathcal{V}(\mathbf{G}_0) < 0$ then $\hat{\mathbf{x}}$ strictly dominates $\hat{\mathbf{x}}_0$ on $\mathcal{U}$;
4. $\hat{\mathbf{x}}$ is admissible on $\mathcal{U}$.

Before proving the theorem we note that the minimum in (6) is well defined since the objective is continuous and coercive [32].

Proof: We prove each of the statements 1-4:

1. Since $\epsilon(\mathbf{G}, \mathbf{x}) - \epsilon(\mathbf{G}_0, \mathbf{x})$ is strictly convex in $\mathbf{G}$ (because $\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$ is strictly convex) $\mathcal{M}(\mathbf{G}, \mathbf{G}_0)$ is also strictly convex and the minimum of $\mathcal{M}(\mathbf{G}, \mathbf{G}_0)$ is unique.

2. Since $\mathcal{M}(\mathbf{G}_0, \mathbf{G}_0) = 0$ and $\widehat{\mathbf{G}}$ is unique, $\mathcal{V}(\mathbf{G}_0) = 0$ implies that $\widehat{\mathbf{G}} = \mathbf{G}_0$. The converse is trivially true.

Now, if $\hat{\mathbf{x}}_0$ is admissible, then for every $\mathbf{G}$ there exists an $\mathbf{x} \in \mathcal{U}$ such that $\epsilon(\mathbf{G}, \mathbf{x}) \geq \epsilon(\mathbf{G}_0, \mathbf{x})$. Thus, $\mathcal{M}(\mathbf{G}, \mathbf{G}_0) \geq 0$ for all $\mathbf{G}$, and $\mathcal{V}(\mathbf{G}_0) \geq 0$. Since $\mathcal{M}(\mathbf{G}_0, \mathbf{G}_0) = 0$, we also have that $\mathcal{V}(\mathbf{G}_0) \leq 0$, from which we conclude that $\mathcal{V}(\mathbf{G}_0) = 0$. Conversely, if $\mathcal{V}(\mathbf{G}_0) = 0$, then since $\widehat{\mathbf{G}}$ is unique, for any $\mathbf{G} \neq \mathbf{G}_0$, we have that $\mathcal{M}(\mathbf{G}, \mathbf{G}_0) > 0$, and $\mathbf{G}_0$ is admissible.

3. If $\mathcal{V}(\mathbf{G}_0) < 0$, then $\epsilon(\widehat{\mathbf{G}}, \mathbf{x}) < \epsilon(\mathbf{G}_0, \mathbf{x})$ for all $\mathbf{x} \in \mathcal{U}$, and $\widehat{\mathbf{G}}$ strictly dominates $\mathbf{G}_0$.

4. From part 2 it suffices to show that $\mathcal{V}(\widehat{\mathbf{G}}) = 0$. Since $\widehat{\mathbf{G}}$ is the unique minimum, for any $\mathbf{G} \neq \widehat{\mathbf{G}}$,

$$\sup_{\mathbf{x} \in \mathcal{U}} \{\epsilon(\mathbf{G}, \mathbf{x}) - \epsilon(\mathbf{G}_0, \mathbf{x})\} > \sup_{\mathbf{x} \in \mathcal{U}} \left\{\epsilon(\widehat{\mathbf{G}}, \mathbf{x}) - \epsilon(\mathbf{G}_0, \mathbf{x})\right\}. \tag{8}$$

If $\mathcal{V}(\widehat{\mathbf{G}}) < 0$, then there exists a $\mathbf{G}$ such that

$$\epsilon(\mathbf{G}, \mathbf{x}) < \epsilon(\widehat{\mathbf{G}}, \mathbf{x}), \quad \forall \mathbf{x} \in \mathcal{U}, \tag{9}$$

from which we conclude that

$$\epsilon(\mathbf{G},\mathbf{x})-\epsilon(\mathbf{G}_0,\mathbf{x})<\epsilon(\widehat{\mathbf{G}},\mathbf{x})-\epsilon(\mathbf{G}_0,\mathbf{x}),\quad \forall \mathbf{x}\in\mathcal{U}. \tag{10}$$

But then

$$\sup_{\mathbf{x}\in\mathcal{U}}\left\{\epsilon(\mathbf{G},\mathbf{x})-\epsilon(\mathbf{G}_0,\mathbf{x})\right\}\le \sup_{\mathbf{x}\in\mathcal{U}}\left\{\epsilon(\widehat{\mathbf{G}},\mathbf{x})-\epsilon(\mathbf{G}_0,\mathbf{x})\right\}, \tag{11}$$

which contradicts (8). Thus $\mathcal{V}(\mathbf{G}_0)=0$. ■

Theorem 1 provides a general recipe for determining admissibility of a linear estimator and for constructing admissible and strictly dominating estimators, by solving a convex optimization problem[1]. Therefore, we can now use the many known techniques for dealing with convex problems in order to develop admissibility and domination conditions, as we show in the ensuing sections. For general choices of $\mathcal{U}$, iterative procedures that are designed for saddle point problems can be used to solve (6), such as subgradient algorithms [33] or the prox method [34].

An interesting result of part 3 of the theorem is that if $\hat{\mathbf{x}}_0$ is inadmissible then there exists an estimator that *strictly* dominates it. We also note that the only properties of the error measure $\epsilon(\mathbf{G},\mathbf{x})$ that are used in the proof of the theorem are that the function is continuous, coercive and strictly convex. Thus, Theorem 1 is valid for any error measure with these characteristics.

An immediate consequence of Theorem 1 is that the minimax MSE estimator [10] $\hat{\mathbf{x}}_{\text{MX}}=\mathbf{G}_{\text{MX}}\mathbf{y}$ which is designed to minimize the worst-case MSE over $\mathbf{x}\in\mathcal{U}$, strictly dominates any inadmissible unbiased linear estimator $\hat{\mathbf{x}}_{\text{UB}}=\mathbf{G}_{\text{UB}}\mathbf{y}$. Specifically, by definition $\mathbf{G}_{\text{MX}}$ is the solution to

$$\mathbf{G}_{\text{MX}}=\arg\min_{\mathbf{G}}\sup_{\mathbf{x}\in\mathcal{U}}\epsilon(\mathbf{G},\mathbf{x}). \tag{12}$$

For an unbiased linear estimator the MSE, which we denote by $\epsilon(\mathbf{G}_{\text{UB}})$, does not depend on $\mathbf{x}$. Therefore,

$$\mathbf{G}_{\text{MX}}=\arg\min_{\mathbf{G}}\sup_{\mathbf{x}\in\mathcal{U}}\left\{\epsilon(\mathbf{G},\mathbf{x})-\epsilon(\mathbf{G}_{\text{UB}})\right\}. \tag{13}$$

Combining (13) with part 4 of the theorem it follows that if $\mathbf{G}_{\text{UB}}$ is inadmissible, then $\hat{\mathbf{x}}_{\text{MX}}$ is admissible and strictly dominates $\hat{\mathbf{x}}_{\text{UB}}$.

IV. Dominating Estimators on an Ellipsoid

We now consider the optimization problem (6) for ellipsoidal uncertainty sets of the form

$$\mathcal{U}=\left\{\mathbf{x}\in\mathbb{C}^m\left|\mathbf{x}^*\mathbf{T}\mathbf{x}\le U^2\right.\right\}, \tag{14}$$

with $U>0$ and $\mathbf{T}\succ 0$. Specifically, we develop necessary and sufficient optimality conditions on $\widehat{\mathbf{G}}$ by using Lagrange duality theory, as well as explicit closed from solutions for many special cases. Using these conditions, we derive a closed form expression for the minimax MSE estimator of (12) for high SNR values.

[1]Note that the problem is convex in $\mathbf{G}$ for arbitrary constraint sets $\mathcal{U}$ since the supremum of a convex function over any set $\mathcal{U}$ is convex.

This corrects a previous erroneous result of Pilz [9]. (A detailed discussion of the Pilz estimator is provided in Appendix C.) We also consider efficient numerical solutions for arbitrary choices of $\mathbf{G}_0$ and $\mathbf{T}$. In Section V we use these results to develop an easily verifiable set of necessary and sufficient conditions on an estimator $\hat{\mathbf{x}}$ to be admissible on $\mathcal{U}$.

We begin with the problem of determining whether a given linear estimator $\hat{\mathbf{x}}$ dominates the linear estimator $\hat{\mathbf{x}}_0$ on the set $\mathcal{U}$ defined by (14). Thus, we consider the conditions (5) on $\mathcal{U}$.

From (2) we have that

$$\begin{aligned}\epsilon(\mathbf{G},\mathbf{x}) - \epsilon(\mathbf{G}_0,\mathbf{x}) &= \\ &= \mathbf{x}^*\mathbf{Z}\mathbf{x} + \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) - \mathrm{Tr}(\mathbf{G}_0\mathbf{C}_w\mathbf{G}_0^*),\end{aligned} \tag{15}$$

where we defined

$$\mathbf{Z} = (\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H}) - (\mathbf{I}-\mathbf{G}_0\mathbf{H})^*(\mathbf{I}-\mathbf{G}_0\mathbf{H}). \tag{16}$$

By introducing the change of variable $\mathbf{u} = \mathbf{T}^{1/2}\mathbf{x}$ we have that[2]

$$\begin{aligned}\max_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le U^2} \mathbf{x}^*\mathbf{Z}\mathbf{x} &= \max_{\mathbf{u}^*\mathbf{u}\le U^2} \mathbf{u}^*\mathbf{T}^{-1/2}\mathbf{Z}\mathbf{T}^{-1/2}\mathbf{u} \\ &= U^2 \max\left(\lambda_{\max}(\mathbf{Z}\mathbf{T}^{-1}),0\right),\end{aligned} \tag{17}$$

where we used the fact that the eigenvalues of $\mathbf{AB}$ are equal to the eigenvalues of $\mathbf{BA}$ for any $\mathbf{A}$ and $\mathbf{B}$. Similarly,

$$\min_{\mathbf{x}^*\mathbf{T}\mathbf{x}\le U^2} \mathbf{x}^*\mathbf{Z}\mathbf{x} = U^2 \min\left(\lambda_{\min}(\mathbf{Z}\mathbf{T}^{-1}),0\right). \tag{18}$$

Using (17) and (18), the conditions (5) become

$$\begin{aligned}U^2 \max\left(\lambda_{\max}(\mathbf{Z}\mathbf{T}^{-1}),0\right) &\le \mathrm{Tr}\left(\mathbf{C}_w\left(\mathbf{G}_0^*\mathbf{G}_0 - \mathbf{G}^*\mathbf{G}\right)\right); \\ U^2 \min\left(\lambda_{\min}(\mathbf{Z}\mathbf{T}^{-1}),0\right) &< \mathrm{Tr}\left(\mathbf{C}_w\left(\mathbf{G}_0^*\mathbf{G}_0 - \mathbf{G}^*\mathbf{G}\right)\right).\end{aligned} \tag{19}$$

We note that alternative conditions are derived in [21].

A. Necessary and Sufficient Optimality Conditions

The next problem we address is how to construct an admissible estimator $\hat{\mathbf{x}}$ strictly dominating a given inadmissible estimator $\hat{\mathbf{x}}_0$. From Theorem 1 it follows that $\hat{\mathbf{x}}$ can be constructed as the solution to the problem

$$\min_{\mathbf{G}} \max_{\mathbf{x}\in\mathcal{U}} \left\{\epsilon(\mathbf{G},\mathbf{x}) - \epsilon(\mathbf{G}_0,\mathbf{x})\right\}. \tag{20}$$

Substituting (2) into (20), and noting that $\mathrm{Tr}(\mathbf{G}_0\mathbf{C}_w\mathbf{G}_0^*)$ is independent of $\mathbf{x}$ and $\mathbf{G}$, our problem becomes

$$\begin{aligned}\min_{\mathbf{G}} \max_{\mathbf{x}\in\mathcal{U}} \{&\mathbf{x}^*\left((\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H}) - \mathbf{A}\right)\mathbf{x} \\ &+\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) - \mathrm{Tr}(\mathbf{G}_0\mathbf{C}_w\mathbf{G}_0^*)\},\end{aligned} \tag{21}$$

[2]Here we have replaced sup by max since the set $\mathcal{U}$ is compact.

where we defined

$$\mathbf{A} = (\mathbf{I} - \mathbf{G}_0\mathbf{H})^*(\mathbf{I} - \mathbf{G}_0\mathbf{H}). \tag{22}$$

Lemma 1 below asserts that the unique optimal $\mathbf{G}$ has the form $\mathbf{G} = \mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ for an $m \times m$ matrix $\mathbf{B}$, which reduces the dimensionality of the problem when $m < n$.

Lemma 1: Let the $m \times n$ matrix $\mathbf{G}$ be the solution to the problem (21), where $\mathbf{C}_w \succ 0$, $\mathbf{H}$ is an $n \times m$ matrix with rank m, and $\mathcal{U} = \{\mathbf{x} \in \mathbb{C}^m \,|\mathbf{x}^*\mathbf{T}\mathbf{x} \leq U^2\}$. Then

$$\mathbf{G} = \mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}, \tag{23}$$

where $\mathbf{B}$ is the $m \times m$ matrix that is the solution to

$$\begin{aligned} \min_{\mathbf{B}} \Big\{ & \mathrm{Tr}\left(\mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{B}^*\right) \\ & + \max_{\mathbf{x}\in\mathcal{U}} \mathbf{x}^* \left((\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) - \mathbf{A}\right)\mathbf{x} \Big\}. \end{aligned} \tag{24}$$

Proof: see Appendix A. ■

Using Lemma 1 our problem reduces to finding the matrix $\mathbf{B}$ that is the solution to (24). From (17),

$$\max_{\mathbf{x}\in\mathcal{U}} \mathbf{x}^* \left((\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) - \mathbf{A}\right)\mathbf{x} = U^2 \max\left(\lambda_{\max}, 0\right), \tag{25}$$

where

$$\lambda_{\max} \stackrel{\triangle}{=} \lambda_{\max}(\mathbf{T}^{-1/2}\left((\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) - \mathbf{A}\right)\mathbf{T}^{-1/2}). \tag{26}$$

We can express $\max\left(\lambda_{\max}, 0\right)$ as the solution to

$$\min_{\lambda\geq 0} \lambda \tag{27}$$

subject to

$$\mathbf{T}^{-1/2}\left((\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) - \mathbf{A}\right)\mathbf{T}^{-1/2} \preceq \lambda\mathbf{I}. \tag{28}$$

The problem (24) can then be reformulated as

$$\min_{\mathbf{B},\lambda\geq 0} \left\{\mathrm{Tr}\left(\mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{B}^*\right) + U^2\lambda\right\} \tag{29}$$

subject to (28).

Since the objective in (29) and the constraint (28) are convex in $\mathbf{B}$ and λ, and the problem is strictly feasible, the Karush-Kuhn-Tucker (KKT) conditions [32] assert that $\mathbf{B}$ is optimal if and only if there exists a matrix $\Pi \succeq 0$ and a scalar $\mu \geq 0$ such that the following conditions hold:

1. $d\mathcal{L}/d\mathbf{B} = 0$ and $d\mathcal{L}/d\lambda = 0$ where the Lagrangian $\mathcal{L}$ is defined by

$$\begin{aligned} \mathcal{L} = & \mathrm{Tr}\left(\mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{B}^*\right) + \left(U^2 - \mu - \mathrm{Tr}(\Pi)\right)\lambda \\ & + \mathrm{Tr}\left(\mathbf{T}^{-1/2}\Pi\mathbf{T}^{-1/2}\left((\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) - \mathbf{A}\right)\right). \end{aligned}$$

2. Feasibility: $(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})-\mathbf{A} \preceq \lambda\mathbf{T}$, $\lambda \geq 0$;
3. Complementary slackness:

$$\mathrm{Tr}\left(\mathbf{T}^{-1/2}\Pi\mathbf{T}^{-1/2}\left((\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})-\mathbf{A}\right)\right)=\lambda\mathrm{Tr}(\Pi);$$
$$\mu\lambda = 0.$$

In Appendix B we show that these conditions are equivalent to the conditions in the following theorem:

Theorem 2: Consider the problem of Theorem 1. Let $\hat{\mathbf{x}}_0 = \mathbf{G}_0\mathbf{y}$ be a given linear estimate of $\mathbf{x}$, let $\mathbf{A}=(\mathbf{I}-\mathbf{G}_0\mathbf{H})^*(\mathbf{I}-\mathbf{G}_0\mathbf{H})$, $\mathbf{Q}=\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and $\mathcal{U}=\{\mathbf{x}\in\mathbb{C}^m | \mathbf{x}^*\mathbf{T}\mathbf{x}\leq U^2\}$. Then,

$$\widehat{\mathbf{G}}=\min_{\mathbf{G}}\max_{\mathbf{x}\in\mathcal{U}}\{\epsilon(\mathbf{G},\mathbf{x})-\epsilon(\mathbf{G}_0,\mathbf{x})\}$$

has the form $\widehat{\mathbf{G}}=\mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ where $\mathbf{B}$ satisfies the conditions below for some $\lambda\geq 0$:
1. $\mathbf{QB}=\mathbf{B}^*\mathbf{Q}$;
2. $0 \preceq \mathbf{QB} \prec \mathbf{Q}$;
3. $(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) \preceq \lambda\mathbf{T}+\mathbf{A}$;
4. $(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})\mathbf{B}=(\lambda\mathbf{T}+\mathbf{A})\,\mathbf{B}$;
5. $\mathrm{Tr}(\mathbf{B}\,(\mathbf{I}-\mathbf{B})^{-1}\,\mathbf{Q}^{-1}\mathbf{T})\leq U^2$;
6. $\lambda(\mathrm{Tr}(\mathbf{B}\,(\mathbf{I}-\mathbf{B})^{-1}\,\mathbf{Q}^{-1}\mathbf{T})-U^2)=0$.

Using the optimality conditions of Theorem 2 we can easily show that $\widehat{\mathbf{G}}=0$, or equivalently, $\mathbf{B}=0$, if and only if $\mathbf{A}\succeq\mathbf{I}$. Indeed, if $\mathbf{B}=0$, then conditions 1,2,4 and 5 are trivially satisfied. To satisfy 6 we must have that $\lambda=0$, in which case condition 3 becomes $\mathbf{A}\succeq\mathbf{I}$.

B. Closed Form Admissible and Dominating Estimators

We now introduce two classes of problems for which the optimality conditions of Theorem 2 can be solved explicitly. In Section IV-B.1 we treat the case of jointly diagonalizable matrices, and in Section IV-B.2, the high SNR scenario with $\mathbf{A}\preceq\mathbf{I}$.

B.1 Jointly Diagonalizable Matrices

Suppose that $\mathbf{T},\mathbf{Q}=\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ and $\mathbf{G}_0\mathbf{H}$ have the same eigenvector matrix. Thus, if $\mathbf{Q}$ has an eigendecomposition $\mathbf{Q}=\mathbf{V}\Sigma\mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix and $\Sigma=\mathrm{diag}(\sigma_1,\ldots,\sigma_m)$ with $\sigma_i>0$, then

$$\mathbf{T}=\mathbf{V}\Lambda\mathbf{V}^*; \tag{30}$$
$$\mathbf{G}_0\mathbf{H}=\mathbf{V}\Delta\mathbf{V}^*, \tag{31}$$

where $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_m)$ with $\lambda_i>0$ and $\Delta=\mathrm{diag}(\delta_1,\ldots,\delta_m)$. A closed form expression for the optimal matrix $\mathbf{G}$ in this case is given in the following theorem.

Theorem 3: Consider the problem of Theorem 2. Let $\mathbf{Q}=\mathbf{V}\Sigma\mathbf{V}^*$ where $\Sigma=\mathrm{diag}(\sigma_1,\ldots,\sigma_m)$ with $\sigma_i>0$, and suppose that $\mathbf{T}=\mathbf{V}\Lambda\mathbf{V}^*$ and $\mathbf{G}_0\mathbf{H}=\mathbf{V}\Delta\mathbf{V}^*$ where $\Lambda=\mathrm{diag}(\lambda_1,\ldots,\lambda_m)$ with $\lambda_i>0$ and $\Delta=$

$\text{diag}(\delta_1, \ldots, \delta_m)$. Then

$$\widehat{\mathbf{G}} = \mathbf{VDV}^*(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$$

where $\mathbf{D} = \text{diag}(d_1, \ldots, d_m)$ with

$$d_i = \begin{cases} 1 - \sqrt{\eta_i}, & \eta_i < 1; \\ 0, & \eta_i \geq 1. \end{cases} \tag{32}$$

Here

$$\eta_i = \lambda\lambda_i + |1 - \delta_i|^2, \tag{33}$$

where λ is determined as follows. Define

$$\mathcal{G}(\lambda) = \sum_{i:\eta_i<1} \frac{\lambda_i}{\sigma_i}\left(\frac{1}{\sqrt{\eta_i}} - 1\right) - U^2. \tag{34}$$

Then, $\lambda = 0$ if $\mathcal{G}(0) \leq 0$. Otherwise, λ is the unique value in the range $(0, \alpha)$ for which $\mathcal{G}(\lambda) = 0$, where

$$\alpha = \max_i \frac{1 - |1 - \delta_i|^2}{\lambda_i}. \tag{35}$$

Proof: To prove the theorem we need to show that $\mathbf{B} = \mathbf{VDV}^*$ with d_i given by (32) satisfies the conditions of Theorem 2. Using the fact that $\mathbf{Q} = \mathbf{V}\Sigma\mathbf{V}^*$, $\mathbf{T} = \mathbf{V}\Lambda\mathbf{V}^*$, $\mathbf{G}_0\mathbf{H} = \mathbf{V}\Delta\mathbf{V}^*$ and $\Sigma \succ 0$, these conditions become:

1. $d_i = d_i^*$;
2. $0 \leq d_i < 1$;
3. $(1 - d_i)^2 \leq \lambda\lambda_i + |1 - \delta_i|^2$;
4. $d_i(1 - d_i)^2 = d_i\left(\lambda\lambda_i + |1 - \delta_i|^2\right)$;
5. $\sum_{i=1}^m \frac{\lambda_i d_i}{\sigma_i(1-d_i)} \leq U^2$;
6. $\lambda\left(\sum_{i=1}^m \frac{\lambda_i d_i}{\sigma_i(1-d_i)} - U^2\right) = 0$.

Since $\eta_i \geq 0$, d_i is real and the first condition is satisfied. By definition of d_i, we also have that $d_i \geq 0$. To show that $d_i < 1$ we need to show that $\eta_i > 0$. Since $\lambda \geq 0$, $\eta_i \geq 0$. Furthermore, $\eta_i = 0$ if and only if $\delta_i = 0$ and $\lambda = 0$. But, if $\eta_i = 0$, then $\mathcal{G}(0) \to \infty$ so that λ cannot be equal to 0. Therefore, $\eta_i > 0$ and $d_i < 1$.

To satisfy conditions 3 and 4 we must have that

$$(1 - d_i)^2 = \lambda\lambda_i + |1 - \delta_i|^2, \quad i : d_i \neq 0, \tag{36}$$

and

$$1 \leq \lambda\lambda_i + |1 - \delta_i|^2, \quad i : d_i = 0. \tag{37}$$

Since $\eta_i = \lambda\lambda_i + |1 - \delta_i|^2$, both conditions are satisfied. Finally, the last two conditions are satisfied by our definition of $\mathcal{G}(\lambda)$. It remains to show that if $\mathcal{G}(0) > 0$, then there is a unique value of $0 < \lambda < \alpha$ such that $\mathcal{G}(\lambda) = 0$.

Since η_i is monotonically increasing in $\lambda > 0$, and each term in the sum in the definition of $\mathcal{G}(\lambda)$ is positive, $\mathcal{G}(\lambda)$ is monotonically decreasing in λ as long as there exists at least on i for which $\eta_i < 1$, or equivalently, as long as $\lambda < \alpha$. In addition, $\mathcal{G}(\lambda)$ is continuous. Now, we are assuming that $\mathcal{G}(0) > 0$. Furthermore,

$\mathcal{G}(\lambda) = -U^2$ for $\lambda \geq \alpha$. Therefore, there is a unique $0 < \lambda < \alpha$ such that $\mathcal{G}(\lambda) = 0$. ■

B.2 High SNR

We now treat the case in which the matrices are not necessarily commuting, however $\mathbf{A} \preceq \mathbf{I}$ and the SNR exceeds a threshold. In this setting, we can also obtain a closed form solution, as incorporated in the following theorem.

Theorem 4: Consider the problem of Theorem 2. Let

$$\mathbf{B} = \mathbf{I} - \left(\mathbf{Q}^{-1}(\lambda\mathbf{T} + \mathbf{A})\mathbf{Q}^{-1}\right)^{1/2}\mathbf{Q}, \tag{38}$$

where λ is chosen as follows. If $\mathbf{A}^{-1}$ is defined, and

$$\mathrm{Tr}\left(\mathbf{Q}^{-1}\left(\mathbf{Q}\mathbf{A}^{-1}\mathbf{Q}\right)^{1/2}\mathbf{Q}^{-1}\mathbf{T}\right) \leq \mathrm{Tr}\left(\mathbf{Q}^{-1}\mathbf{T}\right) + U^2$$

then $\lambda = 0$. Otherwise, $\lambda > 0$ is the unique solution to

$$\begin{aligned}\mathrm{Tr}&\left(\mathbf{Q}^{-1}\left(\mathbf{Q}(\lambda\mathbf{T} + \mathbf{A})^{-1}\mathbf{Q}\right)^{1/2}\mathbf{Q}^{-1}\mathbf{T}\right)\\ &= \mathrm{Tr}\left(\mathbf{Q}^{-1}\mathbf{T}\right) + U^2.\end{aligned} \tag{39}$$

We then have that if

$$\lambda_{\max}\left(\mathbf{Q}\left(\mathbf{Q}^{-1}(\lambda\mathbf{T} + \mathbf{A})\mathbf{Q}^{-1}\right)^{1/2}\right) \leq 1 \tag{40}$$

then $\widehat{\mathbf{G}} = \mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ with $\mathbf{B}$ given by (38).

Before proving the theorem we show that the solution is indeed valid for large enough SNR values when $\mathbf{A} \preceq \mathbf{I}$. Let $\mathbf{C}_w = \sigma^2\mathbf{C}$ for some given matrix $\mathbf{C}$, such that $U^2/\mathrm{Tr}(\mathbf{C}) = 1$. The SNR is then defined as $U^2/\mathrm{Tr}(\mathbf{C}_w) = 1/\sigma^2$. Denoting $\mathbf{Q}_0 = \mathbf{H}^*\mathbf{C}^{-1}\mathbf{H}$, (40) becomes

$$\lambda_{\max}(\mathbf{Z}) \leq 1, \tag{41}$$

where

$$\mathbf{Z} = \mathbf{Q}_0^{1/2}\left(\mathbf{Q}_0^{-1}(\lambda\mathbf{T} + \mathbf{A})\mathbf{Q}_0^{-1}\right)^{1/2}\mathbf{Q}_0^{1/2}, \tag{42}$$

and λ is chosen to satisfy

$$\begin{aligned}\mathrm{Tr}&\left(\mathbf{Q}_0^{-1}\left(\mathbf{Q}_0(\lambda\mathbf{T} + \mathbf{A})^{-1}\mathbf{Q}_0\right)^{1/2}\mathbf{Q}_0^{-1}\mathbf{T}\right)\\ &= \mathrm{Tr}\left(\mathbf{Q}_0^{-1}\mathbf{T}\right) + \frac{1}{\sigma^2}U^2.\end{aligned} \tag{43}$$

We now show that for high enough SNR, $\lambda_{\max}(\mathbf{Z}) \leq 1$. To this end we note that $\lambda_{\max}(\mathbf{Z})$ is continuous and monotonically increasing in λ. For high values of λ, $\lambda_{\max}(\mathbf{Z})$ will be larger than one. For $\lambda = 0$, we have that $\lambda_{\max}(\mathbf{Z}) \leq 1$. Indeed, since $\mathbf{A} \preceq \mathbf{I}$,

$$\mathbf{Q}_0^{-1}\mathbf{A}\mathbf{Q}_0^{-1} \preceq \mathbf{Q}_0^{-2}. \tag{44}$$

Using the fact that if $\mathbf{W} \preceq \mathbf{X}$ then $\mathbf{W}^{1/2} \preceq \mathbf{X}^{1/2}$ [35, Theroem 1.1], we conclude that $\mathbf{Z} \preceq \mathbf{I}$. Therefore, there exists a value λ_0 such that for $\lambda \leq \lambda_0$, we have $\lambda_{\max}(\mathbf{Z}) \leq 1$. Now, from (43) it follows that as the SNR increases, λ decreases. Thus, for high enough SNR values (40) will be satisfied for any choice of $\mathbf{T}$, $\mathbf{Q}_0$ and $\mathbf{A} \preceq \mathbf{I}$.

We also note that $\mathbf{B}$ of (38) is invariant to scaling of the matrix $\mathbf{T}$ and the constant U^2, *i.e.,* if $\mathbf{T}$ and U^2 are multiplied by a constant c, then $\mathbf{B}$ remains the same. This follows from the fact that from (39), $\lambda\mathbf{T}$ is invariant to this scaling.

We now prove Theorem 4.

Proof: To prove the theorem we need to show that $\mathbf{B}$ of (38) satisfies the optimality conditions of Theorem 2.

Since $\mathbf{Q}$, $\mathbf{T}$ and $\mathbf{A}$ are Hermitian, the first condition is satisfied. From the definition of λ we have that $\lambda\mathbf{T} + \mathbf{A} \succ 0$ so that $\mathbf{BQ} \prec \mathbf{I}$. The assumption (40) implies that

$$\mathbf{Q}^{1/2} \left(\mathbf{Q}^{-1}(\lambda\mathbf{T} + \mathbf{A})\mathbf{Q}^{-1}\right)^{1/2} \mathbf{Q}^{1/2} \preceq \mathbf{I}, \tag{45}$$

so that $\mathbf{QB} \succeq 0$. Now,

$$(\mathbf{I} - \mathbf{B})^*(\mathbf{I} - \mathbf{B}) = \lambda\mathbf{T} + \mathbf{A} \tag{46}$$

and conditions 3 and 4 are also satisfied. Conditions 5 and 6 follow from the fact that

$$\mathbf{B}(\mathbf{I} - \mathbf{B})^{-1} = \mathbf{Q}^{-1} \left(\mathbf{Q}(\lambda\mathbf{T} + \mathbf{A})^{-1}\mathbf{Q}\right)^{1/2} - \mathbf{I}, \tag{47}$$

so that

$$\begin{aligned} \mathrm{Tr}\left(\mathbf{B}(\mathbf{I} - \mathbf{B})^{-1}\mathbf{Q}^{-1}\mathbf{T}\right) &= -\mathrm{Tr}\left(\mathbf{Q}^{-1}\mathbf{T}\right) \\ &+ \mathrm{Tr}\left(\mathbf{Q}^{-1} \left(\mathbf{Q}(\lambda\mathbf{T} + \mathbf{A})^{-1}\mathbf{Q}\right)^{1/2} \mathbf{Q}^{-1}\mathbf{T}\right). \end{aligned} \tag{48}$$

It remains to show that there exists a $\lambda > 0$ satisfying (39). This follows from the fact that the left hand side of (39) is monotonically decreasing in λ, is equal to 0 for $\lambda \to \infty$ and is larger than the right hand side when $\lambda \to 0$. ■

A simplified special case of Theorem 4 arises when $\mathbf{A} = \beta\mathbf{T}$ for some $\beta \geq 0$:

Corollary 5: Consider the problem of Theorem 2 with $\mathbf{A} = \beta\mathbf{T}$ for some $0 < \beta \leq 1/\lambda_{\max}(\mathbf{T})$. Let

$$\mathbf{B} = \mathbf{I} - \alpha \left(\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1}\right)^{1/2} \mathbf{Q}, \tag{49}$$

where α is chosen as follows. If $\beta > 0$ and

$$\sqrt{\beta} \geq \frac{\mathrm{Tr}\left((\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1})^{1/2}\right)}{U^2 + \mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{T})}$$

then $\alpha = \sqrt{\beta}$. Otherwise,

$$\alpha = \frac{\mathrm{Tr}\left((\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1})^{1/2}\right)}{U^2 + \mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{T})}.$$

We then have that if

$$\lambda_{\min}\left(\mathbf{Q}^{-1}(\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q})^{1/2}\right) \geq \alpha$$

then $\widehat{\mathbf{G}} = \mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ with $\mathbf{B}$ given by (49).

B.3 Minimax MSE Estimator

Corollary 5 can be used to develop a closed form solution for the minimax MSE estimator of (12) for high SNR. Specifically, from (13) it follows that the minimax MSE estimator can be obtained as the solution to $\min_{\mathbf{G}} \max_{\mathbf{x}\in\mathcal{U}} \{\epsilon(\mathbf{G},\mathbf{x}) - \epsilon(\mathbf{G}_{\text{UB}})\}$, where $\mathbf{G}_{\text{UB}}\mathbf{H} = \mathbf{I}$. In this case, $\mathbf{A} = 0 = \beta\mathbf{T}$ with $\beta = 0$. From Corollary 5 we then have that if

$$\lambda_{\min}\left(\mathbf{Q}^{-1}(\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q})^{1/2}\right) \geq \alpha \tag{50}$$

where

$$\alpha = \frac{\operatorname{Tr}\left((\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1})^{1/2}\right)}{U^2 + \operatorname{Tr}(\mathbf{Q}^{-1}\mathbf{T})}, \tag{51}$$

then the optimal minimax MSE matrix $\mathbf{G}$ is

$$\mathbf{G} = (\mathbf{I} - \alpha\left(\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1}\right)^{1/2}\mathbf{Q})\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}. \tag{52}$$

Since $\mathbf{A} \preceq \mathbf{I}$, our general discussion implies that the solution (52) will be valid at high SNR for any $\mathbf{Q}_0$ and $\mathbf{T}$.

In [10] a closed form expression for the minimax MSE estimator was derived for the case of commuting matrices. The minimax MSE estimator for general matrices was developed under a certain condition by Pilz in [9], and coincides with the solution of [10] when the matrices commute. However, the solution of [9] is incorrect in general (the error occurs in the proof of Theorem 4), and leads to a larger worst-case MSE then our estimator (52). To demonstrate this, in Fig. 1 we plot the worst-case MSE as a function of the SNR defined by $-10\log\sigma^2$ where $\mathbf{C}_w = \sigma^2\mathbf{I}$ resulting from the minimax MSE estimator of (52) and that proposed in [9], for $m = 3$, $U^2 = 1$, $\mathbf{T} = \operatorname{diag}(1, 0.8, 0.3)$ and

$$\mathbf{H}^*\mathbf{H} = \begin{bmatrix} 1 & 0.6 & 0.4 \\ 0.6 & 1 & 0.6 \\ 0.4 & 0.6 & 1 \end{bmatrix}. \tag{53}$$

Clearly, the two estimators are different. Furthermore, our estimator leads to a smaller worst-case MSE, proving that the estimator of [9] cannot be minimax. A detailed discussion of the Pilz estimator and its relation with the minimax MSE estimator of (52) is provided in Appendix C.

C. SDP Formulation

We have seen in the previous section that in some special cases, Theorem 2 can be used to develop explicit solutions to (21). We now treat the general case, and show that the optimal $\mathbf{G}$ can be found numerically by solving an SDP [30], [31], which is the problem of minimizing a linear objective subject to linear matrix inequality constraints, *i.e.,* constraints of the form $\mathbf{Z}(\mathbf{x}) \succeq 0$, where the matrix $\mathbf{Z}$ depends linearly on $\mathbf{x}$.

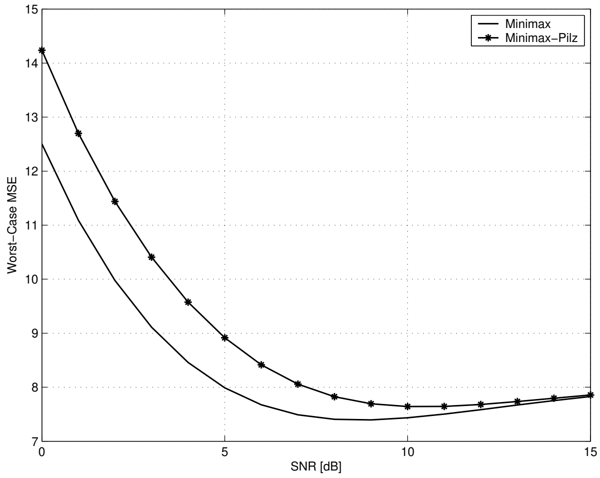

Fig. 1. Worst-case MSE in estimating $\mathbf{x}$ as a function of SNR using the minimax estimator of (52) and that proposed in [9].

The advantage of this formulation is that it readily lends itself to efficient computational methods which are guaranteed to converge to the global optimum.

The problem (21) was shown in Section IV to be equivalent to minimizing (29) subject to (28), which can be written as

$$\min_{\tau,\mathbf{B},\lambda\geq 0} \tau \tag{54}$$

subject to

$$\begin{aligned} \mathrm{Tr}\left(\mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{B}^*\right) + U^2\lambda &\leq \tau \\ \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})\mathbf{T}^{-1/2} &\preceq \lambda\mathbf{I} + \mathbf{T}^{-1/2}\mathbf{A}\mathbf{T}^{-1/2}. \end{aligned} \tag{55}$$

Let $\mathbf{b} = \mathrm{vec}((\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1/2}\mathbf{B}^*)$, where $\mathbf{m} = \mathrm{vec}(\mathbf{M})$ denotes the vector obtained by stacking the columns of $\mathbf{M}$. With this notation the constraints (55) become

$$\begin{aligned} \mathbf{b}^*\mathbf{b} + U^2\lambda &\leq \tau \\ \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})\mathbf{T}^{-1/2} &\preceq \lambda\mathbf{I} + \mathbf{T}^{-1/2}\mathbf{A}\mathbf{T}^{-1/2}. \end{aligned} \tag{56}$$

Next we exploit the following lemma [36, p. 28]:

Lemma 2 (Schur's complement) Let

$$\mathbf{M} = \begin{bmatrix} \mathbf{X} & \mathbf{Y}^* \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}$$

be a Hermitian matrix. Then $\mathbf{M} \succeq 0$ if and only if $\mathbf{Z} \succeq 0$, $\mathbf{X} - \mathbf{Y}^*\mathbf{Z}^\dagger\mathbf{Y} \succeq 0$ and $\mathbf{Y}^*(\mathbf{I} - \mathbf{Z}\mathbf{Z}^\dagger) = 0$. Equivalently, $\mathbf{M} \succeq 0$ if and only if $\mathbf{X} \succeq 0$, $\mathbf{Z} - \mathbf{Y}\mathbf{X}^\dagger\mathbf{Y}^* \succeq 0$ and $\mathbf{Y}(\mathbf{I} - \mathbf{X}\mathbf{X}^\dagger) = 0$.

From Lemma 2 it follows that the constraints (56) are equivalent to the linear matrix inequalities

$$\begin{bmatrix} \tau - U^2\lambda & \mathbf{b}^* \\ \mathbf{b} & \mathbf{I} \end{bmatrix} \succeq 0$$

$$\begin{bmatrix} \lambda\mathbf{I} + \mathbf{T}^{-1/2}\mathbf{A}\mathbf{T}^{-1/2} & \mathbf{T}^{-1/2}(\mathbf{I}-\mathbf{B})^* \\ (\mathbf{I}-\mathbf{B})\mathbf{T}^{-1/2} & \mathbf{I} \end{bmatrix} \succeq 0. \tag{57}$$

The problem of (21) can thus be formulated as the SDP $\min_{\tau,\mathbf{B},\lambda\geq 0} \tau$ subject to (57). Denoting the optimal τ and $\mathbf{B}$ by $\hat{\tau}$ and $\widehat{\mathbf{B}}$ we have that $\mathcal{V}(\mathbf{G}_0) = \hat{\tau} - \mathrm{Tr}(\mathbf{G}_0\mathbf{C}_w\mathbf{G}_0^*)$, and $\widehat{\mathbf{G}} = \widehat{\mathbf{B}}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$.

V. Admissible Estimators on an Ellipsoid

We now use Theorem 2 to develop necessary and sufficient conditions on $\hat{\mathbf{x}}_0$ to be admissible on $\mathcal{U}$ of (14).

Theorem 6: Let $\hat{\mathbf{x}}_0 = \mathbf{G}_0\mathbf{y}$ be a given linear estimate of $\mathbf{x}$ in the model (1), let $\mathbf{B} = \mathbf{G}_0\mathbf{H}$ and let $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. Then $\hat{\mathbf{x}}_0$ is admissible on $\mathcal{U} = \{\mathbf{x} \in \mathbb{C}^m | \mathbf{x}^*\mathbf{T}\mathbf{x} \leq U^2\}$ if and only if $\mathbf{G}_0 = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ for a matrix $\mathbf{B}$ satisfying:

1. $\mathbf{Q}\mathbf{B} = \mathbf{B}^*\mathbf{Q}$;
2. $0 \preceq \mathbf{Q}\mathbf{B} \prec \mathbf{Q}$;
3. $\mathrm{Tr}\left(\mathbf{B}(\mathbf{I}-\mathbf{B})^{-1}\mathbf{Q}^{-1}\mathbf{T}\right) \leq U^2$.

The admissibility conditions of the theorem can be shown to be equivalent to those derived in [20], [22]. However, our method of proof is different. Since our proof is direct and straightforward, we include it here for completeness.

Proof: To prove the theorem it is sufficient to show that $\mathbf{G}_0$ is the optimal solution to (20), or equivalently, that $\mathbf{B} = \mathbf{G}_0\mathbf{H}$ satisfies the conditions of Theorem 2 with $\mathbf{A} = (\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})$ for some $\lambda \geq 0$ if and only if it satisfies conditions 1-3.

Since conditions 1-3 are a subset of the conditions of Theorem 2, it follows immediately that any $\mathbf{B}$ satisfying Theorem 2 also satisfies conditions 1-3. On the other hand, if $\mathbf{B}$ satisfies conditions 1-3, then it also satisfies the conditions of Theorem 2 with $\lambda = 0$, completing the proof. ∎

Given an admissible estimator $\hat{\mathbf{x}}_0$, we can construct a whole class of admissible estimators by simply scaling $\hat{\mathbf{x}}_0$, as incorporated in the following proposition.

Proposition 1: If $\hat{\mathbf{x}}_0$ is admissible on $\mathcal{U}$ of (14), then $\alpha\hat{\mathbf{x}}_0$ is admissible for any $0 \leq \alpha < 1$.

Proof: From Theorem 6, $\hat{\mathbf{x}}_0 = \mathbf{B}_0\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}$ for some $\mathbf{B}_0$, so that $\hat{\mathbf{x}} = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}$ where $\mathbf{B} = \alpha\mathbf{B}_0$. Since $\mathbf{Q}\mathbf{B}_0 = \mathbf{B}_0^*\mathbf{Q}$, we have immediately that $\mathbf{Q}\mathbf{B} = \mathbf{B}^*\mathbf{Q}$. Similarly, since $0 \leq \alpha < 1$, the fact that $0 \preceq \mathbf{Q}\mathbf{B}_0 \prec \mathbf{Q}$ implies that $0 \preceq \mathbf{Q}\mathbf{B} \prec \mathbf{Q}$. It remains to show that if $\mathbf{B}_0$ satisfies condition 3, then so does $\alpha\mathbf{B}_0$. Now,

$$\alpha\mathbf{B}_0(\mathbf{I}-\alpha\mathbf{B}_0)^{-1}\mathbf{Q}^{-1} = (\mathbf{Q}-\alpha\mathbf{Q}\mathbf{B}_0)^{-1} - \mathbf{Q}^{-1}. \tag{58}$$

Since $0 \leq \alpha < 1$, and $\mathbf{Q} - \alpha\mathbf{Q}\mathbf{B}_0$ is Hermitian,

$$\mathbf{Q} - \alpha\mathbf{Q}\mathbf{B}_0 \succeq \mathbf{Q} - \mathbf{Q}\mathbf{B}_0 \succ 0, \tag{59}$$

so that

$$(\mathbf{Q} - \alpha\mathbf{Q}\mathbf{B}_0)^{-1} \preceq (\mathbf{Q} - \mathbf{Q}\mathbf{B}_0)^{-1}. \tag{60}$$

Therefore,

$$\begin{aligned} \alpha\mathbf{B}_0(\mathbf{I} - \alpha\mathbf{B}_0)^{-1}\mathbf{Q}^{-1} &\preceq (\mathbf{Q} - \mathbf{Q}\mathbf{B}_0)^{-1} - \mathbf{Q}^{-1} \\ &= \mathbf{B}_0(\mathbf{I} - \mathbf{B}_0)^{-1}\mathbf{Q}^{-1}. \end{aligned} \tag{61}$$

We now rely on the following lemma (see, *e.g.*, [37]):

Lemma 3: Let $\mathbf{T}, \mathbf{W}$ and $\mathbf{M}$ be Hermitian nonnegative definite matrices with $\mathbf{W} \preceq \mathbf{T}$. Then $\mathrm{Tr}(\mathbf{M}\mathbf{W}) \leq \mathrm{Tr}(\mathbf{M}\mathbf{T})$.

Using Lemma 3 together with (61) we have that

$$\begin{aligned} &\mathrm{Tr}\left(\alpha\mathbf{B}_0(\mathbf{I} - \alpha\mathbf{B}_0)^{-1}\mathbf{Q}^{-1}\mathbf{T}\right) \leq \\ &\qquad \mathrm{Tr}\left(\mathbf{B}_0(\mathbf{I} - \mathbf{B}_0)^{-1}\mathbf{Q}^{-1}\mathbf{T}\right) \leq U^2, \end{aligned} \tag{62}$$

completing the proof. ∎

Another general class of admissible estimators is given in the following proposition.

Proposition 2: Consider the problem of Theorem 6. Let $\mathbf{Q} = \mathbf{V}\Sigma\mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ with $\sigma_i > 0$. Then $\hat{\mathbf{x}} = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ is admissible for any $\mathbf{B} = \mathbf{V}\Delta\mathbf{V}^*$ where $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_m)$ and the values δ_i satisfy

$$\begin{gathered} 0 \leq \delta_i < 1, \quad 1 \leq i \leq m; \\ \textstyle\sum_{i=1}^m \frac{z_i\delta_i}{(1-\delta_i)\sigma_i} \leq U^2, \end{gathered}$$

where z_i is the ith diagonal element of $\mathbf{V}^*\mathbf{T}\mathbf{V}$.

Proof: To prove the proposition we need to show that $\mathbf{B} = \mathbf{V}\Delta\mathbf{V}^*$ satisfies the conditions of Theorem 6. The first two conditions are trivially satisfied. The last condition follows from the fact that

$$\begin{aligned} \mathrm{Tr}\left(\mathbf{B}(\mathbf{I} - \mathbf{B})^{-1}\mathbf{Q}^{-1}\mathbf{T}\right) &= \mathrm{Tr}\left(\Delta(\mathbf{I} - \Delta)^{-1}\Sigma^{-1}\mathbf{V}^*\mathbf{T}\mathbf{V}\right) \\ &= \sum_{i=1}^m \frac{z_i\delta_i}{(1-\delta_i)\sigma_i}. \end{aligned} \tag{63}$$

∎

VI. Examples

We now consider some examples of admissible and strictly dominating estimators on the ellipsoidal constraint set of (14).

A. Tikhonov Estimation

A popular estimator in applications is the Tikhonov estimator [5], [4], which has the form

$$\hat{\mathbf{x}}_{\mathrm{TIK}} = (\mathbf{H}\mathbf{C}_w^{-1}\mathbf{H} + \mathbf{M})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}, \tag{64}$$

for some $\mathbf{M} \succ 0$. This estimator can be obtained as the solution to the regularized least-squares problem

$$\min_{\hat{\mathbf{x}}} \left\{ (\mathbf{y} - \mathbf{H}\hat{\mathbf{x}})^*\mathbf{C}_w^{-1}(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}) + \hat{\mathbf{x}}^*\mathbf{M}\hat{\mathbf{x}} \right\}. \tag{65}$$

Using Theorem 6 we can develop conditions on $\mathbf{M}$ such that $\hat{\mathbf{x}}_{\mathrm{TIK}}$ is admissible:

Proposition 3: The Tikhonov estimator is admissible on $\mathcal{U} = \{\mathbf{x} \in \mathbb{C}^m | \mathbf{x}^*\mathbf{T}\mathbf{x} \leq U^2\}$ if and only if $\mathrm{Tr}(\mathbf{T}\mathbf{M}^{-1}) \leq U^2$.

Proof: To prove the proposition we need to show that the conditions of Theorem 6 are satisfied if and only if $\mathrm{Tr}(\mathbf{T}\mathbf{M}^{-1}) \leq U^2$.

Denoting $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$, we can express $\hat{\mathbf{x}}_{\mathrm{TIK}}$ as $\hat{\mathbf{x}}_{\mathrm{TIK}} = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$, with $\mathbf{B} = (\mathbf{Q} + \mathbf{M})^{-1}\mathbf{Q}$. We have immediately that $\mathbf{Q}\mathbf{B} = \mathbf{B}^*\mathbf{Q}$ for any choice of $\mathbf{M}$. Next, from the fact that $\mathbf{Q} + \mathbf{M} \succ 0$,

$$\mathbf{Q}\mathbf{B} = \mathbf{Q}(\mathbf{Q} + \mathbf{M})^{-1}\mathbf{Q} \succ 0. \tag{66}$$

Finally, since $(\mathbf{Q}+\mathbf{M})^{-1} \prec \mathbf{Q}^{-1}$ we have that $\mathbf{Q}\mathbf{B} \prec \mathbf{Q}$. Therefore, $\hat{\mathbf{x}}_{\mathrm{TIK}}$ is admissible if and only if it satisfies the trace condition (condition 3).

Now,

$$\mathbf{I} - \mathbf{B} = \mathbf{I} - (\mathbf{Q} + \mathbf{M})^{-1}\mathbf{Q} = (\mathbf{Q} + \mathbf{M})^{-1}\mathbf{M}. \tag{67}$$

Therefore,

$$\mathbf{B}(\mathbf{I} - \mathbf{B})^{-1} = (\mathbf{I} + \mathbf{Q}^{-1}\mathbf{M})^{-1}(\mathbf{M}^{-1}\mathbf{Q} + \mathbf{I}) = \mathbf{M}^{-1}\mathbf{Q}, \tag{68}$$

where the last equality can be verified by multiplying both sides of the equation by $\mathbf{I} + \mathbf{Q}^{-1}\mathbf{M}$. Using (68),

$$\mathrm{Tr}\left(\mathbf{B}(\mathbf{I} - \mathbf{B})^{-1}\mathbf{Q}^{-1}\mathbf{T}\right) = \mathrm{Tr}\left(\mathbf{M}^{-1}\mathbf{T}\right), \tag{69}$$

completing the proof. ∎

Suppose now that $\mathbf{M}$ does not satisfy the constraint $\mathrm{Tr}\left(\mathbf{M}^{-1}\mathbf{T}\right) \leq U^2$, so that $\hat{\mathbf{x}}_{\mathrm{TIK}}$ is inadmissible. To obtain an admissible estimator, one approach is to modify $\hat{\mathbf{x}}_{\mathrm{TIK}}$ by multiplying $\mathbf{M}$ by a scalar α satisfying $\alpha \geq \mathrm{Tr}(\mathbf{T}\mathbf{M}^{-1})/U^2$, resulting in the admissible Tikhonov estimator

$$\hat{\mathbf{x}}_{\mathrm{TIK}}^{\mathrm{ADM}} = \left(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H} + \alpha\mathbf{M}\right)^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}. \tag{70}$$

Note, however, that we are not guaranteed that the estimator (70) dominates the original Tikhonov estimator $\hat{\mathbf{x}}_{\mathrm{TIK}}$. Indeed, in Figs. 2 and 3 we plot the MSE of $\hat{\mathbf{x}}_{\mathrm{TIK}}$ and $\hat{\mathbf{x}}_{\mathrm{TIK}}^{\mathrm{ADM}}$ for $U^2 = .5$ and $U^2 = 2$ respectively, as a function of the SNR defined by $-10\log\sigma^2$, where $\mathbf{C}_w = \sigma^2\mathbf{I}$. The parameters are chosen as $m = 3$, $\mathbf{T} = \mathbf{I}$,

$\mathbf{M} = (0.6m/U^2)\mathbf{I}$,

$$\mathbf{H}^*\mathbf{H} = \begin{bmatrix} 1 & 0.6 & 0.4 \\ 0.6 & 1 & 0.6 \\ 0.4 & 0.6 & 1 \end{bmatrix}, \tag{71}$$

and $\mathbf{x} = U^2[1\ 1\ 1]/\sqrt{3}$. It is evident from the figures that $\hat{\mathbf{x}}_{\text{TIK}}^{\text{ADM}}$ does not dominate $\hat{\mathbf{x}}_{\text{TIK}}$.

An estimator that is admissible and strictly dominates $\hat{\mathbf{x}}_{\text{TIK}}$ can be obtained from Theorem 1 as $\hat{\mathbf{x}} = \widehat{\mathbf{G}}\mathbf{y}$ where

$$\widehat{\mathbf{G}} = \arg\min_{\mathbf{G}} \max_{\mathbf{x}\in\mathcal{U}} \{\epsilon(\mathbf{G},\mathbf{x}) - \epsilon(\mathbf{G}_0,\mathbf{x})\} \tag{72}$$

with $\mathbf{G}_0 = (\mathbf{Q}+\mathbf{M})^{-1}\mathbf{Q}\mathbf{H}^*\mathbf{C}_w^{-1}$. We refer to this estimator as the minimax admissible estimator. In the case in which $\mathbf{T} = \mathbf{I}$ and $\mathbf{M} = \beta\mathbf{I}$ for some $\beta > 0$, we have that $\mathbf{B} = \mathbf{G}_0\mathbf{H}$, $\mathbf{T}$ and $\mathbf{Q}$ have the same eigenvector matrix. We can then use Theorem 3 to obtain a closed form solution for the minimax admissible estimator. It can be seen from the theorem that the resulting estimator is no longer a Tikhonov estimator in general.

The MSE resulting from the minimax admissible estimator is plotted as a function of the SNR in Figs. 2 and 3. As can be seen from the figures, this estimator dominates the original Tikhonov estimator for all SNR.

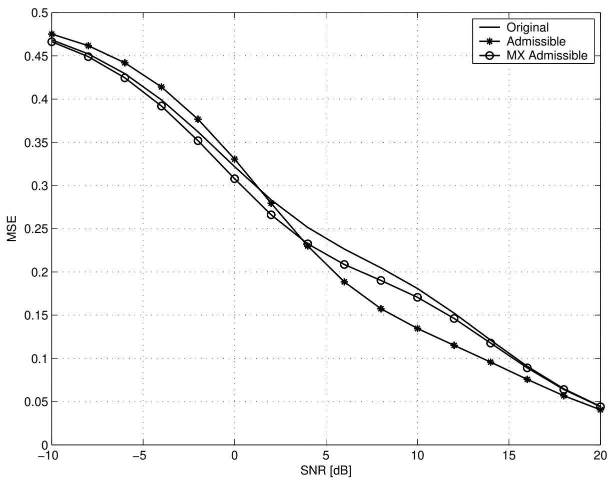

Fig. 2. MSE in estimating $\mathbf{x}$ as a function of SNR using the Tikhonov, admissible Tikhonov and minimax admissible estimators, for $U^2 = 0.5$.

B. Shrunken Estimator

Another popular estimator is the shrunken estimator [6], which is a shrinkage of the least-squares estimator, and is given by

$$\hat{\mathbf{x}}_{\text{SH}} = \alpha\hat{\mathbf{x}}_{\text{LS}} = \alpha(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y} \tag{73}$$

for some $0 \le \alpha < 1$. For this estimator, $\mathbf{B} = \alpha\mathbf{I}$, and therefore it follows from Theorem 6 that $\hat{\mathbf{x}}_{\text{SH}}$ is admissible if and only if $0 \le \alpha \le U^2/(U^2 + \text{Tr}(\mathbf{Q}^{-1}\mathbf{T}))$.

We now use Theorem 3 to find a strictly dominating and admissible estimator in the case in which $\mathbf{T} = \mathbf{I}$ and $\alpha > U^2/(U^2 + \text{Tr}(\mathbf{Q}^{-1}))$. Since $\delta_i = \alpha$, $\eta_i = \eta$ is independent of i and is given by $\eta = \lambda + (1-\alpha)^2$. Thus,

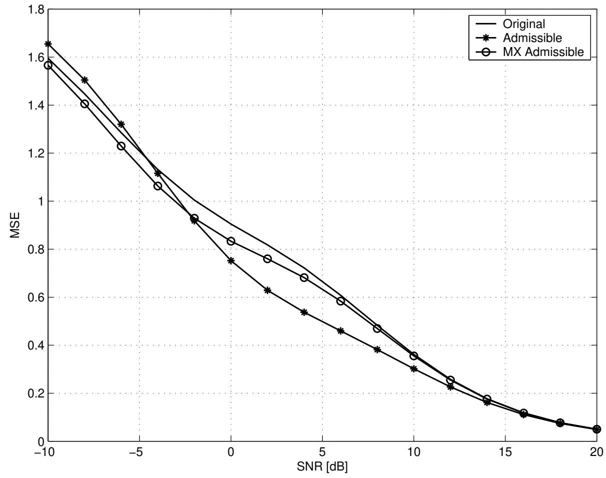

Fig. 3. MSE in estimating $\mathbf{x}$ as a function of SNR using the Tikhonov, the admissible Tikhonov and the minimax admissible estimators, for $U^2 = 2$.

λ is chosen such that

$$\left(\frac{1}{\sqrt{\eta}} - 1\right) \sum_{i=1}^{m} \frac{1}{\sigma_i} = U^2, \tag{74}$$

which results in

$$\sqrt{\eta} = \frac{\mathrm{Tr}(\mathbf{Q}^{-1})}{U^2 + \mathrm{Tr}(\mathbf{Q}^{-1})}. \tag{75}$$

We conclude that the minimax admissible estimator is always a shrunken estimator with shrinkage

$$\beta = 1 - \sqrt{\eta} = \frac{U^2}{U^2 + \mathrm{Tr}(\mathbf{Q}^{-1})}. \tag{76}$$

The resulting shrunken estimator is equivalent to the minimax MSE estimator derived in [10] for the case $\mathbf{T} = \mathbf{I}$.

VII. Admissible and Dominating Estimators on The Entire Space

We now treat the case in which $\mathcal{U} = \mathbb{C}^m$, so that $\mathbf{x}$ is not restricted. The main difference between this setting and the ellipsoidal constraint case is that, as we will see below, when $\mathbf{x}$ is not restricted the optimization problem we end up with is not strictly feasible. Therefore, some manipulations are necessary in order to develop a solution.

A. Dominating Estimators

Given an inadmissible estimator $\hat{\mathbf{x}}_0$, we can use Theorem 1 to construct an admissible estimator $\hat{\mathbf{x}}$ strictly dominating $\hat{\mathbf{x}}_0$ on $\mathbb{C}^m$ by solving the problem

$$\min_{\mathbf{G}} \left\{ \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) + \sup_{\mathbf{x}} \mathbf{x}^*\mathbf{Z}\mathbf{x} \right\} \tag{77}$$

where $\mathbf{Z}$ is defined in (16). Now,

$$\sup_{\mathbf{x}} \mathbf{x}^*\mathbf{Z}\mathbf{x} = \begin{cases} 0, & \mathbf{Z} \preceq 0; \\ \infty, & \text{otherwise.} \end{cases} \tag{78}$$

Therefore, (77) is equivalent to

$$\min_{\mathbf{G}} \operatorname{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) \tag{79}$$

subject to

$$(\mathbf{I}-\mathbf{G}\mathbf{H})^*(\mathbf{I}-\mathbf{G}\mathbf{H}) \preceq (\mathbf{I}-\mathbf{G}_0\mathbf{H})^*(\mathbf{I}-\mathbf{G}_0\mathbf{H}) \stackrel{\triangle}{=} \mathbf{A}. \tag{80}$$

Using arguments similar to Lemma 1 we can show that the optimal solution has the form $\widehat{\mathbf{G}} = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$, where $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_w\mathbf{H}$ and $\mathbf{B}$ is the solution to

$$\min_{\mathbf{B}} \operatorname{Tr}(\mathbf{B}\mathbf{Q}^{-1}\mathbf{B}^*) \tag{81}$$

subject to

$$(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) \preceq \mathbf{A}. \tag{82}$$

The constraint (82) is not strictly feasible if $\mathbf{A}$ is rank deficient, so that we cannot apply the KKT conditions. To obtain a strictly feasible problem we note that using Lemma 2, (82) is equivalent to

$$\begin{aligned} (\mathbf{I}-\mathbf{B})\mathbf{A}^\dagger(\mathbf{I}-\mathbf{B})^* &\preceq \mathbf{I}; \\ (\mathbf{I}-\mathbf{B})(\mathbf{I}-\mathbf{A}\mathbf{A}^\dagger) &= 0. \end{aligned} \tag{83}$$

Denoting by $\mathbf{P}_{\mathcal{N}(\mathbf{A})}$ the orthogonal projection onto the null space $\mathcal{N}(\mathbf{A})$ of $\mathbf{A}$, the constraints (83) can be written as

$$\begin{aligned} \mathbf{A}^{\dagger/2}(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})\mathbf{A}^{\dagger/2} &\preceq \mathbf{I}; \\ (\mathbf{I}-\mathbf{B})\mathbf{P}_{\mathcal{N}(\mathbf{A})} &= 0, \end{aligned} \tag{84}$$

where we used the notation $\mathbf{A}^{\dagger/2} = (\mathbf{A}^\dagger)^{1/2}$. From (84) it follows that $\mathbf{I}-\mathbf{B} = \mathbf{X}\mathbf{P}_{\mathcal{R}}$ for some matrix $\mathbf{X}$, where $\mathbf{P}_{\mathcal{R}} \stackrel{\triangle}{=} \mathbf{P}_{\mathcal{R}(\mathbf{A})} = \mathbf{I}-\mathbf{P}_{\mathcal{N}(\mathbf{A})}$ is the orthogonal projection onto the range space $\mathcal{R}(\mathbf{A})$ of $\mathbf{A}$. Substituting

$$\mathbf{B} = \mathbf{I}-\mathbf{X}\mathbf{P}_{\mathcal{R}} \tag{85}$$

into (81) and (84), our problem becomes

$$\min_{\mathbf{X}} \operatorname{Tr}\left((\mathbf{X}\mathbf{P}_{\mathcal{R}}-\mathbf{I})\mathbf{Q}^{-1}(\mathbf{X}\mathbf{P}_{\mathcal{R}}-\mathbf{I})^*\right) \tag{86}$$

subject to

$$\mathbf{A}^{\dagger/2}\mathbf{X}^*\mathbf{X}\mathbf{A}^{\dagger/2} \preceq \mathbf{I}, \tag{87}$$

which is a strictly feasible convex optimization problem. Therefore, we can now use the KKT conditions to find optimality conditions, resulting in the following theorem.

Theorem 7: Consider the problem of Theorem 1. Let $\hat{\mathbf{x}}_0 = \mathbf{G}_0\mathbf{y}$ be a given linear estimate of $\mathbf{x}$, let $\mathbf{A} = (\mathbf{I} - \mathbf{G}_0\mathbf{H})^*(\mathbf{I} - \mathbf{G}_0\mathbf{H})$ and $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. Denote the orthogonal projection onto $\mathcal{R}(\mathbf{A})$ by $\mathbf{P}_{\mathcal{R}}$ and let $\mathbf{M} = (\mathbf{Q}(\mathbf{I} - \mathbf{B}))^\dagger - \mathbf{P}_{\mathcal{R}}\mathbf{Q}^{-1}\mathbf{P}_{\mathcal{R}}$. Then,

$$\widehat{\mathbf{G}} = \min_{\mathbf{G}} \sup_{\mathbf{x}} \{\epsilon(\mathbf{G}, \mathbf{x}) - \epsilon(\mathbf{G}_0, \mathbf{x})\}$$

has the form $\widehat{\mathbf{G}} = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ where $\mathbf{B}$ satisfies

1. $\mathbf{Q}\mathbf{B} = \mathbf{B}^*\mathbf{Q}$;
2. $0 \preceq \mathbf{Q}\mathbf{B} \preceq \mathbf{Q}$;
3. $(\mathbf{I} - \mathbf{B})\mathbf{P}_{\mathcal{R}} = \mathbf{I} - \mathbf{B}$;
4. $(\mathbf{I} - \mathbf{B})^*(\mathbf{I} - \mathbf{B}) \preceq \mathbf{A}$;
5. $(\mathbf{I} - \mathbf{B})^*(\mathbf{I} - \mathbf{B})\mathbf{M} = \mathbf{A}\mathbf{M}$.

Proof: see Appendix D. ■

In Theorems 8 and 9 below we use the conditions of Theorem 7 to derive explicit closed form solutions for the same special cases treated in Section IV-B. For arbitrary choices of $\mathbf{A}$ and $\mathbf{Q}$ an SDP formulation similar to that presented in Section IV-C can be used.

Theorem 8: Consider the problem of Theorem 7. Let $\mathbf{Q} = \mathbf{V}\Sigma\mathbf{V}^*$ where $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_m)$ with $\sigma_i > 0$, and suppose that $\mathbf{G}_0\mathbf{H} = \mathbf{V}\Delta\mathbf{V}^*$ where $\Delta = \text{diag}(\delta_1, \ldots, \delta_m)$. Then

$$\widehat{\mathbf{G}} = \mathbf{V}\mathbf{D}\mathbf{V}^*(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$$

where $\mathbf{D} = \text{diag}(d_1, \ldots, d_m)$ with

$$d_i = \begin{cases} 1 - |1 - \delta_i|, & |1 - \delta_i| < 1; \\ 0, & |1 - \delta_i| \geq 1. \end{cases} \tag{88}$$

Note that this theorem is analogous to Theorem 3 where here we replace λ by 0.

Proof: To prove Theorem 8 we need to show that $\mathbf{B} = \mathbf{V}\mathbf{D}\mathbf{V}^*$ with d_i given by (88) satisfies the optimality conditions of Theorem 7. Using the fact that $\mathbf{Q} = \mathbf{V}\Sigma\mathbf{V}^*$, $\mathbf{G}_0\mathbf{H} = \mathbf{V}\Delta\mathbf{V}^*$ and $\Sigma \succ 0$ we have that $\mathbf{P}_{\mathcal{R}} = \mathbf{V}\mathbf{Z}\mathbf{V}^*$ where $\mathbf{Z} = \text{diag}(z_1, \ldots, z_m)$ with

$$z_i = \begin{cases} 1, & \delta_i \neq 1; \\ 0, & \delta_i = 1, \end{cases} \tag{89}$$

and $\mathbf{M} = \mathbf{V}\Sigma^{-1}((\mathbf{I} - \mathbf{D})^\dagger - \mathbf{Z})\mathbf{V}^*$. The conditions then become:

1. $d_i = d_i^*$;
2. $0 \leq d_i \leq 1$;
3. $d_i = 1$ for all i such that $\delta_i = 1$;
4. $(1 - d_i)^2 \leq |1 - \delta_i|^2$;
5. $(1 - d_i)^2 = |1 - \delta_i|^2$ for all $\{i : d_i \neq 0, \delta_i \neq 1\}$,

which are clearly satisfied. ■

Next we consider non-commuting matrices with bounded eigenvalues:

Theorem 9: Consider the problem of Theorem 7, and let

$$\mathbf{B} = \mathbf{I} - \left(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}^{-1}\right)^{1/2}\mathbf{Q}. \tag{90}$$

We then have that if

$$\lambda_{\max}\left(\mathbf{Q}\left(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}^{-1}\right)^{1/2}\right) \leq 1 \tag{91}$$

then $\widehat{\mathbf{G}} = \mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ with $\mathbf{B}$ given by (90).

Proof: To prove the theorem we need to show that $\mathbf{B}$ given by (90) satisfies the optimality conditions of Theorem 7.

Since $\mathbf{Q}$ and $\mathbf{A}$ are Hermitian, the first condition is satisfied. Furthermore, using the fact that $\mathbf{A} \succeq 0$ we have that $\mathbf{QB} \preceq \mathbf{I}$. The assumption (91) implies that

$$\mathbf{Q}^{1/2}\left(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}^{-1}\right)^{1/2}\mathbf{Q}^{1/2} \preceq \mathbf{I}, \tag{92}$$

so that $\mathbf{QB} \succeq 0$. Now,

$$(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) = \mathbf{A} \tag{93}$$

and conditions 4 and 5 are also satisfied. To prove condition 3 we need to show that $(\mathbf{I}-\mathbf{B})\mathbf{z} = \left(\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}^{-1}\right)^{1/2}\mathbf{Q}\mathbf{z} = 0$ for any $\mathbf{z} \in \mathcal{N}(\mathbf{A})$; this is equivalent to $\mathbf{Q}^{-1}\mathbf{A}\mathbf{Q}^{-1}\mathbf{Q}\mathbf{z} = 0$ which is clearly satisfied, completing the proof. ■

B. Admissible Estimators

We now use Theorem 7 to develop necessary and sufficient conditions on $\hat{\mathbf{x}}_0$ to be admissible for all $\mathbf{x}$.

Theorem 10: Let $\hat{\mathbf{x}}_0 = \mathbf{G}_0\mathbf{y}$ be a given linear estimate of $\mathbf{x}$ in the model (1), let $\mathbf{B} = \mathbf{G}_0\mathbf{H}$ and let $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. Then $\hat{\mathbf{x}}_0$ is admissible on $\mathbb{C}^m$ if and only if $\mathbf{G}_0 = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ for a matrix $\mathbf{B}$ satisfying

1. $\mathbf{QB} = \mathbf{B}^*\mathbf{Q}$;
2. $0 \preceq \mathbf{QB} \preceq \mathbf{Q}$.

Proof: The theorem follows by noting that $\mathbf{B} = \mathbf{G}_0\mathbf{H}$ satisfies the conditions of Theorem 7 where $\mathbf{A} = (\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})$ if and only if it satisfies conditions 1-2. ■

A general class of admissible estimators is given in the following proposition.

Proposition 4: Consider the problem of Theorem 10. Let $\mathbf{Q} = \mathbf{V}\Sigma\mathbf{V}^*$, where $\mathbf{V}$ is a unitary matrix and $\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$ with $\sigma_i > 0$. Then $\hat{\mathbf{x}} = \mathbf{B}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}$ is admissible for any $\mathbf{B} = \mathbf{V}\Delta\mathbf{V}^*$ where $\Delta = \mathrm{diag}(\delta_1, \ldots, \delta_m)$ and the values δ_i satisfy $0 \leq \delta_i \leq 1$.

Proof: The proof follows by showing that the resulting estimator satisfies the conditions of Theorem 10. ■

VIII. Conclusion

We addressed the important issue of comparing the MSE performance of linear estimators by considering a unified approach for developing admissible estimators strictly dominating a given linear estimator, and for characterizing the admissibility of linear estimators. The results can be applied to arbitrary constraint sets on the parameter vector to be estimated, and general error measures. As we showed, the admissibility of a linear estimator as well as the construction of a linear estimator strictly dominating an inadmissible strategy can both be treated by considering a certain convex optimization problem. The machinery of convex optimization can then be utilized to analyze the MSE performance for specific uncertainty sets.

Our results can be used in practice to help select an appropriate estimator from the multitude of methods proposed in the literature. We also suggest methods for designing new estimators with lower MSE in the case in which the conventional approaches are inadmissible.

An interesting extension of this work is to apply our general procedure to other parameter sets that appear in applications. For example, in an image processing context the elements x_i of $\mathbf{x}$, which represent pixel values, are always nonnegative so that $\mathcal{U} = \{\mathbf{x} \in \mathbb{C}^m | x_i \geq 0\}$.

Appendix

I. Proof of Lemma 1

To prove the lemma we note that the objective in (21) depends on $\mathbf{G}$ only through $\mathbf{GH}$ and $\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$. For any choice of $\mathbf{G}$,

$$\begin{aligned}\mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*) &= \mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*)\\ &\quad+\mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}(\mathbf{I}-\mathbf{P})\mathbf{C}_w^{1/2}\mathbf{G}^*)\\ &\geq \mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}^*), \end{aligned} \tag{94}$$

where

$$\mathbf{P} = \mathbf{C}_w^{-1/2}\mathbf{H}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1/2} \tag{95}$$

is the orthogonal projection onto $\mathcal{R}(\mathbf{C}_w^{-1/2}\mathbf{H})$. In addition, $\mathbf{GH} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2}\mathbf{H}$ since $\mathbf{P}\mathbf{C}_w^{-1/2}\mathbf{H} = \mathbf{C}_w^{-1/2}\mathbf{H}$. Now, suppose that $\mathbf{G}$ is an optimal solution to (21), and let $\mathbf{G}' = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2}$. Then $\mathbf{G}'\mathbf{H} = \mathbf{GH}$ and $\mathrm{Tr}(\mathbf{G}'\mathbf{C}_w\mathbf{G}'^*) = \mathrm{Tr}(\mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{1/2}\mathbf{G}) \leq \mathrm{Tr}(\mathbf{G}\mathbf{C}_w\mathbf{G}^*)$. Since from Theorem 1 the solution is unique we must have that $\mathbf{G} = \mathbf{G}'$, or $\mathbf{G} = \mathbf{G}\mathbf{C}_w^{1/2}\mathbf{P}\mathbf{C}_w^{-1/2}$. Using (95),

$$\begin{aligned}\mathbf{G} &= \mathbf{GH}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\\ &= \mathbf{B}(\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H})^{-1}\mathbf{H}^*\mathbf{C}_w^{-1},\end{aligned} \tag{96}$$

for some $m \times m$ matrix $\mathbf{B}$. The proof of the lemma then follows from substituting $\mathbf{G}$ of (96) into (21).

II. KKT Conditions

In this appendix, we show that the KKT conditions can be reduced to the conditions of Theorem 2. To this end, we express the Lagrange multiplier Π as a function of $\mathbf{B}$, and then translate the conditions on Π to conditions on $\mathbf{B}$.

Differentiating $\mathcal{L}$ with respect to λ and equating to 0,

$$\mathrm{Tr}(\Pi) = U^2 - \mu. \tag{97}$$

Since $\mu \geq 0$, (97) implies that $\mathrm{Tr}(\Pi) \leq U^2$.

Differentiating $\mathcal{L}$ with respect to $\mathbf{B}$ and equating to 0,

$$\mathbf{B} = \mathbf{T}^{-1/2}\Pi\mathbf{T}^{-1/2}\left(\mathbf{Q}^{-1} + \mathbf{T}^{-1/2}\Pi\mathbf{T}^{-1/2}\right)^{-1}, \tag{98}$$

where $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$. Note, that since $\mathbf{Q} \succ 0$ and $\mathbf{T}^{-1/2}\Pi\mathbf{T}^{-1/2} \succeq 0$, the inverse in (98) is always defined. With $\mathbf{B}$ given by (98),

$$\mathbf{I} - \mathbf{B} = \left(\mathbf{I} + \widetilde{\Pi}\mathbf{Q}\right)^{-1}, \tag{99}$$

where we denoted $\widetilde{\Pi} = \mathbf{T}^{-1/2}\Pi\mathbf{T}^{-1/2}$ and used the matrix inversion lemma [37]. To show that the inverse exists, suppose to the contrary that there exists a $\mathbf{z} \neq 0$ such that

$$(\mathbf{I} + \widetilde{\Pi}\mathbf{Q})\mathbf{z} = 0. \tag{100}$$

Then, multiplying both sides of (100) by $\mathbf{z}^*\mathbf{Q}$,

$$\mathbf{z}^*\mathbf{Q}\widetilde{\Pi}\mathbf{Q}\mathbf{z} = -\mathbf{z}^*\mathbf{Q}\mathbf{z}. \tag{101}$$

Since $\widetilde{\Pi} \succeq 0$ and $\mathbf{Q}\widetilde{\Pi}\mathbf{Q} \succeq 0$, (101) implies that $\mathbf{z}^*\mathbf{Q}\widetilde{\Pi}\mathbf{Q}\mathbf{z} = \mathbf{z}^*\mathbf{Q}\mathbf{z} = 0$, which is impossible since $\mathbf{Q} \succ 0$.

Taking the inverse on both sides of (99) and using the matrix inversion lemma,

$$\widetilde{\Pi} = \left((\mathbf{I} - \mathbf{B})^{-1} - \mathbf{I}\right)\mathbf{Q}^{-1} = \mathbf{B}\left(\mathbf{I} - \mathbf{B}\right)^{-1}\mathbf{Q}^{-1}. \tag{102}$$

Since Π is Hermitian and nonnegative definite, so is $\widetilde{\Pi}$. Thus any $\mathbf{B}$ such that $\widetilde{\Pi}$ of (102) is Hermitian and nonnegative definite can be expressed in the form of (98).

The requirement $\widetilde{\Pi} = \widetilde{\Pi}^*$ is equivalent to

$$\left(\mathbf{I} - \mathbf{B}\right)^{-1}\mathbf{Q}^{-1} = \mathbf{Q}^{-1}\left(\mathbf{I} - \mathbf{B}^*\right)^{-1}, \tag{103}$$

which, taking the inverse of both sides, becomes $\mathbf{Q}\mathbf{B} = \mathbf{B}^*\mathbf{Q}$. The condition $\widetilde{\Pi} \succeq 0$ reduces to

$$\left(\mathbf{I} - \mathbf{B}\right)^{-1}\mathbf{Q}^{-1} \succeq \mathbf{Q}^{-1}. \tag{104}$$

Since $\mathbf{Q}^{-1} \succ 0$, (104) implies that $\left(\mathbf{I} - \mathbf{B}\right)^{-1}\mathbf{Q}^{-1} \succ 0$, which is equivalent to $\mathbf{Q} - \mathbf{Q}\mathbf{B} \succ 0$. Taking the inverse on both sides of (104) we also have that $\mathbf{Q} \succeq \mathbf{Q} - \mathbf{Q}\mathbf{B}$, which together imply that $0 \preceq \mathbf{Q}\mathbf{B} \prec \mathbf{Q}$. Thus, $\mathbf{B}$ and λ satisfy the first KKT condition if $\mathbf{B}^*\mathbf{Q} = \mathbf{Q}\mathbf{B}$, $0 \preceq \mathbf{Q}\mathbf{B} \prec \mathbf{Q}$, and

$$\mathrm{Tr}(\Pi) = \mathrm{Tr}\left(\mathbf{B}\left(\mathbf{I} - \mathbf{B}\right)^{-1}\mathbf{Q}^{-1}\mathbf{T}\right) \leq U^2. \tag{105}$$

We now consider the complementary slackness condition. To this end let

$$\mathbf{Z} = \lambda\mathbf{I} - \mathbf{T}^{-1/2}\left((\mathbf{I} - \mathbf{B})^*(\mathbf{I} - \mathbf{B}) - \mathbf{A}\right)\mathbf{T}^{-1/2}. \tag{106}$$

From feasibility we have that $\mathbf{Z} \succeq 0$, and from complementary slackness $\mathrm{Tr}(\mathbf{Z}\Pi) = 0$. Since $\Pi \succeq 0$, this implies that $\mathbf{Z}\Pi = 0$, or

$$\lambda \mathbf{T}\widetilde{\Pi} = ((\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) - \mathbf{A})\,\widetilde{\Pi}. \tag{107}$$

Substituting (102) into (107), the condition becomes

$$(\lambda \mathbf{T} + \mathbf{A})\,\mathbf{B} = (\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})\mathbf{B}. \tag{108}$$

Finally, since from (97) we have that $\mu = U^2 - \mathrm{Tr}(\Pi)$, the second complementary slackness condition becomes $\lambda(U^2 - \mathrm{Tr}(\Pi)) = 0$, completing the proof.

III. Discussion of the Results of [9]

In this appendix we discuss the minimax MSE solution of [9]. In particular, we show that this solution coincides with the minimax MSE estimator of (52) only in the case in which the matrices $\mathbf{T}$ and $\mathbf{Q}$ are jointly diagonalizable.

In [9, Theorem 4], Pilz considers the minimax MSE estimator that minimizes the worst-case weighted MSE $E\{(\mathbf{G}\mathbf{y}-\mathbf{x})^*\mathbf{U}(\mathbf{G}\mathbf{y}-\mathbf{x})\}$ for some weighting matrix $\mathbf{U}$. The maximum is over all $\mathbf{x}$ satisfying $(\mathbf{x}-\mathbf{x}_0)^*\mathbf{T}(\mathbf{x}-\mathbf{x}_0) \leq 1$ where $\mathbf{x}_0$ is an arbitrary fixed vector. Let $\mathbf{Q} = \mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{H}$ have an eigendecomposition $\mathbf{Q} = \mathbf{S}\mathbf{C}\mathbf{S}^*$ where $\mathbf{S}$ is unitary and $\mathbf{C} = \mathrm{diag}(c_1, \ldots, c_m)$. Then, in the case $\mathbf{U} = \mathbf{I}$ and $\mathbf{x}_0 = 0$, which is the case considered in this paper, the proposed estimator is given by

$$\hat{\mathbf{x}}_{\mathrm{P}} = \mathbf{B}_{\mathrm{P}}\mathbf{Q}^{-1}\mathbf{H}^*\mathbf{C}_w^{-1}\mathbf{y}, \tag{109}$$

where

$$\mathbf{B}_{\mathrm{P}} = \left(\mathbf{I} + \mathbf{Q}^{-1}\mathbf{M}^{-1}\right)^{-1}. \tag{110}$$

Here

$$\mathbf{M} = \mathbf{Q}^{-1}\mathbf{S}\mathbf{D}\mathbf{S}^*\mathbf{Q}^{-1}, \tag{111}$$

and $\mathbf{D}$ is a diagonal matrix with diagonal elements

$$d_i = \frac{1 + \sum_{j=1}^m c_j z_j}{\sqrt{z_i}\sum_{j=1}^m \sqrt{z_j}} - c_i, \quad 1 \leq i \leq m, \tag{112}$$

where z_i is the ith diagonal element of the matrix

$$\mathbf{Z} = \mathbf{S}^*\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1}\mathbf{S}. \tag{113}$$

Substituting (111) into (110),

$$\mathbf{B}_{\mathrm{P}} = \left(\mathbf{I} + \mathbf{S}\mathbf{D}^{-1}\mathbf{C}\mathbf{S}^*\right)^{-1} = \mathbf{I} - \left(\mathbf{I} + \mathbf{S}\mathbf{D}\mathbf{C}^{-1}\mathbf{S}^*\right)^{-1}, \tag{114}$$

where we used the facts that $\mathbf{Q} = \mathbf{S}\mathbf{C}\mathbf{S}^*$ and for any invertible matrix $\mathbf{A}$, $(\mathbf{I}+\mathbf{A})^{-1} = \mathbf{I} - (\mathbf{I}+\mathbf{A}^{-1})^{-1}$. The solution $\hat{\mathbf{x}}_{\text{P}}$ is claimed to be valid as long as

$$\max_{1\le i\le m} c_i\sqrt{z_i} \le \frac{1+\sum_{j=1}^m c_j z_j}{\sum_{j=1}^m \sqrt{z_j}}. \tag{115}$$

We now show that the estimator $\hat{\mathbf{x}}_{\text{P}}$ is similar in structure to the minimax MSE estimator of (52), however, the two coincide only in the case in which $\mathbf{T}$ and $\mathbf{Q}$ are jointly diagonalizable, *i.e.,* only when $\mathbf{S}^*\mathbf{T}\mathbf{S}$ is a diagonal matrix.

Denoting by $\mathbf{P}$ the diagonal matrix with diagonal elements z_i, the matrix $\mathbf{D}$ can be written as

$$\mathbf{D} = \frac{1}{\beta}\mathbf{P}^{-1/2} - \mathbf{C}, \tag{116}$$

where

$$\beta = \frac{\operatorname{Tr}\left(\mathbf{P}^{1/2}\right)}{1+\operatorname{Tr}(\mathbf{C}\mathbf{P})} = \frac{\operatorname{Tr}\left((\mathbf{S}\mathbf{P}\mathbf{S}^*)^{1/2}\right)}{1+\operatorname{Tr}(\mathbf{Q}\mathbf{S}\mathbf{P}\mathbf{S}^*)}. \tag{117}$$

Here we used the fact that $(\mathbf{S}\mathbf{X}\mathbf{S}^*)^{1/2} = \mathbf{S}\mathbf{X}^{1/2}\mathbf{S}^*$ for any $\mathbf{X}\succeq 0$. Substituting $\mathbf{D}$ of (116) into (114),

$$\mathbf{B}_{\text{P}} = \mathbf{I} - \beta\mathbf{S}\mathbf{P}^{1/2}\mathbf{C}\mathbf{S}^* = \mathbf{I} - \beta\mathbf{S}\mathbf{P}^{1/2}\mathbf{S}^*\mathbf{Q}. \tag{118}$$

Comparing (118) with (52) we see that the two expressions are similar. They coincide if and only if

$$\mathbf{S}\mathbf{P}^{1/2}\mathbf{S}^* = \frac{\alpha}{\beta}\left(\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1}\right)^{1/2}, \tag{119}$$

or, equivalently,

$$\mathbf{P} = \frac{\alpha^2}{\beta^2}\mathbf{S}^*\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1}\mathbf{S} = \frac{\alpha^2}{\beta^2}\mathbf{Z}, \tag{120}$$

where $\mathbf{Z}$ and α are defined in (113) and (51), respectively. By definition, $\mathbf{P}$ is the diagonal matrix with diagonal elements z_i. Therefore, (120) is satisfied if and only if $\mathbf{Z}$ is diagonal, and $\alpha=\beta$. Now, we can write $\mathbf{Z}$ as $\mathbf{Z} = \mathbf{C}^{-1}\mathbf{S}^*\mathbf{T}\mathbf{S}\mathbf{C}^{-1}$. Since $\mathbf{C}$ is diagonal, it follows that $\mathbf{Z}$ is diagonal only if $\mathbf{S}^*\mathbf{T}\mathbf{S}$ is diagonal. In this case, $\mathbf{P}=\mathbf{Z}$ and β of (117) reduces to

$$\beta = \frac{\operatorname{Tr}\left((\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1})^{1/2}\right)}{1+\operatorname{Tr}(\mathbf{Q}^{-1}\mathbf{T})} = \alpha. \tag{121}$$

Furthermore, if $\mathbf{Z}$ is diagonal, then the condition (115) can be written as

$$\mathbf{C}\mathbf{Z}^{1/2} \preceq \frac{1+\operatorname{Tr}(\mathbf{C}\mathbf{Z})}{\operatorname{Tr}\left(\mathbf{Z}^{1/2}\right)}\mathbf{I} = \frac{1}{\alpha}\mathbf{I}. \tag{122}$$

Since $\mathbf{C}\mathbf{Z}^{1/2} = \mathbf{S}^*\mathbf{Q}(\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1})^{1/2}\mathbf{S}$, (122) becomes

$$\begin{aligned}&\lambda_{\max}\left(\mathbf{S}^*\mathbf{Q}(\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1})^{1/2}\mathbf{S}\right)\\ &\quad= \lambda_{\max}\left(\mathbf{Q}(\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1})^{1/2}\right) \le \frac{1}{\alpha},\end{aligned} \tag{123}$$

or, equivalently,

$$\alpha \leq \frac{1}{\lambda_{\max}\left(\mathbf{Q}(\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1})^{1/2}\right)} = \lambda_{\min}\left(\mathbf{Q}^{-1}(\mathbf{Q}\mathbf{T}^{-1}\mathbf{Q})^{1/2}\right), \tag{124}$$

which is the same as (50).

We conclude that the solution in [9] coincides with the minimax MSE estimator of (52) only if $\mathbf{Q}$ and $\mathbf{T}$ are jointly diagonalizable. Otherwise, our estimator of (52) leads to smaller worst-case performance, as demonstrated in the example in Fig. 1. Note, that the case of jointly diagonalizable matrices has been treated in [10], in which it was shown that a closed form solution exists for all choices of $\mathbf{Q}$ and $\mathbf{T}$, not only for matrices satisfying (120).

We now discuss the source of the error in the proof of [9, Theroem 4]. The problem considered is that of minimizing $\mathrm{Tr}(\mathbf{S}^*\mathbf{X}\mathbf{S}+\mathbf{C})^{-1}$ subject to $\mathrm{Tr}(\mathbf{Q}^{-1}\mathbf{T}\mathbf{Q}^{-1}\mathbf{X}) = 1$, where $\mathbf{X} \succeq 0$. The author claims that

$$\mathrm{Tr}(\mathbf{S}^*\mathbf{X}\mathbf{S}+\mathbf{C})^{-1} \geq \mathrm{Tr}(\mathbf{L}+\mathbf{C})^{-1} \tag{125}$$

where $\mathbf{L}$ is a diagonal matrix with diagonal elements equal to the diagonal elements of $\mathbf{S}^*\mathbf{X}\mathbf{S}$, with equality if and only if $\mathbf{S}^*\mathbf{X}\mathbf{S}$ is diagonal. The author then concludes that the optimal choice of $\mathbf{X}$ is such that $\mathbf{S}^*\mathbf{X}\mathbf{S}$ is diagonal. This conclusion is indeed correct if the problem is not constrained. However, since the problem is constrained, we cannot draw such a conclusion, since this condition may result in a stricter constraint on the diagonal values resulting in a higher objective.

IV. Proof of Theorem 7

From the KKT conditions applied to the problem of (86) and (87) it follows that $\mathbf{X}$ is optimal if and only if there exists a matrix $\Pi \succeq 0$ such that:

1. $d\mathcal{L}/d\mathbf{X} = 0$ where the Lagrangian $\mathcal{L}$ is defined as

$$\begin{aligned}\mathcal{L} = &\ \mathrm{Tr}\left((\mathbf{X}\mathbf{P}_{\mathcal{R}}-\mathbf{I})\mathbf{Q}^{-1}(\mathbf{X}\mathbf{P}_{\mathcal{R}}-\mathbf{I})^*\right)\\ &+\mathrm{Tr}\left(\Pi\left(\mathbf{A}^{\dagger/2}\mathbf{X}^*\mathbf{X}\mathbf{A}^{\dagger/2}-\mathbf{I}\right)\right).\end{aligned}$$

2. Feasibility: $\mathbf{A}^{\dagger/2}\mathbf{X}^*\mathbf{X}\mathbf{A}^{\dagger/2} \preceq \mathbf{I}$;
3. Complementary slackness: $\mathrm{Tr}(\Pi\mathbf{A}^{\dagger/2}\mathbf{X}^*\mathbf{X}\mathbf{A}^{\dagger/2}) = \mathrm{Tr}(\Pi)$.

Differentiating $\mathcal{L}$ with respect to $\mathbf{X}$ and equating to 0,

$$\mathbf{X}\left(\mathbf{P}_{\mathcal{R}}\mathbf{Q}^{-1}\mathbf{P}_{\mathcal{R}}+\mathbf{A}^{\dagger/2}\Pi\mathbf{A}^{\dagger/2}\right) = \mathbf{Q}^{-1}\mathbf{P}_{\mathcal{R}}. \tag{126}$$

We next exploit the following lemma.

Lemma 4: Let $\mathbf{W}, \mathbf{Z} \succeq 0$. Then $\mathcal{R}(\mathbf{W}+\mathbf{Z}) = \mathcal{R}(\mathbf{W}) \cup \mathcal{R}(\mathbf{Z})$.

Proof: To prove the lemma it suffices to show that

$$\mathcal{N}(\mathbf{W}+\mathbf{Z}) = \mathcal{N}(\mathbf{W}) \cap \mathcal{N}(\mathbf{Z}). \tag{127}$$

Indeed, since $\mathcal{R}(\mathbf{D}) = \mathcal{N}(\mathbf{D})^{\perp}$ for any Hermitian matrix $\mathbf{D}$, if (127) is satisfied then $\mathcal{R}(\mathbf{W}+\mathbf{Z}) = (\mathcal{N}(\mathbf{W}) \cap \mathcal{N}(\mathbf{Z}))^{\perp} = \mathcal{N}(\mathbf{W})^{\perp} \cup \mathcal{N}(\mathbf{Z})^{\perp} = \mathcal{R}(\mathbf{W}) \cup \mathcal{R}(\mathbf{Z})$.

Now, clearly $(\mathbf{W}+\mathbf{Z})\mathbf{x} = 0$ for any $\mathbf{x} \in \mathcal{N}(\mathbf{W}) \cap \mathcal{N}(\mathbf{Z})$. Let $\mathbf{x}$ be such that $(\mathbf{W}+\mathbf{Z})\mathbf{x} = 0$. Then,

$$\mathbf{x}^*\mathbf{W}\mathbf{x} = -\mathbf{x}^*\mathbf{Z}\mathbf{x}. \tag{128}$$

Since $\mathbf{x}^*\mathbf{W}\mathbf{x} \geq 0$ and $-\mathbf{x}^*\mathbf{Z}\mathbf{x} \leq 0$, we conclude that (128) can hold true only if $\mathbf{W}\mathbf{x} = 0$ and $\mathbf{Z}\mathbf{x} = 0$ so that $\mathbf{x} \in \mathcal{N}(\mathbf{W}) \cap \mathcal{N}(\mathbf{Z})$. ■

Since $\mathcal{R}(\mathbf{P}_{\mathcal{R}}\mathbf{Q}^{-1}\mathbf{P}_{\mathcal{R}}) = \mathcal{R}(\mathbf{A})$, and $\mathcal{R}(\mathbf{A}^{\dagger/2}\Pi\mathbf{A}^{\dagger/2}) \subseteq \mathcal{R}(\mathbf{A})$, from Lemma 4 it follows that $\mathcal{R}(\mathbf{P}_{\mathcal{R}}\mathbf{Q}^{-1}\mathbf{P}_{\mathcal{R}} + \mathbf{A}^{\dagger/2}\Pi\mathbf{A}^{\dagger/2}) = \mathcal{R}(\mathbf{A})$. From (126) we then have that

$$\mathbf{Q}\mathbf{X}\mathbf{P}_{\mathcal{R}} = \left(\mathbf{P}_{\mathcal{R}}\mathbf{Q}^{-1}\mathbf{P}_{\mathcal{R}} + \mathbf{A}^{\dagger/2}\Pi\mathbf{A}^{\dagger/2}\right)^{\dagger}, \tag{129}$$

or equivalently,

$$(\mathbf{Q}\mathbf{X}\mathbf{P}_{\mathcal{R}})^{\dagger} = \mathbf{P}_{\mathcal{R}}\mathbf{Q}^{-1}\mathbf{P}_{\mathcal{R}} + \mathbf{A}^{\dagger/2}\Pi\mathbf{A}^{\dagger/2}, \tag{130}$$

and

$$\mathcal{R}(\mathbf{Q}\mathbf{X}\mathbf{P}_{\mathcal{R}}) = \mathcal{R}(\mathbf{A}). \tag{131}$$

Thus $\mathbf{X}$ satisfies the first KKT condition if and only if there exists a $\Pi \succeq 0$ such that (129) or (130) are satisfied.

The requirement $\Pi = \Pi^*$ is equivalent to $\mathbf{Q}\mathbf{X}\mathbf{P}_{\mathcal{R}} = \mathbf{P}_{\mathcal{R}}\mathbf{X}^*\mathbf{Q}$, or in terms of the matrix $\mathbf{B}$,

$$\mathbf{Q}\mathbf{B} = \mathbf{B}^*\mathbf{Q}. \tag{132}$$

The constraint $\Pi \succeq 0$ reduces to

$$(\mathbf{Q}\mathbf{X}\mathbf{P}_{\mathcal{R}})^{\dagger} - \mathbf{P}_{\mathcal{R}}\mathbf{Q}^{-1}\mathbf{P}_{\mathcal{R}} \succeq 0, \tag{133}$$

which using Lemma 2 is equivalent to

$$\mathbf{Q}\mathbf{X}\mathbf{P}_{\mathcal{R}} \succeq 0; \tag{134}$$

$$\mathbf{Q} \succeq \mathbf{P}_{\mathcal{R}}\mathbf{Q}\mathbf{X}\mathbf{P}_{\mathcal{R}}; \tag{135}$$

$$\mathcal{R}(\mathbf{P}_{\mathcal{R}}\mathbf{Q}\mathbf{X}\mathbf{P}_{\mathcal{R}}) = \mathcal{R}(\mathbf{A}). \tag{136}$$

The condition (136) follows from (131). Substituting (85) into (134), the condition becomes $\mathbf{Q}\mathbf{B} \preceq \mathbf{Q}$, and (135) reduces to

$$\mathbf{P}_{\mathcal{R}}\mathbf{Q}(\mathbf{I}-\mathbf{B})\mathbf{P}_{\mathcal{R}} \preceq \mathbf{Q}. \tag{137}$$

Now, from (84) we have that $(\mathbf{I}-\mathbf{B})\mathbf{P}_{\mathcal{R}} = \mathbf{I}-\mathbf{B}$. In addition, from (132),

$$\mathbf{P}_{\mathcal{R}}\mathbf{Q}(\mathbf{I}-\mathbf{B}) = (\mathbf{Q}(\mathbf{I}-\mathbf{B})\mathbf{P}_{\mathcal{R}})^* = (\mathbf{Q}(\mathbf{I}-\mathbf{B}))^* = \mathbf{Q}(\mathbf{I}-\mathbf{B}). \tag{138}$$

Thus, $\mathbf{P}_{\mathcal{R}}\mathbf{Q}(\mathbf{I}-\mathbf{B})\mathbf{P}_{\mathcal{R}} = \mathbf{Q}(\mathbf{I}-\mathbf{B})$ and (137) is equivalent to $\mathbf{QB} \succeq 0$. We conclude that the first KKT condition is satisfied if and only if $\mathbf{QB} = \mathbf{B}^*\mathbf{Q}$, $0 \preceq \mathbf{QB} \preceq \mathbf{Q}$ and $(\mathbf{I}-\mathbf{B})\mathbf{P}_{\mathcal{R}} = \mathbf{I}-\mathbf{B}$.

Using the fact that $\mathbf{X}\mathbf{A}^{\dagger/2} = \mathbf{X}\mathbf{P}_{\mathcal{R}}\mathbf{A}^{\dagger/2}$ and $(\mathbf{I}-\mathbf{B})\mathbf{P}_{\mathcal{R}} = \mathbf{I}-\mathbf{B}$ the second KKT condition is equivalent to $(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) \preceq \mathbf{A}$. The third condition can be written as

$$\mathrm{Tr}(\widetilde{\Pi}(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})) = \mathrm{Tr}(\Pi), \tag{139}$$

where $\widetilde{\Pi} = \mathbf{A}^{\dagger/2}\Pi\mathbf{A}^{\dagger/2}$. Writing Π as $\Pi = \Pi\mathbf{P}_{\mathcal{R}} + \Pi(\mathbf{I}-\mathbf{P}_{\mathcal{R}}) = \Pi\mathbf{A}^{\dagger/2}\mathbf{A}\mathbf{A}^{\dagger/2} + \Pi(\mathbf{I}-\mathbf{P}_{\mathcal{R}})$, (139) becomes

$$\mathrm{Tr}\left(\widetilde{\Pi}\left((\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})-\mathbf{A}\right)\right) = \mathrm{Tr}(\Pi(\mathbf{I}-\mathbf{P}_{\mathcal{R}})). \tag{140}$$

Since $\widetilde{\Pi} \succeq 0$ and $\mathbf{A}-(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B}) \succeq 0$, the left-hand side of (140) is non-positive, while the right-hand side is non-negative since $\Pi \succeq 0$ and $\mathbf{I}-\mathbf{P}_{\mathcal{R}} \succeq 0$. Therefore, (140) is satisfied if and only if

$$(\mathbf{I}-\mathbf{B})^*(\mathbf{I}-\mathbf{B})\widetilde{\Pi} = \mathbf{A}\widetilde{\Pi}. \tag{141}$$

Using the fact that from (130),

$$\widetilde{\Pi} = (\mathbf{Q}(\mathbf{I}-\mathbf{B}))^\dagger - \mathbf{P}_{\mathcal{R}}\mathbf{Q}^{-1}\mathbf{P}_{\mathcal{R}}, \tag{142}$$

completes the proof of the theorem.

References

[1] A.C. Aitken, "On least squares and linear combination of observations," *Proc. Roy. Soc. Edinburgh Sect.*, vol. A 55, pp. 42–48, 1935.

[2] G. Zyskind and F. B. Martin, "On best linear estimation and a general Gauss-Markoff theorem in linear models with arbitrary negative covariance structure," *SIAM J. Appl. Math.*, vol. 17, pp. 1190–1202, 1969.

[3] C. R. Rao, "Unified theory of least squares," *Comm. Statist.*, vol. 1, pp. 1–8, 1973.

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

[5] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems*, Washington, DC: V.H. Winston, 1977.

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

[7] Y. C. Eldar and A. V. Oppenheim, "Covariance shaping least-squares estimation," *IEEE Trans. Signal Processing*, vol. 51, pp. 686–697, Mar. 2003.

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

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

[10] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 53, pp. 168–181, Jan. 2005.

[11] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Linear minimax regret estimation of deterministic parameters with bounded data uncertainties," *IEEE Trans. Signal Processing*, vol. 52, pp. 2177–2188, Aug. 2004.

[12] A. Beck, Y. C. Eldar, and A. Ben-Tal, "Minimax mean-squared error estimation of multichannel signals," submitted to *IEEE Trans. Inform. Theory.*

[13] A. Beck, A. Ben-Tal, and Y. C. Eldar, "Robust mean-squared error estimation of multiple signals in linear systems affected by model and noise uncertainties," *Math Prog.*, to appear.

[14] Z. Ben-Haim and Y. C. Eldar, "Maximum set estimators with bounded estimation error," to appear in *IEEE Trans. Signal Processing.*

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

[16] A. Cohen, "All admissible estimates of the mean vector," *Ann. Math. Statist.*, vol. 37, pp. 458–463, 1966.

[17] C. R. Rao, "Estimation of a parameter in linear models," *Ann. Statist.*, vol. 4, pp. 1023–1037, 1976.

[18] K. Hoffmann, "Admissible improvements of the least squares estimator," *Statistics*, vol. 11, pp. 373–388, 1980.

[19] J. Gross and A. Markiewicz, "Characterizations of admissible linear estimators in the linear model," *Linear Algebra and its Applications*, vol. 388, pp. 239–248, 2004.

[20] K. Hoffmann, "Admissibility of linear estimation with respect to restricted parameter sets," *Math. Oper. Statist. Ser. Statist.*, vol. 8, pp. 425–438, 1977.

[21] T. Mathew, "Admissible linear estimation in singular models with respect to restricted parameter set," *Commun. Statist. Theory and Methods*, vol. 14, no. 2, pp. 491–498, 1985.

[22] K. Hoffmann, "All admissible linear estimators of the regression parameter vector in the case of an arbitrary parameter subset," *Journal of Statistical Planning and Inference*, vol. 48, pp. 371–377, 1995.

[23] C.-Y. Lu and N.-Z. Shi, "Admissible linear estimators in linear models with respect to inequality constraints," *Linear Algebra and its Applications*, vol. 354, pp. 187–194, 2002.

[24] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proc. Third Berkeley Symp. Math. Statist. Prob.* 1956, vol. 1, pp. 197–206, University of California Press, Berkeley.

[25] W. James and C. Stein, "Estimation of quadratic loss," in *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 1961, vol. 1, pp. 361–379, University of California Press, Berkeley.

[26] W. E. Strawderman, "Proper Bayes minimax estimators of multivariate normal mean," *Ann. Math. Statist.*, vol. 42, pp. 385–388, 1971.

[27] K. Alam, "A family of admissible minimax estimators of the mean of a multivariate normal distribution," *Ann. Statist.*, vol. 1, pp. 517–525, 1973.

[28] J. O. Berger, "Admissible minimax estimation of a multivariate normal mean with arbitrary quadratic loss," *Ann. Statist.*, vol. 4, no. 1, pp. 223–226, Jan. 1976.

[29] Z. Ben-Haim and Y. C. Eldar, "Blind minimax estimators: Improving on least squares estimation," to appear in *IEEE Workshop on Statistical signal Processing (SSP'05).*

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

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

[32] D. P. Bertsekas, *Nonlinear Programming*, Belmont MA: Athena Scientific, second edition, 1999.

[33] T. Larsson, M. Patriksson, and A.-B. Strömberg, "On the convergence of conditional ϵ-subgradient methods for convex programs and convex-concave saddle-point problems," *European J. Oper. Res.*, vol. 151, no. 3, pp. 461–473, 2003.

[34] A. Nemirovski, "Prox-method with rate of convergence $o(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems," *SIAM J. Opt.*, vol. 15, pp. 229–251, 2004.

[35] X. Zhan, *Matrix Inequalities*, Berline, Germany: Springer-Verlag, Inc., 2002.

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

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