Robust Estimation of a Random Parameter in a Gaussian Linear Model With Joint Eigenvalue and Elementwise Covariance Uncertainties

Roni Mittelman, *Member, IEEE*, and Eric L. Miller, *Senior Member, IEEE*

***Abstract*—We consider the estimation of a Gaussian random vector x observed through a linear transformation H and corrupted by additive Gaussian noise with a known covariance matrix, where the covariance matrix of x is known to lie in a given region of uncertainty that is described using bounds on the eigenvalues and on the elements of the covariance matrix. Recently, two criteria for minimax estimation called difference regret (DR) and ratio regret (RR) were proposed and their closed form solutions were presented assuming that the eigenvalues of the covariance matrix of x are known to lie in a given region of uncertainty, and assuming that the matrices $H^T C_w^{-1} H$ and C_x are jointly diagonalizable, where C_w and C_x denote the covariance matrices of the additive noise and of x respectively. In this work we present a new criterion for the minimax estimation problem which we call the generalized difference regret (GDR), and derive a new minimax estimator which is based on the GDR criterion where the region of uncertainty is defined not only using upper and lower bounds on the eigenvalues of the parameter's covariance matrix, but also using upper and lower bounds on the individual elements of the covariance matrix itself. Furthermore, the new estimator does not require the assumption of joint diagonalizability and it can be obtained efficiently using semidefinite programming. We also show that when the joint diagonalizability assumption holds and when there are only eigenvalue uncertainties, then the new estimator is identical to the difference regret estimator. The experimental results show that we can obtain improved mean squared error (MSE) results compared to the MMSE, DR, and RR estimators.**

***Index Terms*—Covariance uncertainty, linear estimation, minimax estimators, minimum mean squared error (MMSE) estimation, regret, robust estimation.**

I. Introduction

THE classic solution to estimating a Gaussian random vector $\boldsymbol{x}$ that is observed through a linear transformation and corrupted by Gaussian noise is obtained using the minimum mean squared error (MMSE) estimator which assumes full knowledge of the covariance matrix of the random vector $\boldsymbol{x}$ and the covariance matrix of the observation noise. Specifically, let

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

where $\boldsymbol{y} \in \mathbb{R}^n$ is the observation, $\boldsymbol{H} \in \mathbb{R}^{n\times m}$, and $\boldsymbol{x} \in \mathbb{R}^m$, $\boldsymbol{w} \in \mathbb{R}^n$ are independent zero mean Gaussian random vectors with covariance matrices $\boldsymbol{C_x}$ and $\boldsymbol{C_w}$, respectively, then given an observation vector $\boldsymbol{y}$ the MMSE estimate of $\boldsymbol{x}$ takes the form [1]

$$\hat{\boldsymbol{x}} = \boldsymbol{C_x}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{C_x}\boldsymbol{H}^T + \boldsymbol{C_w})^{-1}\boldsymbol{y}. \tag{2}$$

In many applications it is reasonable to expect that the estimate of the covariance matrix of the observation noise is accurate. However the estimate of the covariance matrix of $\boldsymbol{x}$ may often be highly inaccurate and lead to severe performance degradation when using the MMSE estimator. Therefore, in practice it is necessary to require the estimator to be robust with respect to such uncertainties. The common approach to achieve such robustness is through the use of a minimax estimator which minimizes the worst case performance over some criterion in the region of uncertainty [3], [4].

One such performance measure is the mean squared error (MSE), where the estimator is chosen such that the worst case MSE in the region of uncertainty of the covariance matrix of $\boldsymbol{x}$ is minimized. However, as was noted in [1] this choice may be too pessimistic and therefore the performance of an estimator designed this way may be unsatisfactory. Instead it is proposed in [1] to minimize the worst case *difference regret* (DR) which is defined as the difference between the MSE when using a linear estimator of the form $\hat{\boldsymbol{x}} = \boldsymbol{G}\boldsymbol{y}$ and the MSE when using the MMSE estimator matched to a covariance matrix $\boldsymbol{C_x}$, where $\boldsymbol{G}$ is a matrix with the appropriate dimensions. The motivation for this choice is that the worst case DR criterion is less pessimistic than the worst case MSE criterion. Similarly, the *ratio regret* (RR) estimator proposed in [2], minimized the worst case RR which is defined as the ratio between the MSE when using a linear estimator of the form $\hat{\boldsymbol{x}} = \boldsymbol{G}\boldsymbol{y}$ and the MSE when using the MMSE estimator matched to a covariance matrix $\boldsymbol{C_x}$. The motivation for the RR estimator is similar to the DR where the MSE is measured in decibels. The DR and RR estimators presented in [1] and [2] assume that the eigenvector matrix of $\boldsymbol{C_x}$ is known and is identical to the eigenvector matrix of $\boldsymbol{H}^T\boldsymbol{C_w}^{-1}\boldsymbol{H}$, which is also called the *jointly diagonalizable matrices* assumption. Furthermore, the region of uncertainty is expressed using upper and lower bounds on each of the eigenvalues of $\boldsymbol{C_x}$.

Manuscript received March 04, 2009; accepted September 29, 2009. First published November 06, 2009; current version published February 10, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Thierry Blu. This work was supported by the Center for Subsurface Sensing and Imaging Systems under the Engineering Research Centers Program of the National Science Foundation (Award Number EEC-9986821).

R. Mittelman is with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: rmittelm@umich.edu).

E. L. Miller is with the Department of Electrical and Computer Engineering, Tufts University, Medford, MA 02155 USA (e-mail: elmiller@ece.tufts.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2036063

In this paper, we develop a new criterion for the robust estimation problem which we call the generalized difference regret (GDR). Rather than subtracting the MSE when using the MMSE estimator matched to a covariance matrix $\boldsymbol{C_x}$ from the MSE when using an estimator $\boldsymbol{G}$, for the GDR we subtract another function of $\boldsymbol{C_x}$ and $\boldsymbol{C_w}$. More specifically, we develop a collection of qualifications that this function should satisfy, which are aimed at guaranteeing the scale invariance of the obtained estimator and ensuring that the GDR criterion is not more pessimistic than the MSE criterion. Functions satisfying these criteria are termed *admissible regret functions*. While the choice of an admissible regret function is far from unique, in this paper, we make one suggestion which we call the linearized epigraph (LE) admissible regret function, and use it as the basis for the development of a new robust estimator.

The estimator we propose here generalizes the ideas in both [1] and [2] in a number of ways and can, thus, be used to address a far broader range of estimation problems. Most importantly, our approach does not require the joint diagonalizability assumption and allows for uncertainty in both the eigenvalues as well as the individual elements of $\boldsymbol{C_x}$. Our LE-GDR scheme can also be computed easily using semidefinite programming. When considering only eigenvalue uncertainties and using the jointly diagonalizable matrices assumption, we show that the resulting estimator is identical to the DR estimator. This result gives insight into why the new criterion is an effective tool for designing robust estimators, and helps to explain the experimental results.

We test the LE-GDR estimator using two examples. First we consider the same example used in [1] and [2], when the covariance matrix is obtained from a stationary process and where the MSE is computed using the same samples that are used to find the robust estimator, and also use it for cases in which the jointly diagonalizable matrices assumption does not hold. Subsequently, we consider using the LE-GDR estimator in an estimation problem in a sensor network, where unlike the previous example different samples are used to compute the MSE and to find the estimator. A major concern in sensor networks applications is the power loss due to the communication of messages between the sensor nodes rather than the energy lost during computation [5], [6]. We show that the LE-GDR estimator can be used to reduce the number of samples which have to be transmitted to a centralized location in order to estimate a covariance matrix which is required in order to use the MMSE estimator. The experimental results of the new estimator show improved MSE compared to presently available methods.

The remainder of this paper is organized as follows. In Section II, we give the background on the DR and RR estimators, on semidefinite programming, and on minimax theory. In Section III, we present the GDR criterion for minimax estimation and the LE admissible regret function which is then used with the GDR criterion to derive the LE-GDR estimator with joint eigenvalue and elementwise covariance matrix uncertainties. Section IV presents an example of the LE-GDR estimator using a stationary covariance matrix and different choices for the matrix $\boldsymbol{H}$, and Section V presents the application of the LE-GDR estimator to a robust estimation problem in a sensor network. Section VI concludes this paper.

II. Background

Throughout this paper we denote vectors in $\mathbb{R}^n$ by boldface lower-case letters, and matrices in $\mathbb{R}^{n\times m}$ by boldface upper-case letters. The notation $\boldsymbol{A} \preceq \boldsymbol{B}$ means that $\boldsymbol{B}-\boldsymbol{A}$ is a positive semidefinite matrix, and $\boldsymbol{A} \prec \boldsymbol{B}$ means that $\boldsymbol{B}-\boldsymbol{A}$ is a positive definite matrix. The notation $\boldsymbol{A} \leq \boldsymbol{B}$ means that $A_{ij} \leq B_{ij}$ for all $i = 1,\ldots,n$ and $j = 1,\ldots,m$, and $\boldsymbol{I}$ denotes the identity matrix with appropriate dimensions, and $(\cdot)^T$ denotes the transpose of a matrix. The pseudo inverse of a matrix is denoted by $(\cdot)^\dagger$, and $\hat{(\cdot)}$ denotes an estimator. The trace of the matrix $\boldsymbol{A}$ is denoted by $Tr(\boldsymbol{A})$, and $\mathrm{diag}(\boldsymbol{x})$ denotes a diagonal matrix with the diagonal elements of the vector $\boldsymbol{x}$. A multivariate Gaussian distribution with mean $\boldsymbol{m}$ and covariance matrix $\boldsymbol{\Sigma}$ is denoted by $\mathcal{N}(\boldsymbol{m},\boldsymbol{\Sigma})$.

A. Minimax Regret Estimators

The aim of the minimax regret estimators is to achieve robustness to the uncertainty in the covariance matrix by finding a linear estimator of the form $\hat{\boldsymbol{x}} = \boldsymbol{G}\boldsymbol{y}$ that minimizes the worst performance of the regret in the region of uncertainty of the covariance matrix $\boldsymbol{C_x}$. Specifically, let $\Re(\boldsymbol{C_x},\boldsymbol{G})$ denote the regret, and let $\Omega \subset S_+$, where S_+ denotes the set of positive semidefinite matrices, denote the region of uncertainty of $\boldsymbol{C_x}$. The minimax estimator is then obtained by solving

$$\boldsymbol{G} = \arg\min_{\boldsymbol{G}} \max_{\boldsymbol{C_x}\in\Omega} \Re(\boldsymbol{C_x},\boldsymbol{G}). \tag{3}$$

The DR and RR criteria are defined as the difference and the ratio between the MSE when using an estimator of the form $\hat{\boldsymbol{x}} = \boldsymbol{G}\boldsymbol{y}$ and the MSE when using the MMSE estimator. The MSE when estimating $\boldsymbol{x}$ using a linear estimator of the form $\hat{\boldsymbol{x}} = \boldsymbol{G}\boldsymbol{y}$ is given by [1]

$$\mathbb{E}\left(\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|^2\right) = Tr\left(\boldsymbol{C_x}(\boldsymbol{I}-\boldsymbol{GH})^T(\boldsymbol{I}-\boldsymbol{GH})\right) + Tr(\boldsymbol{G}\boldsymbol{C_w}\boldsymbol{G}^T). \tag{4}$$

The MSE when using the MMSE estimator takes the form [1]

$$MSE^0 = Tr\left(\left(\boldsymbol{H}^T\boldsymbol{C_w^{-1}}\boldsymbol{H}+\boldsymbol{C_x^{-1}}\right)^{-1}\right). \tag{5}$$

Both the difference and ratio estimators presented in [1] and [2] assume that the region of uncertainty Ω is expressed as uncertainties in the eigenvalues of the covariance matrix $\boldsymbol{C_x}$ assuming that the eigenvectors are known. Specifically, let $\boldsymbol{V}$ denote the eigenvectors matrix of $\boldsymbol{C_x}$, and let u_i and ℓ_i denote upper and lower bounds on the eigenvalues δ_i $i = 1,\ldots,m$, then $\Omega = \{\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T|\boldsymbol{\Delta} = \mathrm{diag}(\boldsymbol{\delta}), \ell_i \leq \delta_i \leq u_i\}$.

1) Difference Regret Estimator: The DR is defined as the difference between (4) and (5)

$$\begin{aligned}\Re(\boldsymbol{C_x},\boldsymbol{G}) &= \mathbb{E}\left(\|\hat{\boldsymbol{x}}-\boldsymbol{x}\|^2\right) - MSE^0\\ &= Tr\left(\boldsymbol{C_x}(\boldsymbol{I}-\boldsymbol{GH})^T(\boldsymbol{I}-\boldsymbol{GH})\right) + Tr(\boldsymbol{G}\boldsymbol{C_w}\boldsymbol{G}^T)\\ &\quad - Tr\left(\left(\boldsymbol{H}^T\boldsymbol{C_w^{-1}}\boldsymbol{H}+\boldsymbol{C_x^{-1}}\right)^{-1}\right).\end{aligned} \tag{6}$$

Assuming that $\boldsymbol{H}^T\boldsymbol{C_w^{-1}}\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$ where $\boldsymbol{\Lambda}$ is a diagonal matrix with the diagonal elements λ_i $i = 1, \ldots, m$, it is shown in [1] that

$$\boldsymbol{G} = \boldsymbol{V}\boldsymbol{D}\Lambda^{-1}\boldsymbol{V}^T\boldsymbol{H}^T\boldsymbol{C_w^{-1}} \tag{7}$$

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

$$d_i = 1 - \frac{1}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}} \tag{8}$$

and where $\zeta_i = (u_i + \ell_i)/2$ and $\epsilon_i = (u_i - \ell_i)/2$.

The DR estimator can also be interpreted as the MMSE estimator (2) with an equivalent covariance matrix $\boldsymbol{C_x} = \boldsymbol{V}\boldsymbol{X}\boldsymbol{V}^T$ where $\boldsymbol{X}$ is a diagonal matrix with the diagonal elements

$$x_i = \frac{1}{\lambda_i}\left(\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2} - 1\right). \tag{9}$$

2) Ratio Regret Estimator: The RR is defined as the ratio between (4) and (5). Assuming that $\boldsymbol{H}^T\boldsymbol{C_w^{-1}}\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$ where Λ is a diagonal matrix with the positive diagonal elements λ_i $i = 1, \ldots, m$, it is shown in [2] that the RR estimator also takes the form in (7), where $\boldsymbol{D}$ is an $m \times m$ diagonal matrix with diagonal elements d_i that are given by

$$d_i = \begin{cases} 1 - \frac{\sqrt{\gamma}}{\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}}, & 1 \le \gamma \le \gamma_i^0 \\ \frac{\lambda_i(\zeta_i - \epsilon_i)}{1+\lambda_i(\zeta_i-\epsilon_i)}, & \gamma \ge \gamma_i^0 \end{cases} \tag{10}$$

where

$$\gamma_i^0 = \frac{1+\lambda_i(\zeta_i+\epsilon_i)}{1+\lambda_i(\zeta_i-\epsilon_i)} \tag{11}$$

and where γ is chosen using a line search such that $\sum_{i=1}^n t_i(\gamma) = 0$, where $t_i(\gamma)$ is given in (12)

$$t_i(\gamma) = \begin{cases} \frac{1}{\lambda_i} - \frac{2\sqrt{\gamma}}{\lambda_i\sqrt{(1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2}} \\ \quad + \frac{\gamma\left(\lambda_i^2\left(\epsilon_i^2 - \zeta_i^2\right)+1\right)}{\lambda_i\left((1+\lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2\right)} & 1 \le \gamma \le \gamma_i^0 \\ \frac{(\gamma-1)(\epsilon_i - \zeta_i)}{1+\lambda_i(\zeta_i-\epsilon_i)} & \gamma \ge \gamma_i^0. \end{cases} \tag{12}$$

B. Semidefinite Programming

Convex optimization problems deal with minimization of a convex objective function over a convex domain. Unlike general nonlinear problems, convex optimization problems can be solved efficiently using interior point methods in polynomial complexity [7]. One subclass of the convex optimization problems that is used in this paper is semidefinite programming which takes the form [8], [9]

$$\min_{\boldsymbol{x}} \boldsymbol{c^T x} \tag{13}$$

subject to

$$\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{F_0} + x_1\boldsymbol{F_1} + \ldots + x_n\boldsymbol{F_n} \succeq 0 \tag{14}$$

where $\boldsymbol{F_0}, \boldsymbol{F_1}, \ldots, \boldsymbol{F_n}$ are symmetric matrices, $x_1, \ldots, x_n$ denote the elements of $\boldsymbol{x} \in \mathbb{R}^n$, $\boldsymbol{c} \in \mathbb{R}^n$, and the generalized inequality is with respect to the positive semidefinite cone. The standard form of a semidefinite program can easily be extended to include linear equality constraints [8].

The following Lemma is often used in order to transform an optimization problem into the semidefinite programming form.

Lemma 1: (Schur's Complement [10]): Let

$$\boldsymbol{M} = \begin{bmatrix} \boldsymbol{A} & \boldsymbol{B}^T \\ \boldsymbol{B} & \boldsymbol{C} \end{bmatrix}$$

be a Hermitian matrix with $\boldsymbol{C} \succ 0$ (i.e., $\boldsymbol{C}$ is a positive definite matrix). Then $\boldsymbol{M} \succeq 0$ if and only if $\boldsymbol{A} - \boldsymbol{B}^T\boldsymbol{C^{-1}}\boldsymbol{B} \succeq 0$.

C. Minimax Theory

Minimax theory deals with optimization problems of the form

$$\min_{v\in D}\max_{u\in C} K(u,v) \tag{15}$$

where C and D denote two nonempty sets and $K : C \times D \longrightarrow [-\infty, \infty]$. The solution of such optimization problems is not straightforward in the general case, however, if the objective function K satisfies certain conditions, then there exist minimax theorems that can facilitate the solution. In particular, if the objective function has a saddle point then it must be a solution of the minimax problem (although it may not be a unique solution).

Definition 1: [11] Let C and D denote two nonempty sets and let $K : C \times D \longrightarrow [-\infty, \infty]$, then a point $(\tilde{u}, \tilde{v}) \in C \times D$ is called a saddle point of K with respect to maximizing over C and minimizing over D if

$$K(u,\tilde{v}) \le K(\tilde{u},\tilde{v}) \le K(\tilde{u},v), \quad \forall u \in C, \quad \forall v \in D.$$

An important Lemma that states sufficient conditions for a function to have a saddle point is given here.

Lemma 2: [11] Let C and D be two non-empty closed convex sets in $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, and let K be a continuous finite concave-convex function on $C \times D$ (i.e., $|K(u,v)| < \infty$, concave in C and convex in D). If either C or D is bounded, one has

$$\inf_{v\in D}\sup_{u\in C} K(u,v) = \sup_{u\in C}\inf_{v\in D} K(u,v). \tag{16}$$

It can also be shown that if the conditions in Lemma 2 are satisfied then the solution to (16) is a saddle point [11]. Most importantly since the order of the maximization and minimization can be interchanged, the solution of the minimax problem can be simplified in many cases.

III. Minimax Estimation With Joint Eigenvalue and Elementwise Covariance Uncertainties Based on the GDR Criterion

In this section, we propose a new criterion for the minimax problem which we call the generalized difference regret (GDR) criterion, and subsequently we use this criterion to develop a new robust estimator which has two major differences compared to the DR and RR estimators. It does not necessitate the jointly diagonalizable matrices assumption, and the region of uncertainty can be defined as the intersection of the eigenvalue and elementwise uncertainty regions.

As was demonstrated in [1], the MSE is a very conservative criterion for the minimax estimation problem and performs poorly, therefore the DR criterion was motivated as being less pessimistic than the MSE criterion. We define the GDR as the difference between the MSE when using an estimator $\boldsymbol{G}$ and a function $f(\boldsymbol{C_x}, \boldsymbol{C_w}, \ldots)$ which is a function of $\boldsymbol{C_x}, \boldsymbol{C_w}$ and potentially some other parameters

$$\Re_g(\boldsymbol{C_x}, \boldsymbol{G}) = Tr\left(\boldsymbol{C_x}(\boldsymbol{I} - \boldsymbol{GH})^T(\boldsymbol{I} - \boldsymbol{GH})\right) + Tr(\boldsymbol{G}\boldsymbol{C_w}\boldsymbol{G}^T) - f(\boldsymbol{C_x}, \boldsymbol{C_w}, \ldots). \quad (17)$$

It can be seen that if we take $f(\boldsymbol{C_x}, \boldsymbol{C_w}, \ldots)$ equal to the MSE when using the MMSE estimator matched to a covariance matrix $\boldsymbol{C_x}$ (5), then we obtain the DR as a special case of the GDR criterion. More generally, we consider functions $f(\boldsymbol{C_x}, \boldsymbol{C_w}, \ldots)$ that satisfy the qualifications given in the following.

Definition 2: A function $f(\boldsymbol{C_x}, \boldsymbol{C_w}, \ldots)$ is called an *admissible regret function* if it satisfies the following:

1) $f(\boldsymbol{C_x}, \boldsymbol{C_w}, \ldots) \geq 0$
2) $f(\alpha\boldsymbol{C_x}, \alpha\boldsymbol{C_w}, \ldots) = \alpha f(\boldsymbol{C_x}, \boldsymbol{C_w}, \ldots), \forall \alpha > 0$.

The first qualification ensures that the GDR in (17) is not greater than the MSE when using an estimator $\boldsymbol{G}$ as in (4), and is therefore not more pessimistic than the MSE criterion. Using the second qualification, we have that the GDR criterion satisfies

$$\Re_g(\alpha\boldsymbol{C_x}, \alpha\boldsymbol{C_w}) = \alpha\Re_g(\boldsymbol{C_x}, \boldsymbol{C_w}), \quad \forall \quad \alpha > 0 \quad (18)$$

and, therefore, the second qualification ensures that the obtained estimator is invariant to the scaling of $\boldsymbol{C_x}$ and $\boldsymbol{C_w}$.

In order to derive an admissible regret function we also argue that it is advisable to choose a convex function as it would lead to a GDR criterion which is convex-concave and, therefore, using the results of Lemma 2 the solution of the minimax problem becomes much simplified. In order to obtain our admissible regret function, we make some modifications to (5) such that it is in the form of a Schur's complement and is linear in $\boldsymbol{\delta}$. First we note that (5) can be rewritten as

$$\begin{aligned} Tr&\left(\left(\boldsymbol{H}^T\boldsymbol{C_w^{-1}}\boldsymbol{H} + \boldsymbol{C_x^{-1}}\right)^{-1}\right) \\ &= Tr\left((\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}^T\boldsymbol{V} + \boldsymbol{\Delta}^{-1})^{-1}\right) \\ &= Tr\left(\boldsymbol{\Delta}(\boldsymbol{Q}\boldsymbol{Q}^T + \boldsymbol{I})^{-1}\right) \end{aligned} \quad (19)$$

where $\boldsymbol{Q} = \boldsymbol{\Delta}^{1/2}\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}^{1/2}$. Since a function $f(\boldsymbol{x})$ is convex if and only if its epigraph $\{(\boldsymbol{x}, t)|f(\boldsymbol{x}) \leq t\}$ is a convex set, we note that using Lemma 1 the epigraph of (5) takes the form

$$\left\{(\boldsymbol{\delta}, t)\middle|\begin{bmatrix}\boldsymbol{X} & \sqrt{\boldsymbol{\Delta}} \\ \sqrt{\boldsymbol{\Delta}} & \boldsymbol{Q}\boldsymbol{Q}^T + \boldsymbol{I}\end{bmatrix} \succeq 0, Tr(\boldsymbol{X}) = t\right\} \quad (20)$$

where $\sqrt{\boldsymbol{\Delta}}$ is a diagonal matrix with the diagonal elements $\sqrt{\delta_i}$, $i = 1, \ldots, m$. The set given in (20) is not convex because the matrix inequality is not linear in $\boldsymbol{\delta}$. Our approach is to linearize the matrix inequality as follows.

1) We replace each of the diagonal elements of $\sqrt{\boldsymbol{\Delta}}$ with the line that connects the points $(\ell_i, \sqrt{\ell_i})$ and $(u_i, \sqrt{u_i})$.
2) We assume that $\boldsymbol{Q} = \boldsymbol{Q}^T$ which is a relaxed version of the jointly diagonalizable matrices assumption since it always holds if $\boldsymbol{V} = \boldsymbol{T}$, however, it may also hold in other cases, for example, if $\boldsymbol{\Delta} = \boldsymbol{\Lambda}$ and $\boldsymbol{V}^T\boldsymbol{T} = \boldsymbol{T}^T\boldsymbol{V}$.

The epigraph for the new function therefore takes the form

$$\left\{(\boldsymbol{\delta}, t)\middle|\begin{bmatrix}\boldsymbol{X} & \boldsymbol{P} \\ \boldsymbol{P} & \boldsymbol{Q}^T\boldsymbol{Q} + \boldsymbol{I}\end{bmatrix} \succeq 0, Tr(\boldsymbol{X}) = t\right\} \quad (21)$$

where $\boldsymbol{P}$ is a diagonal matrix with the diagonal elements $\sqrt{\ell_i} + \kappa_i(\delta_i - \ell_i)$, and where $\kappa_i = (\sqrt{u_i} + \sqrt{\ell_i})^{-1}$.

The function whose epigraph is (21) is shown in Lemma 3 to be an admissible regret function. We call this function the linearized epigraph (LE) admissible regret function.

Lemma 3: Let $\boldsymbol{H}^T\boldsymbol{C_w^{-1}}\boldsymbol{H} = \boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}^T$ where $\boldsymbol{\Lambda}$ is a diagonal matrix with the nonnegative elements $\{\lambda_i\}_{i=1}^m$ and where $\boldsymbol{T}$ is a unitary matrix. Let

$$f(\boldsymbol{\delta}, \boldsymbol{C_w}, \boldsymbol{\ell}, \boldsymbol{u}, \boldsymbol{H}) = Tr\left(\boldsymbol{P}^2(\boldsymbol{\Lambda}^{1/2}\boldsymbol{T}^T\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}^{1/2} + \boldsymbol{I})^{-1}\right) \quad (22)$$

where $\boldsymbol{P}$ is a diagonal matrix with the diagonal elements $\sqrt{\ell_i} + \kappa_i(\delta_i - \ell_i)$, and where $\kappa_i = (\sqrt{u_i} + \sqrt{\ell_i})^{-1}$. We then have that $f(\boldsymbol{\delta}, \boldsymbol{C_w}, \boldsymbol{\ell}, \boldsymbol{u}, \boldsymbol{H})$ is an admissible regret function and convex in $\boldsymbol{\delta}$.

Proof: The nonnegativity of $f(\boldsymbol{\delta}, \boldsymbol{C_w}, \boldsymbol{\ell}, \boldsymbol{u}, \boldsymbol{H})$ follows since $\boldsymbol{P}(\boldsymbol{\Lambda}^{1/2}\boldsymbol{T}^T\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}^{1/2} + \boldsymbol{I})^{-1}\boldsymbol{P}$ is a positive semidefinite matrix. To prove the second qualification of Definition 2 we note that $\boldsymbol{\Lambda}^{1/2}\boldsymbol{T}^T\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}^{1/2}$ is invariant to the scaling of $\boldsymbol{\delta}$ and $\boldsymbol{C_w}$, and that the scaling of $\boldsymbol{\ell}$ and $\boldsymbol{u}$ is the same as that of $\boldsymbol{\delta}$. Therefore, we have

$$\begin{aligned} &f(\alpha\boldsymbol{\delta}, \alpha\boldsymbol{C_w}, \alpha\boldsymbol{\ell}, \alpha\boldsymbol{u}, \boldsymbol{H}) \quad (23) \\ &\quad = Tr\left((\boldsymbol{P}_\alpha)^2(\boldsymbol{\Lambda}^{1/2}\boldsymbol{T}^T\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}^{1/2} + \boldsymbol{I})^{-1}\right) \\ &\quad = \alpha f(\boldsymbol{\delta}, \boldsymbol{C_w}, \boldsymbol{\ell}, \boldsymbol{u}, \boldsymbol{H}) \quad (24) \end{aligned}$$

where the $\boldsymbol{P}_\alpha$ is a diagonal matrix with the diagonal elements $\sqrt{\alpha\ell_i} + (\sqrt{\alpha\ell_i} + \sqrt{\alpha u_i})^{-1}(\alpha\delta_i - \alpha\ell_i)$. The convexity of $f(\boldsymbol{\delta}, \boldsymbol{C_w}, \boldsymbol{\ell}, \boldsymbol{u}, \boldsymbol{H})$ in $\boldsymbol{\delta}$ follows since the epigraph is a convex set. ■

Next, we derive in theorem 1 the new minimax estimator that uses the GDR criterion with the LE admissible regret function.

Theorem 1: Let $\boldsymbol{x}$ denote the unknown parameter vector in the linear Gaussian model $\boldsymbol{y} = \boldsymbol{Hx} + \boldsymbol{w}$ where $\boldsymbol{H} \in \mathbb{R}^{n\times m}$ and where $\boldsymbol{x} \in \mathbb{R}^m$ and $\boldsymbol{w} \in \mathbb{R}^n$ are independent zero mean Gaussian random vectors with covariance matrices $\boldsymbol{C_x}$ and $\boldsymbol{C_w}$, respectively. Let $\boldsymbol{U}$ and $\boldsymbol{L}$ denote elementwise upper and lower bounds on the elements of $\boldsymbol{C_x}$ such that $\boldsymbol{L} \leq \boldsymbol{C_x} \leq \boldsymbol{U}$, and let $\boldsymbol{V}$ denote a unitary matrix such that $\boldsymbol{C_x} = \boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T$ where $\boldsymbol{\Delta}$ is a diagonal matrix with the diagonal elements δ_i such that $0 \leq \ell_i \leq \delta_i \leq u_i, i = 1, \ldots, m$. Furthermore, let $\boldsymbol{H}^T\boldsymbol{C_w^{-1}}\boldsymbol{H} = \boldsymbol{T}\boldsymbol{\Lambda}\boldsymbol{T}^T$ where $\boldsymbol{\Lambda}$ is a diagonal matrix with the diagonal elements $\lambda_i \geq 0, i = 1, \ldots, m$ and where $\boldsymbol{T}$ is a unitary matrix. Then the solution to the problem

$$\min_{\hat{\boldsymbol{x}}=\boldsymbol{Gy}} \max_{\boldsymbol{\delta}\in\Omega} \Re_g(\boldsymbol{\Delta}, \boldsymbol{G}) \quad (25)$$

where

$$\begin{aligned} \Re_g(\boldsymbol{\Delta}, \boldsymbol{G}) =& Tr\left(\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T(\boldsymbol{I} - \boldsymbol{GH})^T(\boldsymbol{I} - \boldsymbol{GH})\right) \\ &+ Tr(\boldsymbol{G}\boldsymbol{C_w}\boldsymbol{G}^T) \\ &- Tr\left(\boldsymbol{P}^2(\boldsymbol{\Lambda}^{1/2}\boldsymbol{T}^T\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}^{1/2} + \boldsymbol{I})^{-1}\right) \end{aligned} \quad (26)$$

and where $\Omega = \{\boldsymbol{\delta}_0 | \boldsymbol{L} \leq \boldsymbol{V}\mathrm{diag}(\boldsymbol{\delta}_0)\boldsymbol{V}^T \leq \boldsymbol{U}, \ell_i \leq \delta_{0,i} \leq u_i\}$, takes the form

$$\boldsymbol{G} = \boldsymbol{V}\boldsymbol{\Delta}^*\boldsymbol{V}^T\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{V}\boldsymbol{\Delta}^*\boldsymbol{V}^T\boldsymbol{H}^T + \boldsymbol{C}_{\boldsymbol{w}})^{-1} \tag{27}$$

where the diagonal elements $\boldsymbol{\delta}^*$ of $\boldsymbol{\Delta}^*$ can be obtained as follows:

1) $\boldsymbol{\delta}^*$ can be obtained as the optimal solution for $\boldsymbol{\delta}$ of the semidefinite program

$$\min_{\boldsymbol{Z}_1,\boldsymbol{Z}_2,\boldsymbol{\delta}} Tr(\boldsymbol{Z}_1 + \boldsymbol{Z}_2 - \Delta) \tag{28}$$

$$\begin{aligned} \text{subject to} & \\ & \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{H}^T \\ \boldsymbol{H}\boldsymbol{V}\boldsymbol{\Delta} & \boldsymbol{C}_{\boldsymbol{w}} + \boldsymbol{H}\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{H}^T \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} \boldsymbol{Z}_2 & \boldsymbol{P} \\ \boldsymbol{P} & \boldsymbol{\Lambda}^{1/2}\boldsymbol{T}^T\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}^{1/2} + \boldsymbol{I} \end{bmatrix} \succeq 0 \\ & \ell_i \leq \delta_i \leq u_i \\ & \boldsymbol{L} \leq \boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T \leq \boldsymbol{U} \end{aligned} \tag{29}$$

where $\boldsymbol{P}$ is defined as in Lemma 3.

2) If $\boldsymbol{V} = \boldsymbol{T}$, then $\boldsymbol{\delta}^*$ can be obtained as the optimal solution for $\boldsymbol{\delta}$ of the semidefinite program

$$\min_{\boldsymbol{z}_1,\boldsymbol{z}_2,\boldsymbol{\delta}} \sum_{i=1}^{m} \left(z_1^{(i)}/\lambda_i + z_2^{(i)} \right) \tag{30}$$

$$\begin{aligned} \text{subject to} & \\ & \begin{bmatrix} z_1^{(i)} & 1 \\ 1 & 1 + \lambda_i\delta_i \end{bmatrix} \succeq 0 \\ & \begin{bmatrix} z_2^{(i)} & \sqrt{\ell_i} + \kappa_i(\delta_i - \ell_i) \\ \sqrt{\ell_i} + \kappa_i(\delta_i - \ell_i) & 1 + \lambda_i\delta_i \end{bmatrix} \succeq 0 \\ & \ell_i \leq \delta_i \leq u_i \\ & \boldsymbol{L} \leq \boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T \leq \boldsymbol{U} \end{aligned} \tag{31}$$

where $\kappa_i = (\sqrt{u_i} + \sqrt{\ell_i})^{-1}$.

Proof: In order to show that the estimator takes the form in (27) we note that $\Re_g(\boldsymbol{\Delta}, \boldsymbol{G})$ in (26) and the minimax problem (25) satisfy all the conditions of Lemma 2 and therefore the order of minimization and maximization can be interchanged. Minimizing (26) with respect to $\boldsymbol{G}$ leads to a solution in the form of the MMSE estimator with a covariance matrix given by $\boldsymbol{C}_{\boldsymbol{x}} = \boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T$, and specifically

$$\boldsymbol{G} = \boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{H}^T + \boldsymbol{C}_{\boldsymbol{w}})^{-1}. \tag{32}$$

Substituting (32) into (26) then leads to the objective for the maximization, which is simply the difference between the MSE when using the MMSE estimator (5) with $\boldsymbol{C}_{\boldsymbol{x}} = \boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T$ and the LE admissible regret function in (22),

$$\max_{\boldsymbol{\delta}\in\Omega} \left\{ Tr\left(\left(\boldsymbol{H}^T\boldsymbol{C}_{\boldsymbol{w}}^{-1}\boldsymbol{H} + \boldsymbol{V}\boldsymbol{\Delta}^{-1}\boldsymbol{V}^T \right)^{-1} \right) - Tr\left(\boldsymbol{P}^2(\boldsymbol{\Lambda}^{1/2}\boldsymbol{T}^T\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}^{1/2} + \boldsymbol{I})^{-1} \right) \right\}. \tag{33}$$

Additionally, we have $Tr((\boldsymbol{H}^T\boldsymbol{C}_{\boldsymbol{w}}^{-1}\boldsymbol{H} + \boldsymbol{V}\boldsymbol{\Delta}^{-1}\boldsymbol{V}^T)^{-1}) = Tr(((\boldsymbol{H}\boldsymbol{V})^T\boldsymbol{C}_{\boldsymbol{w}}^{-1}\boldsymbol{H}\boldsymbol{V} + \boldsymbol{\Delta}^{-1})^{-1})$ and using the matrix inversion Lemma [8] we have

$$\left((\boldsymbol{H}\boldsymbol{V})^T\boldsymbol{C}_{\boldsymbol{w}}^{-1}\boldsymbol{H}\boldsymbol{V} + \boldsymbol{\Delta}^{-1} \right)^{-1} = \boldsymbol{\Delta} - \boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{H}^T(\boldsymbol{C}_{\boldsymbol{w}} + \boldsymbol{H}\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{H}^T)^{-1}\boldsymbol{H}\boldsymbol{V}\boldsymbol{\Delta}. \tag{34}$$

We can now rewrite (33) as

$$\min_{\boldsymbol{Z}_1,\boldsymbol{Z}_2,\boldsymbol{\delta}} Tr(\boldsymbol{Z}_1 + \boldsymbol{Z}_2 - \boldsymbol{\Delta}) \tag{35}$$

$$\begin{aligned} \text{subject to} & \\ & \boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{H}^T(\boldsymbol{C}_{\boldsymbol{w}} + \boldsymbol{H}\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{H}^T)^{-1}\boldsymbol{H}\boldsymbol{V}\boldsymbol{\Delta} \preceq \boldsymbol{Z}_1 \\ & \boldsymbol{P}(\boldsymbol{\Lambda}^{1/2}\boldsymbol{T}^T\boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T\boldsymbol{T}\boldsymbol{\Lambda}^{1/2} + \boldsymbol{I})^{-1}\boldsymbol{P} \preceq \boldsymbol{Z}_2 \\ & \ell_i \leq \delta_i \leq u_i \\ & \boldsymbol{L} \leq \boldsymbol{V}\boldsymbol{\Delta}\boldsymbol{V}^T \leq \boldsymbol{U} \end{aligned} \tag{36}$$

and using Lemma 1 we obtain the semidefinite program in (28) and (29), which proves 1.

In order to prove 2 we use $\boldsymbol{V} = \boldsymbol{T}$ in (33) which simplifies to

$$\max_{\boldsymbol{\delta}\in\Omega} \sum_{i=1}^{m} \left(\frac{1}{\lambda_i} - \frac{1/\lambda_i}{1+\lambda_i\delta_i} - \frac{\left(\sqrt{\ell_i} + \kappa_i(\delta_i - \ell_i)\right)^2}{1+\lambda_i\delta_i} \right). \tag{37}$$

By adding the inequalities

$$\frac{1}{1+\lambda_i\delta_i} \leq z_1^{(i)}$$
$$\frac{\left(\sqrt{\ell_i} + \kappa_i(\delta_i - \ell_i)\right)^2}{1+\lambda_i\delta_i} \leq z_2^{(i)}$$

and using Lemma 1, it follows that the δ_i's are obtained using the semidefinite program given by (30) and (31). ■

The computational complexity of the semidefinite program in for the general case is $O(m^4)$ whereas the computational complexity of the semidefinite program when the jointly diagonalizable matrices assumption holds is $O(m^3)$ [9]. Therefore, if joint diagonalizability holds it can be used to reduce the computational complexity. Furthermore, the semidefinite program can be solved efficiently and accurately using standard toolboxes, e.g., [12].

It is important to emphasize that since the solution of the minimax problem is obtained without the joint diagonalizability assumption, the LE-GDR estimator can be used generally also when joint diagonalizability does not hold. This is also verified by the experimental results that are given in the next Section.

A. Equivalence Between the LE-GDR Estimator With Eigenvalue Alone Uncertainties and the Difference Regret Estimator for the Jointly Diagonalizable Matrices Case

Although a closed form solution of the DR estimator assuming that $\boldsymbol{H}^T\boldsymbol{C}_{\boldsymbol{w}}^{-1}\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$ and with eigenvalue alone uncertainty region was presented in [1], it is interesting to derive the closed form solution to the LE-GDR estimator under the same assumptions since it reveals an interesting property of the LE-GDR estimator. In order to derive the closed form solution we maximize the objective in (37) with respect to $\boldsymbol{\delta}$ over the uncertainty set $\Omega = \{\boldsymbol{\delta}_0 | \ell_i \leq \delta_{0,i} \leq u_i\}$. If the maximum of the objective is obtained inside the uncertainty interval, then it is also the solution to the constrained problem. Solving for the maximum of the unconstrained problem, we have that the solution must satisfy the quadratic equation

$$\delta_i^2 + \frac{2}{\lambda_i}\delta_i + \frac{(\sqrt{\ell_i} - \kappa_i\ell_i)(2\kappa_i - \lambda_i\sqrt{\ell_i} + \lambda_i\ell_i\kappa_i) - 1}{\kappa_i^2\lambda_i} = 0 \tag{38}$$

and its solution takes the form

$$\begin{aligned}\delta_i =& -\frac{1}{\lambda_i} + \frac{1}{\lambda_i}\\ &\times\sqrt{1 - \frac{\lambda_i}{\kappa_i^2}(\sqrt{\ell_i} - \kappa_i\ell_i)(2\kappa_i - \lambda_i\sqrt{\ell_i} + \lambda_i\ell_i\kappa_i) + \frac{\lambda_i}{\kappa_i^2}}\\ =& \frac{1}{\lambda_i}\left(\sqrt{1 + \lambda_i u_i + \lambda_i\ell_i + \lambda_i^2\ell_i u_i} - 1\right).\end{aligned} \tag{39}$$

It is straightforward to verify that (39) satisfies $\ell_i \le \delta_i \le u_i$ and therefore it is also the solution to the constrained problem. Furthermore, if we define $\zeta_i = (u_i + \ell_i)/2$ and $\epsilon_i = (u_i - \ell_i)/2$ then we obtain that

$$\delta_i = \frac{1}{\lambda_i}\left(\sqrt{(1 + \lambda_i\zeta_i)^2 - \lambda_i^2\epsilon_i^2} - 1\right) \tag{40}$$

which is identical to the solution that is obtained for the DR estimator (9).

This result indicates that if the elementwise bounds are very loose (as may be the case in high SNR scenarios), and if the jointly diagonalizable matrices assumption holds then the performance is going to be identical to that of the DR estimator. It also gives us insight into why the LE-GDR criterion performs well experimentally, since it leads to the same solution as the DR criterion under the same assumptions in this case.

IV. Example of the LE-GDR Estimator

The example that we consider here is an estimation problem with the model given in (1), where $\boldsymbol{x}$ is a length m segment of a zero mean stationary first order autoregressive process with parameter ρ and where the covariance matrix of $\boldsymbol{w}$ is $\boldsymbol{C_w} = \sigma^2\boldsymbol{I}$ where σ is assumed to be known. The autocorrelation function of $\boldsymbol{x}$ therefore takes the form

$$\mathbb{E}(x_i x_j) = \rho^{|j-i|}. \tag{41}$$

The covariance matrix of $\boldsymbol{x}$, which is denoted by $\boldsymbol{C_x}$, is unknown and is estimated from the available noisy measurements vector $\boldsymbol{y}$ using the estimator

$$\hat{\boldsymbol{C}}_{\boldsymbol{x}} = \left[\boldsymbol{H}^\dagger(\hat{\boldsymbol{C}}_{\boldsymbol{y}} - \boldsymbol{C_w})\boldsymbol{H}^{\dagger T}\right]_+ = \left[\boldsymbol{H}^\dagger(\hat{\boldsymbol{C}}_{\boldsymbol{y}} - \sigma^2\boldsymbol{I})\boldsymbol{H}^{\dagger T}\right]_+ \tag{42}$$

where $[\boldsymbol{C}]_+$ is obtained by replacing all the negative eigenvalues of $\boldsymbol{C}$ with zero. Specifically, let $\boldsymbol{C} = \boldsymbol{UQU^{-1}}$ where $\boldsymbol{Q}$ is a diagonal matrix, then $[\boldsymbol{C}]_+ = \boldsymbol{U\bar{Q}U^{-1}}$ where $\bar{\boldsymbol{Q}}$ is a diagonal matrix with the elements $\bar{q}_{ii} = \max(0, q_{ii})$. Let $\{\boldsymbol{y}^{(t)}\}_{t=1}^N$ denote the N sample vectors available to estimate the covariance matrix, then the estimate of the covariance matrix of $\boldsymbol{y}$ takes the form [1]

$$\hat{\boldsymbol{C}}_{\boldsymbol{y}}(i,j) = \frac{1}{mN}\sum_{t=1}^{N}\sum_{k=1}^{m-|j-i|} y_k^{(t)} y_{k+|j-i|}^{(t)}. \tag{43}$$

Since the estimators considered in this paper assume that the eigenvector matrix $\boldsymbol{V}$ of the parameter's covariance matrix is known, we set it equal to the eigenvector matrix of $\hat{\boldsymbol{C}}_{\boldsymbol{x}}$ (more on the estimation of the eigenvectors of covariance matrices can be found in [13]). Let ζ_i denote the eigenvalues of $\hat{\boldsymbol{C}}_{\boldsymbol{x}}$ then similarly to [1], [2] we set the upper and lower bounds for the eigenvalues of the covariance matrix as $u_i = \zeta_i + \epsilon_i$, $\ell_i = \zeta_i - \epsilon_i$, where ϵ_i is proportional to the standard deviation of an estimate $\hat{\sigma}_x^2$ of the variance σ_x^2.

If $\boldsymbol{H} = \boldsymbol{I}$ then we have

$$\hat{\sigma}_x^2 = \frac{1}{m}\sum_{i=1}^{m} y_i^2 - \sigma^2 \tag{44}$$

and the variance of $\hat{\sigma}_x^2$ takes the form

$$\begin{aligned}\mathbb{E}\left\{\left(\hat{\sigma}_x^2 - \sigma_x^2\right)^2\right\} &= \mathbb{E}\left\{\left(\frac{1}{m}\sum_{i=1}^{m}\left(y_i^2 - \sigma_w^2 - \sigma_x^2\right)\right)^2\right\}\\ &= \frac{1}{m^2}\sum_{i,j=1}^{m}\mathbb{E}(t_i t_j)\end{aligned} \tag{45}$$

where $t_i = y_i^2 - \sigma^2 - \sigma_x^2$. Since $\boldsymbol{x}$ and $\boldsymbol{w}$ are Gaussian and independent we have

$$\mathbb{E}\left\{\left(\hat{\sigma}_x^2 - \sigma_x^2\right)^2\right\} = \frac{2}{m^2}\sum_{i,j=1}^{m}\left(\boldsymbol{C_x}(i,j) + \sigma^2\delta_{ij}\right)^2. \tag{46}$$

The expression given in (46) for the variance of the estimate is slightly different from that given in [1] since we did not assume that the covariance matrix is circular which leads to the simplified expression given in [1] as this is only true in the limit when $m \to \infty$ [14].

If $\boldsymbol{H} \neq \boldsymbol{I}$ then we have the following estimator for the variance of the signal

$$\hat{\sigma}_x^2 = \frac{1}{m}\left(Tr(\boldsymbol{H}^\dagger\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{H}^{\dagger T}) - Tr(\boldsymbol{H}^\dagger\boldsymbol{C_w}\boldsymbol{H}^{\dagger T})\right) \tag{47}$$

and the variance of the estimator $\hat{\sigma}_x^2$ is (see Appendix)

$$\begin{aligned}&\mathbb{E}\left\{\left(\hat{\sigma}_x^2 - \sigma_x^2\right)^2\right\}\\ &= \frac{1}{m^2}\sum_{i,j=1}^{m}\mathcal{H}_{i,j}Tr\\ &\quad\times\left(\left(\boldsymbol{\Sigma}\left(\boldsymbol{E}_{i,j} + \boldsymbol{E}_{i,j}^T\right)\boldsymbol{\Sigma} + Tr(\boldsymbol{E}_{i,j}\boldsymbol{\Sigma})\boldsymbol{\Sigma}\right)\boldsymbol{\mathcal{H}}\right)\\ &\quad - \frac{1}{m^2}\left(Tr(\boldsymbol{C_x} + \boldsymbol{H}^\dagger\boldsymbol{C_w}\boldsymbol{H}^{\dagger T})\right)^2\end{aligned} \tag{48}$$

where $\boldsymbol{\mathcal{H}} = \boldsymbol{H}^{\dagger T}\boldsymbol{H}^\dagger$ and where $\boldsymbol{E}_{i,j}$ is an $m \times m$ matrix with all zero entries but for the i,j entry which is 1.

In order to ensure the nonnegativity of the eigenvalues, ϵ_i takes the form

$$\epsilon_i = \min\left(\zeta_i, A \cdot \sqrt{\mathbb{E}\left\{(\hat{\sigma}_x^2 - \sigma_x^2)^2\right\}}\right) \qquad (49)$$

where the estimate $\hat{\boldsymbol{C}}_{\boldsymbol{x}}(i,j)$ is used instead of $\boldsymbol{C}_{\boldsymbol{x}}(i,j)$ in (46) or (48) in order to compute the variance of $\hat{\sigma}_x^2$, and where A is a proportionality constant chosen experimentally. The elementwise bounds are chosen to be proportional to $\hat{\sigma}_x^2$, and inversely proportional to the standard deviation of $\hat{\sigma}_x^2$. Choosing the elements of the covariance matrix to be proportional to the variance is very intuitive since if the variance is large then the elements of the covariance matrix are expected to be larger in their absolute value, and alternatively if the variance is small then the elements of the covariance matrix are expected to be smaller in their absolute value. The motivation for choosing the elementwise uncertainty bounds to be inversely proportional to the standard deviation of $\hat{\sigma}_x^2$ is less intuitive though. We argue that if the standard deviation of $\hat{\sigma}_x^2$ is small then the estimate of the covariance matrix that we have is expected to be fairly good, and, therefore, we would like our estimator to be close to the MMSE estimator which is optimal if the covariance matrix is perfectly known. Therefore, we would like the elementwise bounds to be very loose so that we only employ the eigenvalue uncertainties which lead to an estimator that converges to the MMSE estimator as the upper and lower bounds on the eigenvalues become closer (since the eigenvalue uncertainty region was chosen to be proportional to the standard deviation of $\hat{\sigma}_x^2$ this is indeed the case). On the other hand if the standard deviation of $\hat{\sigma}_x^2$ is large then we cannot obtain a good estimate of the covariance matrix of the random parameter and therefore the elementwise bounds should be very small in their absolute value such that the estimator is close to $G = 0$. We therefore set the elementwise bounds to

$$\boldsymbol{U}(i,j) = -\boldsymbol{L}(i,j) = B\frac{\hat{\boldsymbol{C}}_{\boldsymbol{x}}(1,1)}{\sqrt{\mathbb{E}\left\{(\hat{\sigma}_x^2 - \sigma_x^2)^2\right\}}} \qquad (50)$$

where B is a proportionality constant, and the estimate $\hat{\boldsymbol{C}}_{\boldsymbol{x}}(i,j)$ is used in (46) or (48) instead of $\boldsymbol{C}_{\boldsymbol{x}}(i,j)$ in order to compute the variance of $\hat{\sigma}_x^2$.

In all the experiments that we present in this section, we used $N = 2$ sample vectors in order to estimate the covariance matrix using (43), and used only one of them in order to plot the MSE or maximum squared error versus SNR figures. Since we assume that $\boldsymbol{x}$ is zero mean and the autocorrelation function is given in (41) the SNR is computed using $-10\log_{10}(\sigma^2)$. Fig. 1 shows the MSE versus SNR for $\boldsymbol{H} = \boldsymbol{I}$, where the MSE is averaged over all the components of the vector. This model satisfies the constraint $\boldsymbol{H}^T\boldsymbol{C}_{\boldsymbol{w}}^{-1}\boldsymbol{H} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^T$, which is required by the DR and RR estimators, for any orthonormal matrix $\boldsymbol{V}$. Furthermore we can use the more computationally efficient implementation given in Theorem 1 for this case. The parameters that we used were $m = 10$, $\sigma = 1$, $A = 4$, $B = 1$, $\rho = 0.8$, and the MSE was averaged over 2000 independent experiments for each SNR value. It can be seen that the LE-GDR estimator can improve the MSE compared to all the other estimators. Since the jointly diagonalizable matrices assumption holds for this example it follows from Section III-A that the results obtained using the LE-GDR estimator with eigenvalue alone uncertainties are the same as those obtained using the DR estimator. This explains the convergence of the LE-GDR estimator with the joint elementwise and eigenvalue uncertainties to the DR estimator in high SNRs, since the elementwise uncertainty was chosen to be very large for high SNRs. It can also be seen that the LE-GDR estimator converges to the RR estimator in low SNRs, which can be explained as an effect of the elementwise bounds. Since the elements of the covariance matrix are bounded, then it can be seen from (27) that as the variance of the noise increases the estimator converges to $\boldsymbol{G} = 0$.

Fig. 2 shows the maximum squared error versus the SNR for the same parameters that were used for Fig. 1, where the maximum squared error was computed over all the elements of $\boldsymbol{x}$,

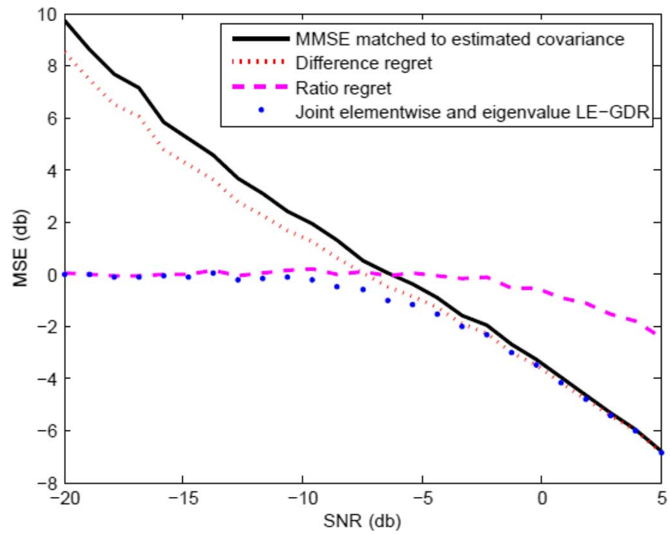

Fig. 1. MSE versus the SNR for the LE-GDR estimator, DR and RR estimators, and the MMSE estimator matched to the estimated covariance, for $\boldsymbol{H} = \boldsymbol{I}$.

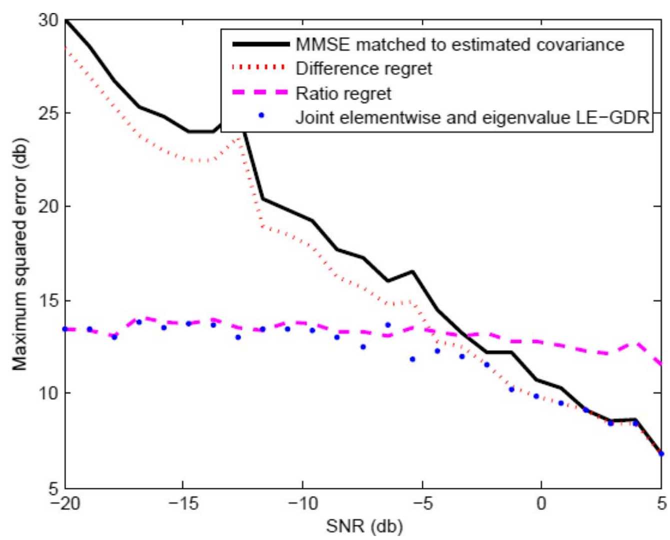

Fig. 2. Maximum squared error versus SNR for the LE-GDR estimator, DR and RR estimators, and the MMSE estimator matched to the estimated covariance, for $\boldsymbol{H} = \boldsymbol{I}$.

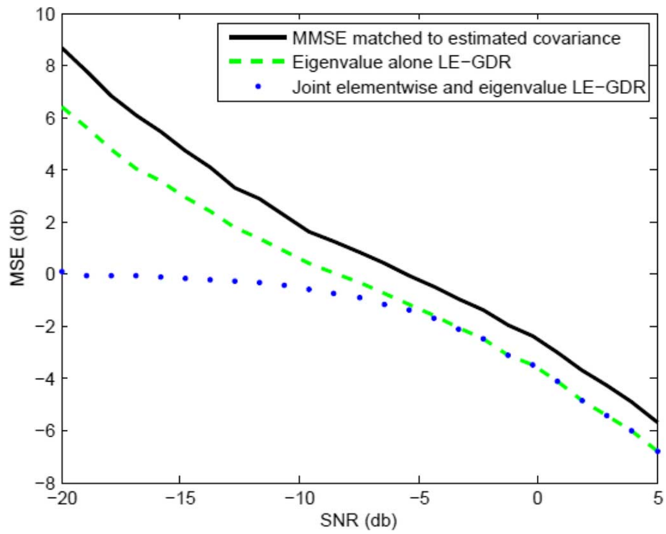

Fig. 3. MSE versus SNR for the LE-GDR estimator and for the MMSE estimator matched to the estimated covariance, with $\boldsymbol{H}$ in a Toeplitz form.

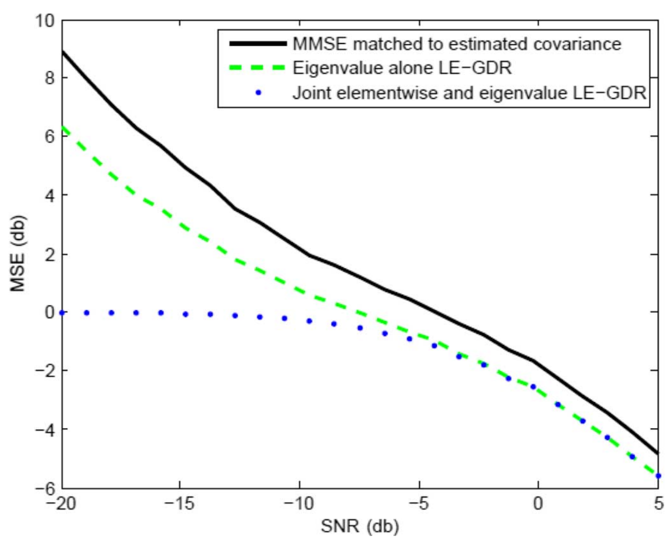

Fig. 4. MSE versus SNR for the LE-GDR estimator and for the MMSE estimator matched to the estimated covariance, with $\boldsymbol{H}$ in a diagonal form.

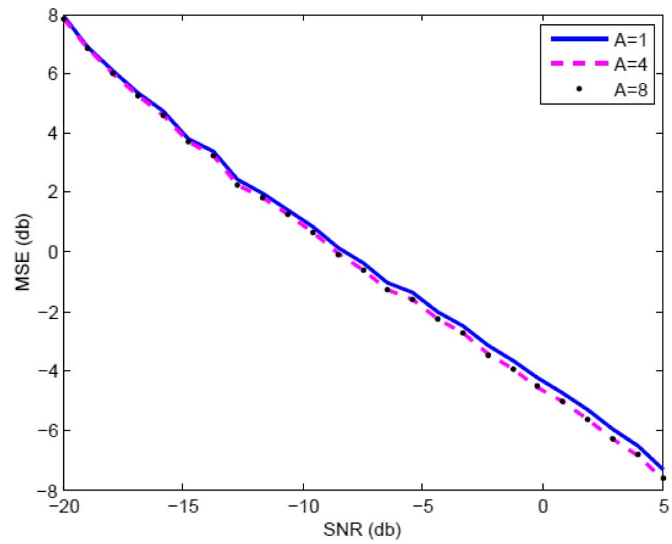

Fig. 5. MSE versus SNR for the LE-GDR estimator with eigenvalue alone uncertainties for different values of A, with $\boldsymbol{H}$ in a Toeplitz form.

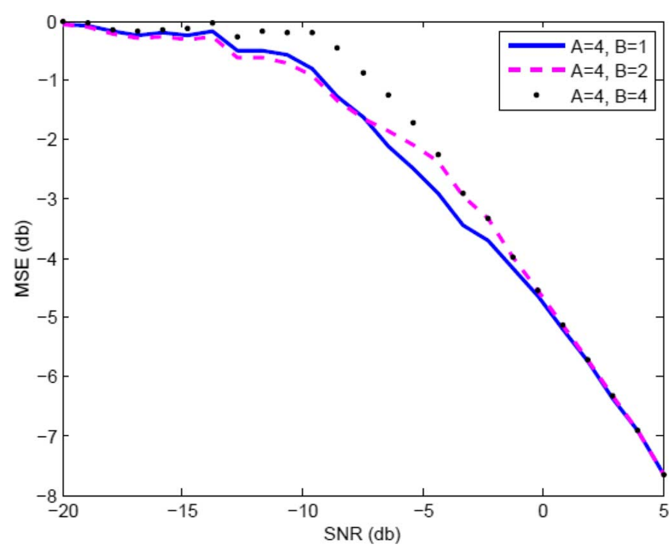

Fig. 6. MSE versus SNR for the LE-GDR estimator with joint elementwise and eigenvalue uncertainties for $A = 4$ and different values for B, with $\boldsymbol{H}$ in a Toeplitz form.

and over 40 000 repetitions of the estimation process. It can be seen that the MMSE estimator that is matched to the estimated covariance has the worse MSE performance among all the estimators since it does not address the uncertainty in the estimated covariance matrix. The MSE of the LE-GDR estimator is generally lower than all the other estimators which confirms the robustness of the new estimator with respect to uncertainties in the covariance matrix.

Figs. 3 and 4 show the MSE versus SNR when $\boldsymbol{H}$ is a Toeplitz matrix and a diagonal matrix, respectively, such that the jointly diagonalizable matrices assumption does not hold. Specifically in Fig. 3, we use a Toeplitz matrix which implements a linear time invariant filter with 4 taps given by $h[0] = 1$, $h[1] = 0.4$, $h[2] = 0.2$, $h[3] = 0.1$, and in Fig. 4 we use the diagonal matrix $\boldsymbol{H} = \mathrm{diag}([1, 0.8, 1, 0.5, 1.3, 1.2, 1.5, 0.7, 2, 1.5]^T)$ where the diagonal elements were chosen arbitrarily. In both figures, we used the parameters $m = 10$, $\sigma = 1$, $A = 4$, $B = 2$, $\rho = 0.8$, and the LE-GDR eigenvalue alone estimator was obtained by removing the elementwise uncertainty constraint from (29). It can be seen from both of the figures that the MSE can be improved significantly when using the LE-GDR estimator compared to using the MMSE estimator.

Finally, in Figs. 5 and 6 we study the effect that the parameters A and B have on the performance of the LE-GDR estimator when using the same experimental setting that was used for Fig. 3. Fig. 5 shows the MSE versus SNR for the LE-GDR estimator with eigenvalue uncertainties alone for different values of the parameter A. It can be seen that the performance is not too sensitive to the exact choice of this parameter. Fig. 6 shows the MSE versus SNR for the LE-GDR estimator with joint elementwise and eigenvalue uncertainties when A is fixed and the parameter B changes. It can be seen that there is greater sensitivity to the exact choice of this parameter.

V. Robust Estimation in a Sensor Network

A sensor network is comprised of many autonomous sensors that are spread in an environment, collecting data and communicating with each other [15]. Each sensor node also has some computational resources and can process the data that it acquires and the transmission that it receives from other sensors independently. Since the sensors are usually battery powered, a major concern in such applications is reducing the energy consumption, especially the energy spent on communication between the sensors, which is significantly larger than any other cause for energy consumption. The straightforward approach to estimation in sensor networks is to transmit all the data collected by the sensors to a centralized location and perform the estimation there, however this approach is very inefficient energy wise since an enormous amount of data has to be transmitted. Instead the more energy efficient approach is to transmit messages between the sensor nodes and have the sensors perform the estimation collectively. Such decentralized estimation can be performed using the distributed algorithms presented in [16] and [17]. Nevertheless these distributed estimation algorithms depend on an estimate of the covariance or inverse covariance matrix, and therefore in practice require an initial stage where many samples are transmitted to a centralized location so that the covariance matrix or inverse covariance matrix can be estimated. The results presented in this paper can be used to improve the estimation performance for a given number of samples that are transmitted to the centralized location and used in order to obtain the estimator. Furthermore, since in the LE-GDR estimator has the same form as the MMSE estimator then one can use the same methods presented in [16], [17] to perform distributed estimation.

The estimation model for the sensor network case is

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{w} \tag{51}$$

where we assume that each node's signal is a scalar (extension to the vector case is straightforward) and the Gaussian random vector $\boldsymbol{x}$ is composed of all the sensors' signals. Similarly, the vector $\boldsymbol{y}$ is composed of all the sensors' noisy observations. The Gaussian random noise vector $\boldsymbol{w}$ where the covariance matrix of $\boldsymbol{w}$ is $\boldsymbol{C_w} = \sigma^2\boldsymbol{I}$. This model is identical to (1) with $\boldsymbol{H} = \boldsymbol{I}$, and therefore satisfies the constraint $\boldsymbol{H}^T\boldsymbol{C_w^{-1}}\boldsymbol{H} = \boldsymbol{V\Lambda V}^T$ which is required by the DR and RR estimators for any orthonormal matrix $\boldsymbol{V}$. Unlike the previous examples, in this example we use a different set of samples for finding the estimator and for testing its performance and therefore the elementwise bounds used in the previous example do not apply in this case. However, since in a sensor network the variance at each sensor can be estimated without transmitting any data (assuming that the observation noise is i.i.d.), we can assume that it is known and use the bound for the elements of the covariance matrix $\boldsymbol{C_x}$ [18]

$$|\boldsymbol{C_x}(i,j)| \le \sigma_{x,i}\sigma_{x,j} \quad \forall\, i,j \in 1,\ldots,m \tag{52}$$

where $\sigma_{x,i}$ denotes the true standard deviation of sensor i, in order to obtain the required elementwise bounds.

In order to simulate the sensors' signals we assume that the covariance matrix is obtained from a Gaussian process (GP) [19], [20] as such modeling is common in sensor networks e.g.,

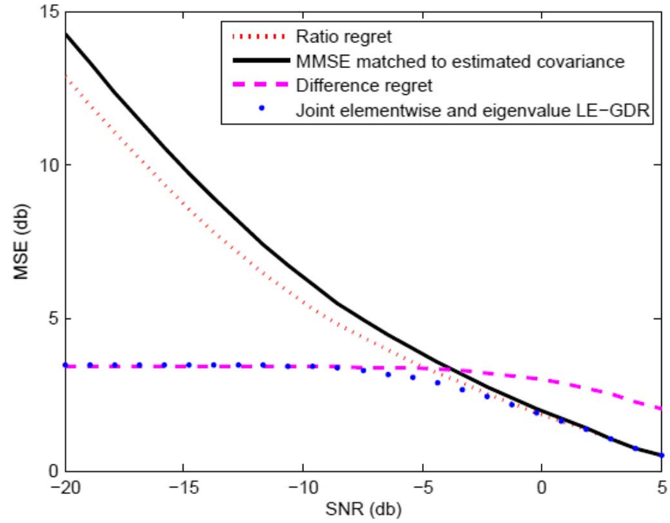

Fig. 7. MSE versus SNR for different estimators for the sensor network example.

[21]. We use a zero mean GP with a neural network covariance function [19] that takes the form

$$k(\boldsymbol{s},\boldsymbol{s}') = \frac{2}{\pi}\sin^{-1}\left(\frac{2\tilde{\boldsymbol{s}}^T\boldsymbol{\Sigma}\tilde{\boldsymbol{s}}'}{\sqrt{(1+2\tilde{\boldsymbol{s}}^T\boldsymbol{\Sigma}\tilde{\boldsymbol{s}})(1+2\tilde{\boldsymbol{s}}'^T\boldsymbol{\Sigma}\tilde{\boldsymbol{s}}')}}\right) \tag{53}$$

where $\tilde{\boldsymbol{s}} = [1\ \boldsymbol{s}^T]^T$, and we used $\boldsymbol{\Sigma} = \mathrm{diag}([10,10,10]^T)$. We generate the positions of $m = 20$ sensors $\{\boldsymbol{s}_j, j = 1,\ldots,m\}$ by sampling a uniform distribution over [−2, 2] for both of the axes. The covariance matrix of the signal vector $\boldsymbol{x}$ is then obtained by $\boldsymbol{C_x}(i,j) = k(\boldsymbol{s}_i,\boldsymbol{s}_j)$, and the measurement vectors $\boldsymbol{y}^{(t)}, t = 1,\ldots,N$ available at the centralized location are generated using (51). The covariance matrix is then estimated from the available samples using

$$\hat{\boldsymbol{C}}_{\boldsymbol{x}} = \left[\frac{1}{N}\sum_{t=1}^{N}\boldsymbol{y}^{(t)}\boldsymbol{y}^{(t)T} - \sigma^2\boldsymbol{I}\right]_+ \tag{54}$$

where σ^2 denotes the variance of the noise which is assumed known, and $[\boldsymbol{C}]_+$ is obtained by replacing the negative eigenvalues of $\boldsymbol{C}$ with zero. Let ζ_i denote the eigenvalues of $\hat{\boldsymbol{C}}_{\boldsymbol{x}}$ then we set the bounds on the eigenvalues to be $\ell_i = 0$, and $u_i = 2\zeta_i$. The bounds on the elements of the covariance matrix are set using (52) to $\boldsymbol{U}(i,j) = \sqrt{\boldsymbol{C_x}(i,i)\boldsymbol{C_x}(j,j)}$ and $\boldsymbol{L}(i,j) = -\sqrt{\boldsymbol{C_x}(i,i)\boldsymbol{C_x}(j,j)}$, where $\boldsymbol{C_x}(i,i)$ denotes the true variance of the signal at sensor node i which as mentioned previously is assumed to be known.

In order to show the usefulness of the LE-GDR estimator for the sensor network problem we assume that we have only $N = 5$ measurement vectors at the centralized location using which we can obtain the robust estimator for $\boldsymbol{x}$. We averaged the MSE shown in Fig. 7 over 2000 experiments, where in each experiment we first generated $N = 5$ measurements from the linear Gaussian model which were used to obtain the robust estimator, and subsequently we computed the MSE using 2000 measurements which were different from those that were used to find the robust estimator. The SNR is computed as $10\log_{10}(\sum_{i=1}^{m}\boldsymbol{C_x}(i,i)/(m\sigma^2))$. It can be seen that the LE-GDR estimator either improves or performs equally as well as the other estimators. Furthermore, since the jointly diagonalizable matrices assumption holds for this example, for high SNRs when the elementwise bounds are very loose we have that the performance of the LE-GDR estimator with joint elementwise and eigenvalue uncertainties converges to that of the DR estimator, as is shown in Section III-A. Similarly to the example in the previous section, it can be seen that the LE-GDR estimator converges to the RR estimator for low SNRs, which is the effect of the elementwise bounds on the covariance matrix.

VI. Conclusion

We presented a new minimax estimator that is robust to an uncertainty region that is described using bounds on the eigenvalues and bounds on the elements of the covariance matrix. The estimator is based on a new criterion which is called the linearized epigraph generalized difference regret (LE-GDR) and can be obtained efficiently using semidefinite programming. Furthermore, the LE-GDR estimator avoids the jointly diagonalizable matrices assumption that is required by both the DR and RR estimators and can therefore be used in more general cases. We also showed that when the jointly diagonalizable matrices assumption holds and when there are only eigenvalue uncertainties, then the LE-GDR estimator is identical to the DR estimator. This result gives motivation into why the proposed criterion is successful, and explains the convergence of the LE-GDR estimator with joint elementwise and eigenvalue uncertainties to the DR estimator in high SNRs when the jointly diagonalizable matrices assumption holds. The experimental results show that the LE-GDR estimator can improve the MSE over the MMSE estimator and the DR and RR estimators. When considering model matrices that do not satisfy the jointly diagonalizable matrices assumption we also showed significant MSE improvement compared to the MMSE estimator.

Appendix
The Variance of the Estimator $\hat{\sigma}^2$ for $\boldsymbol{H} \neq \boldsymbol{I}$

Using (47) the variance of the estimator is

$$\begin{aligned}\mathbb{E}&\left\{\left(\hat{\sigma}_x^2-\sigma_x^2\right)^2\right\}\\ &=\frac{1}{m^2}\mathbb{E}\left\{\left(Tr(\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{H}^{\dagger T}\boldsymbol{H}^{\dagger})-Tr(\boldsymbol{C_x}+\boldsymbol{H}^{\dagger}\boldsymbol{C_w}\boldsymbol{H}^{\dagger T})\right)^2\right\}\\ &=\frac{1}{m^2}\mathbb{E}\left\{\left(Tr(\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{H}^{\dagger T}\boldsymbol{H}^{\dagger})\right)^2\right\}\\ &\quad-\frac{1}{m^2}\left(Tr(\boldsymbol{C_x}+\boldsymbol{H}^{\dagger}\boldsymbol{C_w}\boldsymbol{H}^{\dagger T})\right)^2\end{aligned}\tag{55}$$

denoting $\boldsymbol{\mathcal{H}}=\boldsymbol{H}^{\dagger T}\boldsymbol{H}^{\dagger}$ we have

$$\begin{aligned}\mathbb{E}\left\{\left(Tr(\boldsymbol{y}\boldsymbol{y}^T\boldsymbol{\mathcal{H}})\right)^2\right\}&=\sum_{i,j,k,\ell=1}^{m}\mathbb{E}\{y_iy_jy_ky_\ell\}\mathcal{H}_{i,j}\mathcal{H}_{k,\ell}\\ &=\sum_{i,j=1}^{m}\mathcal{H}_{i,j}Tr\left(\mathbb{E}\{y_iy_j\boldsymbol{y}\boldsymbol{y}^T\}\boldsymbol{\mathcal{H}}\right).\end{aligned}\tag{56}$$

From [22], we have that if $\boldsymbol{y}\sim\mathcal{N}(\boldsymbol{c},\boldsymbol{\Sigma})$ then

$$\begin{aligned}\mathbb{E}\left((\boldsymbol{y}^T\boldsymbol{A}\boldsymbol{y})\boldsymbol{y}\boldsymbol{y}^T\right)&=(\boldsymbol{\Sigma}+\boldsymbol{c}\boldsymbol{c}^T)(\boldsymbol{A}+\boldsymbol{A}^T)(\boldsymbol{\Sigma}+\boldsymbol{c}\boldsymbol{c}^T)\\ &\quad+\boldsymbol{c}^T\boldsymbol{A}\boldsymbol{c}(\boldsymbol{\Sigma}-\boldsymbol{c}\boldsymbol{c}^T)+Tr(\boldsymbol{A}\boldsymbol{\Sigma})(\boldsymbol{\Sigma}+\boldsymbol{c}\boldsymbol{c}^T).\end{aligned}\tag{57}$$

Since $\boldsymbol{y}\sim\mathcal{N}(\boldsymbol{0},\boldsymbol{H}\boldsymbol{C_x}\boldsymbol{H}^T+\boldsymbol{C_w})$ we can use in (57) $\boldsymbol{c}=\boldsymbol{0}$, $\boldsymbol{\Sigma}=\boldsymbol{H}\boldsymbol{C_x}\boldsymbol{H}^T+\boldsymbol{C_w}$, and $\boldsymbol{A}=\boldsymbol{E}_{i,j}$ where $\boldsymbol{E}_{i,j}$ is an $m\times m$ matrix with all zero entries but for the i,j entry which is 1. Therefore, we have

$$\mathbb{E}\{y_iy_j\boldsymbol{y}\boldsymbol{y}^T\}=\boldsymbol{\Sigma}\left(\boldsymbol{E}_{i,j}+\boldsymbol{E}_{i,j}^T\right)\boldsymbol{\Sigma}+Tr(\boldsymbol{E}_{i,j}\boldsymbol{\Sigma})\boldsymbol{\Sigma}.\tag{58}$$

Summarizing (55), (56), and (58), we obtain (48).

Acknowledgment

The authors thank the anonymous reviewers for valuable comments that improved the presentation of this paper.

References

[1] Y. C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation of random parameters," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1931–1946, Jul. 2004.
[2] Y. C. Eldar and N. Merhav, "Minimax MSE-ratio estimation with signal covariance uncertainties," *IEEE Trans. Signal Process.*, vol. 53, no. 4, pp. 1335–1347, Apr. 2005.
[3] S. Verdu and H. V. Poor, "On minimax robustness: A general approach and applications," *IEEE Trans. Inf. Theory*, vol. 30, no. 2, Mar. 1984.
[4] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing: A survey," *Proc. IEEE*, vol. 73, no. 3, Mar. 1985.
[5] R. Mittelman and E. L. Miller, "Nonlinear filtering using a new proposal distribution and the improved fast Gauss transform with tighter performance bounds," *IEEE Trans. Signal Process.*, vol. 56, no. 12, Dec. 2008.
[6] A. T. Ihler, J. W. Fisher, and A. S. Willsky, "Particle filtering under communications constraints," in *Proc. IEEE Statist. Signal Process. Workshop 2005*.
[7] Y. Nesterov and A. Nemirovsky, *Interior-Point Polynomial Algorithms in Convex Programming*. Philadelphia, PA: SIAM, 1994.
[8] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[9] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 40–95, Mar. 1996.
[10] V. Balakrishnan and L. Vandenberghe, "Linear matrix inequalities for signal processing an overview," in *Proc. 32nd Ann. Conf. Inf. Sci. Syst.*, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, Mar. 1998.
[11] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ: Princeton Univ. Press, 1970.
[12] M. Yamashita, K. Fujisawa, and M. Kojima, "Implementation and evaluation of SDPA 6.0 (SemiDefinite Programming Algorithm 6.0)," *Optimiz. Methods Software*, vol. 18, pp. 491–505, 2003.
[13] X. Mestre, "Improved estimation of eigenvalues and eigenvectors of covariance matrices using their sample estimates," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, Nov. 2008.
[14] R. M. Gray, *Toeplitz and Circulant Matrices: A Review*. Boston, MA: Now, 2005.
[15] C. Y. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proc. IEEE*, vol. 91, no. 8, Aug. 2003.
[16] E. B. Sudderth, M. J. Wainwright, and A. S. Willsky, "Embedded trees: Estimation of gaussian processes on graphs with cycles," *IEEE Trans. Signal Process.*, vol. 54, no. 6, Jun. 2006.
[17] V. Delouille, R. Neelamani, and R. G. Baraniuk, "Robust distributed estimation using the embedded subgraph algorithm," *IEEE Trans. Signal Process.*, vol. 54, no. 8, Aug. 2006.
[18] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. Singapore: McGraw-Hill, 1991.
[19] C. E. Rasumussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA: MIT Press, 2006.
[20] M. Seeger, "Gaussian processes for machine learning," *Int. J. Neural Syst.*, vol. 14, no. 2, pp. 69–106, 2004.
[21] A. Krause, A. Singh, and C. Guestrin, "Near optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, 2008.
[22] M. Brookes, The Matrix Reference Manual, 2005 [Online]. Available: http://www.ee.ic.ac.uk/hp/staff/dmb/matrix/intro.html

Roni Mittelman (S'08–M'09) received the B.Sc. and M.Sc. (*cum laude*) degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, and the Ph.D. degree in electrical engineering from Northeastern University, Boston, MA, in 2002, 2006, and 2009 respectively.

Currently, he is a Postdoctoral Fellow with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor. His research interests include statistical signal processing and machine learning.

Eric L. Miller (S'90–M'95–SM'03) received the S.B. degree in 1990, the S.M. degree in 1992, and the Ph.D. degree in 1994, all in electrical engineering and computer science, from the Massachusetts Institute of Technology, Cambridge.

He is currently a Professor with the Department of Electrical and Computer Engineering and an Adjunct Professor of Computer Science at Tufts University, Medford, MA. Since September 2009, he has served as the Associate Dean of Research for Tufts' School of Engineering. His research interests include physics-based tomographic image formation and object characterization, inverse problems in general and inverse scattering in particular, regularization, statistical signal and imaging processing, and computational physical modeling. This work has been carried out in the context of applications including medical imaging, nondestructive evaluation, environmental monitoring and remediation, landmine and unexploded ordnance remediation, and automatic target detection and classification.

Dr. Miller is a member of Tau Beta Pi, Phi Beta Kappa, and Eta Kappa Nu. He received the CAREER Award from the National Science Foundation in 1996 and the Outstanding Research Award from the College of Engineering at Northeastern University in 2002. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and was in the same position for the IEEE TRANSACTIONS ON IMAGE PROCESSING from 1998 to 2002. He was the Co-General Chair of the 2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA.