

Robust least squares methods under bounded data uncertainties



N. Denizcan Vanli^{a,*}, Mehmet A. Donmez^b, Suleyman S. Kozat^a

^a Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey

^b Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, IL, USA

ARTICLE INFO

Article history:

Available online 24 October 2014

Keywords:

Data estimation
Least squares
Robust
Minimax
Regret

ABSTRACT

We study the problem of estimating an unknown deterministic signal that is observed through an unknown deterministic data matrix under additive noise. In particular, we present a minimax optimization framework to the least squares problems, where the estimator has imperfect data matrix and output vector information. We define the performance of an estimator relative to the performance of the optimal least squares (LS) estimator tuned to the underlying unknown data matrix and output vector, which is defined as the regret of the estimator. We then introduce an efficient robust LS estimation approach that minimizes this regret for the worst possible data matrix and output vector, where we refrain from any structural assumptions on the data. We demonstrate that minimizing this worst-case regret can be cast as a semi-definite programming (SDP) problem. We then consider the regularized and structured LS problems and present novel robust estimation methods by demonstrating that these problems can also be cast as SDP problems. We illustrate the merits of the proposed algorithms with respect to the well-known alternatives in the literature through our simulations.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

In this paper, we investigate estimation of an unknown deterministic signal that is observed through a deterministic data matrix under additive noise, which models a wide range of problems in signal processing applications [1–14]. In this framework, the data matrix and the output vector are not exactly known, however, estimates for both of them as well as uncertainty bounds on the estimates are given [2,8,15–19]. Since the model parameters are not known exactly, the performances of the classical LS estimators may significantly degrade, especially when the perturbations on the data matrix and the output vector are relatively high [9,15,16,20–22]. Hence, robust estimation algorithms are needed to obtain a satisfactory performance under such perturbations. This generic framework models several real-life applications, which require estimation of a signal observed through a linear model [9,16]. As an example, this setup models realistic channel equalization scenarios, where the data matrix represents a communication channel and the data vector is the transmitted information. The channel is usually unknown, especially for wireless communications applications, and possibly can be time-varying. Hence, in practical applications, the communication channel is estimated, where this

estimate is usually subject to distortions [9,16]. Under such possible perturbations, robust equalization methods can be used to obtain a more consistent and acceptable performance compared to the LS (or MMSE) equalizer. In this sense, this formulation is comprehensive and can be used in other applications such as in feedback control systems to estimate a desired data under imperfect system knowledge.

A prevalent approach to find robust solutions to such estimation problems is the robust minimax LS method [8,9,16,23–27], in which the uncertainties in the data matrix and the output vector are incorporated into optimization framework via a minimax residual formulation and a worst-case optimization within the uncertainty bounds is performed. Although the robust LS methods are able to minimize the LS error for the worst-case perturbations, they usually provide unsatisfactory results on the average [15,23–27] due to their conservative nature. This issue is significantly exacerbated especially when the actual perturbations do not result in significant performance degradation. Another well-known approach to compensate for errors in the data matrix and the output vector is the total least squares method (TLS) [15], which may yield undesirable results since it employs a conservative approach due to data de-regularization. On the other hand, the data matrix usually has a known special structure, such as Toeplitz and Hankel, in many linear regression problems [9,15]. Hence, in [9,15], the authors illustrate that the performances of the estimators based on minimax approaches improve when such a prior knowledge on data matrix structure is integrated into the problem formulation.

* Corresponding author.

E-mail addresses: vanli@ee.bilkent.edu.tr (N.D. Vanli), donmez2@illinois.edu (M.A. Donmez), kozat@ee.bilkent.edu.tr (S.S. Kozat).

In all these methods, LS estimators under worst case perturbations are introduced to achieve robustness. However, due to this conservative problem formulation, in many practical applications, these approaches yield unsatisfactory performances [2,8,18,28–30].

In order to counterbalance this conservative nature of the robust LS methods [9], we propose a novel robust LS approach that minimizes a worst case “regret” that is defined as the difference between the squared residual error and the smallest attainable squared residual error with an LS estimator [2,8,18,28–30]. By this regret formulation, we seek a linear estimator whose performance is as close as possible to that of the optimal estimator for all possible perturbations on the data matrix and the output vector. Our main goal in proposing the minimax regret formulation is to provide a trade-off between the robust LS methods tuned to the worst possible data parameters (under the uncertainty bounds) and the optimal LS estimator tuned to the underlying unknown model parameters. Minimax regret approaches have been presented in signal processing literature to alleviate the pessimistic nature of the worst case optimization methods [2,8,18,28–30]. In [18,29], linear minimax regret estimators are introduced to minimize the mean squared error (MSE) under imperfect knowledge of channel statistics and true parameters, respectively. In [28], a minimum mean squared error (MMSE) estimation technique under imperfect channel and data knowledge is investigated. In [2], these robust estimation methods are extended to flat fading channels to perform channel equalization. These methods are shown to provide a better average performance compared to the minimax estimators, whereas under large perturbations the robustness of the minimax estimators are superior to these competitive methods. On the other hand, in this paper, the optimization frameworks investigated here are significantly different than [9,16,23–27], where the regret terms are directly adjoined in the cost functions. In particular, unlike [2,18,28,29], where the uncertainties are in the statistics of the transmitted signal or channel parameters, in this paper, the uncertainty is both on the data matrix and the output vector without any statistical assumptions. While in [8], the authors have considered a similar framework, the results of this paper build upon them and provide a complete solution to the regret based robust LS estimation methods unlike [8]. We emphasize that perturbation bounds on the data matrix and the output vector heavily depend on the estimation algorithms employed to obtain them. Since our methods are formulated for given perturbation bounds, different estimation algorithms can be readily incorporated into our framework with the corresponding perturbation bounds [16].

Our main contributions in this paper are as follows. i) We introduce a novel and efficient robust LS estimation method in which we find the transmitted signal by minimizing the worst-case regret, i.e., the worst-case difference between the residual error of the LS estimator and the residual error of the optimal LS estimator tuned to the underlying model. In this sense, we present a robust estimation method that achieves a tradeoff between the robust LS estimation methods and the direct LS estimation method tuned to the estimates of the data matrix and output vector. ii) We next propose a minimax regret formulation for the regularized LS estimation problem. iii) We then introduce a structured robust LS estimation method in which the data matrix is known to have a special structure such as Toeplitz or Hankel. iv) We demonstrate that the robust estimation methods we propose can be cast as SDP problems, hence our methods can be efficiently implemented in real-time [31]. v) In our simulations, we observe that our approaches provide better performance compared to the robust methods that are optimized with respect to the worst-case residual error [9,32], and the conventional methods that directly solve the estimation problem using the perturbed data.

The organization of the paper is as follows. An overview to the problem is provided in Section 2. In Section 3.1, we first introduce

the LS estimation method based on our regret formulation, and then present the regularized LS estimation approach in Section 3.2. We then consider the structured LS approach in Section 3.3 and provide the explicit SDP formulations for all problems. The numerical examples are demonstrated in Section 4. Finally, the paper concludes with certain remarks in Section 5.

2. System overview

2.1. Notation

In this paper, all vectors are column vectors and represented by boldface lowercase letters. Matrices are represented by boldface uppercase letters. For a matrix \mathbf{H} , \mathbf{H}^H is the conjugate transpose, $\|\mathbf{H}\|$ is the spectral norm, \mathbf{H}^+ is the pseudo-inverse, $\mathbf{H} > 0$ represents a positive definite matrix and $\mathbf{H} \geq 0$ represents a positive semi-definite matrix. For a square matrix \mathbf{H} , $\text{Tr}(\mathbf{H})$ is the trace. Naturally, for a vector \mathbf{x} , $\|\mathbf{x}\| = \sqrt{\mathbf{x}^H \mathbf{x}}$ is the ℓ^2 -norm. Here, $\mathbf{0}$ denotes a vector or matrix with all zero elements and the dimensions can be understood from the context. Similarly, \mathbf{I} represents the appropriate sized identity matrix. The operator $\text{vec}(\cdot)$ is the vectorization operator, i.e., it stacks the columns of a matrix of dimension $m \times n$ into an $mn \times 1$ column vector. Finally, the operator \otimes is the Kronecker product [33].

2.2. Problem description

We investigate the problem of estimating an unknown deterministic vector $\mathbf{x} \in \mathbb{C}^n$ which is observed through a deterministic data matrix. However, instead of the actual data matrix and the output vector, their estimates $\mathbf{H} \in \mathbb{C}^{m \times n}$ and $\mathbf{y} \in \mathbb{C}^m$ and uncertainty bounds on these estimates are provided. In this sense, our aim is to find a solution to the following data estimation problem

$$\mathbf{y} \approx \mathbf{H}\mathbf{x},$$

such that

$$\mathbf{y} + \Delta\mathbf{y} = (\mathbf{H} + \Delta\mathbf{H})\mathbf{x},$$

for deterministic perturbations $\Delta\mathbf{H} \in \mathbb{C}^{m \times n}$, $\Delta\mathbf{y} \in \mathbb{C}^m$. Although these perturbations are unknown, a bound on each perturbation is provided, i.e.,

$$\|\Delta\mathbf{H}\| \leq \delta_H \quad \text{and} \quad \|\Delta\mathbf{y}\| \leq \delta_Y,$$

where $\delta_H, \delta_Y \geq 0$. In this sense, we refrain from any assumptions on the data matrix and the output vector, yet consider that the estimates \mathbf{H} and \mathbf{y} are at least accurate to “some degree” but their actual values under these uncertainties are completely unknown to the estimator.

Even in the presence of these uncertainties, the symbol vector \mathbf{x} can be naively estimated by simply substituting the estimates \mathbf{H} and \mathbf{y} into the LS estimator [10]. For the LS estimator we have

$$\hat{\mathbf{x}} = \mathbf{H}^+ \mathbf{y},$$

where \mathbf{H}^+ is the pseudo-inverse of \mathbf{H} [33]. However, this approach yields unsatisfactory results, when the errors in the estimates of the data matrix and the output vector are relatively high [9,18,29,32]. A common approach to find a robust solution is to employ a worst-case residual minimization [9]

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^n} \max_{\|\Delta\mathbf{H}\| \leq \delta_H, \|\Delta\mathbf{y}\| \leq \delta_Y} \left\| (\mathbf{y} + \Delta\mathbf{y}) - (\mathbf{H} + \Delta\mathbf{H})\mathbf{x} \right\|^2,$$

where \mathbf{x} is chosen to minimize the worst-case residual error in the uncertainty region. However, since the solution is found with

respect to the worst possible data matrix and output vector in the uncertainty regions, it may be highly conservative [15,18,29].

Here, we propose a novel LS estimation approach that provides a tradeoff between performance and robustness in order to mitigate the conservative nature of the worst-case residual minimization approach as well as to preserve robustness [18,29]. The regret for not using the optimal LS estimator is defined as the difference between the residual error with an estimate of the input vector and the residual error with the optimal LS estimator, i.e.,

$$\mathcal{R}(\mathbf{x}; \Delta\mathbf{H}, \Delta\mathbf{y}) \triangleq \left\| (\mathbf{y} + \Delta\mathbf{y}) - (\mathbf{H} + \Delta\mathbf{H})\mathbf{x} \right\|^2 - \min_{\mathbf{w} \in \mathbb{C}^n} \left\| (\mathbf{y} + \Delta\mathbf{y}) - (\mathbf{H} + \Delta\mathbf{H})\mathbf{w} \right\|^2. \quad (1)$$

By making such a regret definition, we force our estimator not to construct the symbol vector according to the worst possible scenario considering that it may be too conservative. Instead, we define the regret of any estimator by the difference in the estimation performances of that estimator and the “smartest” estimator knowing both data matrix and output vector in hindsight, so that we achieve a tradeoff between robustness and estimation performance.

We emphasize that the regret defined in (1) is completely different than the regret formulation introduced in [18,29]. In (1), the uncertainty is on the data matrix where the desired data vector \mathbf{x} is completely unknown, unlike [18,29]. We emphasize that we use the residual error $\|(\mathbf{y} + \Delta\mathbf{y}) - (\mathbf{H} + \Delta\mathbf{H})\mathbf{x}\|^2$ instead of the estimation error $\|\hat{\mathbf{x}} - \mathbf{x}\|$ since the estimation error directly depends on the vector \mathbf{x} and cannot be used in the regret formulation since \mathbf{x} is assumed to be unknown in the presence of data uncertainties. Moreover, in our formulation, the estimate $\hat{\mathbf{x}}$ is not constrained to be linear unlike [18,29] since our regret formulation is well-defined without any limitations on the estimated $\hat{\mathbf{x}}$.

In the next sections, the proposed approaches to the robust LS estimation problems are provided. We first introduce the regret based unstructured LS estimation method. We next present the unstructured regularized LS estimation approach in which the worst-case regret is optimized. Finally, we investigate the structured LS estimation approach.

3. Robust least squares estimation methods

3.1. Unstructured robust least squares estimation

In this section, we provide a novel robust unstructured LS estimator based on a certain minimax criterion. We consider the most generic estimation problem

$$\min_{\mathbf{x} \in \mathbb{C}^n} \max_{\|\Delta\mathbf{H}\| \leq \delta_H, \|\Delta\mathbf{y}\| \leq \delta_Y} \mathcal{R}(\mathbf{x}; \Delta\mathbf{H}, \Delta\mathbf{y}), \quad (2)$$

where $\mathcal{R}(\mathbf{x}; \Delta\mathbf{H}, \Delta\mathbf{y})$ is defined as in (1). Now considering the second term in (1), we define $\tilde{\mathbf{H}} \triangleq \mathbf{H} + \Delta\mathbf{H}$, $\tilde{\mathbf{y}} \triangleq \mathbf{y} + \Delta\mathbf{y}$, where $\tilde{\mathbf{H}}$ is a full rank matrix, and denote the estimation performance of the optimal LS estimator for some given $\tilde{\mathbf{H}}$ and $\tilde{\mathbf{y}}$ by

$$f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) \triangleq \min_{\mathbf{w} \in \mathbb{C}^n} \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{w}\|^2.$$

Since we consider an unconstrained minimization over \mathbf{w} , we have [10]

$$\begin{aligned} \mathbf{w}^* &\triangleq \arg \min_{\mathbf{w} \in \mathbb{C}^n} \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{w}\|^2 \\ &= \tilde{\mathbf{H}}^+ \tilde{\mathbf{y}}, \end{aligned} \quad (3)$$

as the optimal data vector minimizing the residual error. Then we have

$$\begin{aligned} f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) &= \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{w}^*\|^2 \\ &= (\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{w}^*)^H (\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{w}^*) \\ &= \tilde{\mathbf{y}}^H (\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{w}^*) \\ &= \tilde{\mathbf{y}}^H \tilde{\mathbf{P}} \tilde{\mathbf{y}}, \end{aligned}$$

where the third line follows from $\tilde{\mathbf{H}}^H \tilde{\mathbf{H}}\mathbf{w}^* = \tilde{\mathbf{H}}^H \tilde{\mathbf{y}}$ [10] and $\tilde{\mathbf{P}} \triangleq \mathbf{I} - \tilde{\mathbf{H}}\tilde{\mathbf{H}}^+$ is the projection matrix of the space perpendicular to the range space of $\tilde{\mathbf{H}}$. If we use the Taylor series expansion based on Wirtinger calculus [33] for $f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})$ around $\tilde{\mathbf{H}} = \mathbf{H}$ and $\tilde{\mathbf{y}} = \mathbf{y}$, then

$$\begin{aligned} f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) &= f(\mathbf{H}, \mathbf{y}) + 2 \operatorname{Re} \left\{ \operatorname{Tr} \left(\nabla f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) \Big|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}}^H [\Delta\mathbf{H} \Delta\mathbf{y}] \right) \right\} \\ &\quad + O(\|[\Delta\mathbf{H} \Delta\mathbf{y}]\|^2). \end{aligned} \quad (4)$$

Note that the first order Taylor approximation is introduced in order to obtain a tractable solution. Clearly, the effect of using this approximation vanishes as $\|[\Delta\mathbf{H} \Delta\mathbf{y}]\|$ decreases and for distortions with larger $\|[\Delta\mathbf{H} \Delta\mathbf{y}]\|$, one can easily use higher order approximations instead. However, we observe through our simulations that even for relatively large perturbations, a satisfactory performance is obtained using this approximation.

We now introduce the following lemma in order to obtain the first order Taylor approximation in (4) in a closed form.

Lemma 1. Let $\tilde{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}$ be a full rank matrix and $\tilde{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y}$, where $\tilde{\mathbf{H}} \in \mathbb{C}^{m \times n}$ and $\tilde{\mathbf{y}} \in \mathbb{C}^m$. Then defining $f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) \triangleq \tilde{\mathbf{y}}^H \tilde{\mathbf{P}} \tilde{\mathbf{y}}$, where $\tilde{\mathbf{P}} \triangleq \mathbf{I} - \tilde{\mathbf{H}}\tilde{\mathbf{H}}^+$, we have

$$\frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \tilde{\mathbf{H}}} \Big|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}} = -\mathbf{P}\mathbf{y}(\mathbf{H}^+\mathbf{y})^H,$$

and

$$\frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \tilde{\mathbf{y}}} \Big|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}} = \mathbf{P}\mathbf{y},$$

where $\mathbf{P} \triangleq \mathbf{I} - \mathbf{H}\mathbf{H}^+$.

Proof of Lemma 1. Since $\tilde{\mathbf{H}}$ is full rank and $m \geq n$, the pseudo-inverse of $\tilde{\mathbf{H}}$ is found by [33]

$$\tilde{\mathbf{H}}^+ \triangleq (\tilde{\mathbf{H}}^H \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^H.$$

Hence, we have [33]

$$\begin{aligned} \mathbf{D} &= \frac{\partial}{\partial \tilde{\mathbf{H}}} (\tilde{\mathbf{y}}^H \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^H \tilde{\mathbf{H}} (\tilde{\mathbf{H}}^H \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^H \tilde{\mathbf{y}}) \Big|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}} \\ &= \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H \mathbf{y} \mathbf{y}^H \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} - \mathbf{y} \mathbf{y}^H \mathbf{H}(\mathbf{H}^H \mathbf{H})^{-1} \\ &= \mathbf{H}\mathbf{H}^+ \mathbf{y}(\mathbf{H}^+\mathbf{y})^H - \mathbf{y}(\mathbf{H}^+\mathbf{y})^H \\ &= -\mathbf{P}\mathbf{y}(\mathbf{H}^+\mathbf{y})^H, \end{aligned} \quad (5)$$

and

$$\begin{aligned} \mathbf{b} &= \frac{\partial}{\partial \tilde{\mathbf{y}}} (\tilde{\mathbf{y}}^H \tilde{\mathbf{y}} - \tilde{\mathbf{y}}^H \tilde{\mathbf{H}} (\tilde{\mathbf{H}}^H \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}^H \tilde{\mathbf{y}}) \Big|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}} \\ &= \mathbf{P}\mathbf{y}, \end{aligned} \quad (6)$$

where the last line of the equality follows since $\mathbf{H}\mathbf{H}^+$ is a symmetric matrix according to the definition of the pseudo-inverse operation. This concludes the proof of Lemma 1. \square

Now turning our attention back to (4), we denote

$$\mathbf{D} \triangleq \frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \tilde{\mathbf{H}}} \Big|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}},$$

and

$$\mathbf{b} \triangleq \left. \frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \tilde{\mathbf{y}}} \right|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}},$$

where we emphasize that the closed form definitions of \mathbf{D} and \mathbf{b} can be obtained from Lemma 1. We then approximate (4) and obtain the first order Taylor approximation as follows

$$\begin{aligned} f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) &\approx f(\mathbf{H}, \mathbf{y}) + 2\text{Re}\{\text{Tr}([\mathbf{D}\mathbf{b}]^H[\Delta\mathbf{H}\Delta\mathbf{y}])\} \\ &= \kappa + 2\text{Re}\{(\text{vec}(\mathbf{D})^H \text{vec}(\Delta\mathbf{H}) + \mathbf{b}^H \Delta\mathbf{y})\} \\ &= \kappa + \mathbf{d}^H \Delta\mathbf{h} + \Delta\mathbf{h}^H \mathbf{d} + \mathbf{b}^H \Delta\mathbf{y} + \Delta\mathbf{y}^H \mathbf{b}, \end{aligned} \quad (7)$$

where $\kappa \triangleq f(\mathbf{H}, \mathbf{y})$, $\mathbf{d} \triangleq \text{vec}(\mathbf{D})$, and $\Delta\mathbf{h} \triangleq \text{vec}(\Delta\mathbf{H})$. Hence we can approximate the regret in (1) as follows

$$\begin{aligned} \mathcal{R}(\mathbf{x}; \Delta\mathbf{H}, \Delta\mathbf{y}) &\approx \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{x}\|^2 \\ &\quad - (\kappa + \mathbf{d}^H \Delta\mathbf{h} + \Delta\mathbf{h}^H \mathbf{d} + \mathbf{b}^H \Delta\mathbf{y} + \Delta\mathbf{y}^H \mathbf{b}). \end{aligned} \quad (8)$$

In the following theorem, we illustrate how the optimization (or equivalently estimation) problem in (8) can be put in an SDP form.

Theorem 1. Let $\mathbf{H} \in \mathbb{C}^{m \times n}$ and $\mathbf{y} \in \mathbb{C}^m$ be the estimates of the data matrix and the output vector, respectively, both having deterministic additive perturbations $\Delta\mathbf{H} \leq \delta_H$ and $\Delta\mathbf{y} \leq \delta_Y$, respectively, i.e., $\tilde{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y}$, where $\tilde{\mathbf{H}}$ is the full rank data matrix, $\tilde{\mathbf{y}}$ is the output vector, and $m \geq n$. Then the problem

$$\min_{\mathbf{x} \in \mathbb{C}^n} \max_{\|\Delta\mathbf{H}\| \leq \delta_H, \|\Delta\mathbf{y}\| \leq \delta_Y} \mathcal{R}(\mathbf{x}; \Delta\mathbf{H}, \Delta\mathbf{y}), \quad (9)$$

where $\mathcal{R}(\mathbf{x}; \Delta\mathbf{H}, \Delta\mathbf{y})$ is defined as in (8), is equivalent to solving the following SDP problem

$$\begin{aligned} &\min \gamma \\ &\text{subject to} \\ &\tau_1 \geq 0, \tau_2 \geq 0, \text{ and} \\ &\begin{bmatrix} \gamma + \kappa - \tau_1 - \tau_2 & (\mathbf{y} - \mathbf{H}\mathbf{x})^H & \delta_Y \mathbf{b}^H & \delta_H \mathbf{d}^H \\ \mathbf{y} - \mathbf{H}\mathbf{x} & \mathbf{I} & -\delta_Y \mathbf{I} & \delta_H \mathbf{X} \\ \delta_Y \mathbf{b} & -\delta_Y \mathbf{I} & \tau_1 \mathbf{I} & \mathbf{0} \\ \delta_H \mathbf{d} & \delta_H \mathbf{X}^H & \mathbf{0} & \tau_2 \mathbf{I} \end{bmatrix} \geq 0, \end{aligned} \quad (10)$$

where \mathbf{X} is the $m \times mn$ matrix defined as $\mathbf{X} \triangleq \mathbf{x}^H \otimes \mathbf{I}$.

The proof of Theorem 1 is provided in Appendix A.

Remark 1. In the proof of Theorem 1, we use Proposition 1 that relies on the lossless S -procedure. However, S -procedure is lossless with two constraints when the corresponding two quadratic (Hermitian) forms on the complex linear space [34]. However, classical S -procedure for quadratic forms is, in general, lossy with two constraints in the real case [35]. Hence, Theorem 1 cannot be extended for real linear space.

Now we can consider two important corollaries of Theorem 1. First, a special case of Theorem 1 in which the uncertainty is only in the data matrix. We emphasize that the perturbation errors only in the data matrix are also common in a wide range of real life applications [10]. Here, we can define the regret as follows

$$\mathcal{R}(\mathbf{x}; \Delta\mathbf{H}) \triangleq \|\mathbf{y} - \tilde{\mathbf{H}}\mathbf{x}\|^2 - \min_{\mathbf{w} \in \mathbb{C}^n} \|\mathbf{y} - \tilde{\mathbf{H}}\mathbf{w}\|^2, \quad (11)$$

and similar to the previous case, we calculate the optimal estimation performance under a given uncertainty bound

$$\begin{aligned} f(\tilde{\mathbf{H}}) &\triangleq \min_{\mathbf{w} \in \mathbb{C}^n} \|\mathbf{y} - \tilde{\mathbf{H}}\mathbf{w}\|^2 \\ &\approx \kappa + 2\text{Re}\{\text{Tr}(\nabla f(\tilde{\mathbf{H}}, \mathbf{y}) \Big|_{\tilde{\mathbf{H}}=\mathbf{H}}^H \Delta\mathbf{H})\} \\ &= \kappa + 2\text{Re}\{\text{vec}(\mathbf{D}^H) \text{vec}(\Delta\mathbf{H})\} \\ &= \kappa + \mathbf{d}^H \Delta\mathbf{h} + \Delta\mathbf{h}^H \mathbf{d}. \end{aligned}$$

Hence we approximate the regret in (11) as follows

$$\mathcal{R}(\mathbf{x}; \Delta\mathbf{H}) \approx \|\mathbf{y} - \tilde{\mathbf{H}}\mathbf{x}\|^2 - (\kappa + \mathbf{d}^H \Delta\mathbf{h} + \Delta\mathbf{h}^H \mathbf{d}). \quad (12)$$

Corollary 1. Let $\mathbf{H} \in \mathbb{C}^{m \times n}$ and $\mathbf{y} \in \mathbb{C}^m$ be the estimates of the data matrix and the output vector, respectively, where $m \geq n$. Suppose there is a bounded uncertainty on the full rank data matrix $\tilde{\mathbf{H}}$, i.e., $\tilde{\mathbf{H}} = \mathbf{H} + \Delta\mathbf{H}$, $\|\Delta\mathbf{H}\| \leq \delta_H$. Then the problem

$$\min_{\mathbf{x} \in \mathbb{C}^n} \max_{\|\Delta\mathbf{H}\| \leq \delta_H} \mathcal{R}(\mathbf{x}; \Delta\mathbf{H}), \quad (13)$$

where $\mathcal{R}(\mathbf{x}; \Delta\mathbf{H})$ is defined as in (12), is equivalent to solving the following SDP problem

$$\begin{aligned} &\min \gamma \\ &\text{subject to} \\ &\tau \geq 0 \text{ and } \begin{bmatrix} \gamma + \kappa - \tau & (\mathbf{y} - \mathbf{H}\mathbf{x})^H & \delta_H \mathbf{d} \\ \mathbf{y} - \mathbf{H}\mathbf{x} & \mathbf{I} & \delta_H \mathbf{X} \\ \delta_H \mathbf{d} & \delta_H \mathbf{X}^H & \tau \mathbf{I} \end{bmatrix} \geq 0. \end{aligned} \quad (14)$$

Outline of the proof of Corollary 1. The proof of Corollary 1 can be explicitly derived from the proof of Theorem 1 by simply setting $\delta_Y = 0$ and $\tau_1 = 0$, hence is omitted. \square

Second, we consider another special case of Theorem 1 in which the uncertainty is only in the output vector. We emphasize that similar to the previous case, this one is also a common case in a wide range of real-life applications [10], and studied under a similar framework in [18]. Here, we can define the regret as follows

$$\mathcal{R}(\mathbf{x}; \Delta\mathbf{y}) \triangleq \|\tilde{\mathbf{y}} - \mathbf{H}\mathbf{x}\|^2 - \min_{\mathbf{w} \in \mathbb{C}^n} \|\tilde{\mathbf{y}} - \mathbf{H}\mathbf{w}\|^2, \quad (15)$$

and similar to the previous case, we calculate the optimal also performance under a given uncertainty bound

$$\begin{aligned} f(\tilde{\mathbf{y}}) &\triangleq \min_{\mathbf{w} \in \mathbb{C}^n} \|\tilde{\mathbf{y}} - \mathbf{H}\mathbf{w}\|^2 \\ &\approx \kappa + 2\text{Re}\{\text{Tr}(\nabla f(\mathbf{H}, \tilde{\mathbf{y}}) \Big|_{\tilde{\mathbf{y}}=\mathbf{y}}^H \Delta\mathbf{y})\} \\ &= \kappa + 2\text{Re}\{\mathbf{b}^H \Delta\mathbf{y}\} \\ &= \kappa + \mathbf{b}^H \Delta\mathbf{y} + \Delta\mathbf{y}^H \mathbf{b}. \end{aligned}$$

Hence we approximate the regret in (15) as follows

$$\mathcal{R}(\mathbf{x}; \Delta\mathbf{y}) \approx \|\tilde{\mathbf{y}} - \mathbf{H}\mathbf{x}\|^2 - (\kappa + \mathbf{b}^H \Delta\mathbf{y} + \Delta\mathbf{y}^H \mathbf{b}). \quad (16)$$

Corollary 2. Let $\mathbf{H} \in \mathbb{C}^{m \times n}$ and $\mathbf{y} \in \mathbb{C}^m$ be the estimates of the data matrix and the output vector, respectively, where $m \geq n$. Suppose there is a bounded uncertainty on the output vector $\tilde{\mathbf{y}}$, i.e., $\tilde{\mathbf{y}} = \mathbf{y} + \Delta\mathbf{y}$, $\|\Delta\mathbf{y}\| \leq \delta_Y$. Then the problem

$$\min_{\mathbf{x} \in \mathbb{C}^n} \max_{\|\Delta\mathbf{y}\| \leq \delta_Y} \mathcal{R}(\mathbf{x}; \Delta\mathbf{y}), \quad (17)$$

where $\mathcal{R}(\mathbf{x}; \Delta\mathbf{y})$ is defined as in (16), is equivalent to solving the following SDP problem

$$\begin{aligned} & \min \gamma \\ & \text{subject to} \\ & \tau \geq 0 \quad \text{and} \quad \begin{bmatrix} \gamma + \kappa - \tau & (\mathbf{y} - \mathbf{H}\mathbf{x})^H & \delta_Y \mathbf{b}^H \\ \mathbf{y} - \mathbf{H}\mathbf{x} & \mathbf{I} & -\delta_Y \mathbf{I} \\ \delta_Y \mathbf{b} & -\delta_Y \mathbf{I} & \tau \mathbf{I} \end{bmatrix} \geq 0. \end{aligned} \quad (18)$$

Outline of the proof of Corollary 2. The proof of Corollary 2 can be explicitly derived from the proof of Theorem 1 by simply setting $\delta_H = 0$ and $\tau_2 = 0$, hence is omitted. \square

Remark 2. Corollaries 1 and 2 follow from the proof of Theorem 1, which relies on the lossless S -procedure. Under the frameworks presented in Corollaries 1 and 2, one can safely extend the same conclusions for the real case also, since S -procedure is lossless for quadratic forms with one constraint both in complex and real spaces [36,37].

3.2. Unstructured robust regularized least squares estimation

In this section, we introduce a worst-case regret optimization approach to solve the regularized LS estimation problem in [32]. The regret for not using the optimal regularized LS estimator is defined by

$$\mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}) \triangleq \{ \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{x}\|^2 + \mu \|\mathbf{x}\|^2 \} - \min_{\mathbf{w} \in \mathbb{C}^n} \{ \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{w}\|^2 + \mu \|\mathbf{w}\|^2 \}, \quad (19)$$

where $\mu > 0$ is the regularization parameter. We emphasize that there are different approaches to choose μ , however, for the focus of this paper, we assume that it is already set before the optimization so that these methods can be readily incorporated in our framework. Hence, we solve the regularized LS estimation problem for an arbitrary $\mu > 0$ and note that we have already covered the $\mu = 0$ case in Section 3.1.

Similar to the previous case, we denote the estimation error of the optimal LS estimator for some estimated data matrix \mathbf{H} and output vector \mathbf{y} by

$$\begin{aligned} f(\mathbf{H}, \mathbf{y}) & \triangleq \min_{\mathbf{w} \in \mathbb{C}^n} \|\mathbf{y} - \mathbf{H}\mathbf{w}\|^2 + \mu \|\mathbf{w}\|^2 \\ & = \|\mathbf{P}^{-1} \mathbf{y}\|^2 \\ & = \mathbf{y}^H \mathbf{P}^{-1} \mathbf{y}, \end{aligned}$$

where $\mathbf{P} \triangleq \mathbf{I} + \mu^{-1} \mathbf{H}\mathbf{H}^H$. Considering the first order Taylor series expansion based on Wirtinger calculus [33] for $f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})$ around $\tilde{\mathbf{H}} = \mathbf{H}$ and $\tilde{\mathbf{y}} = \mathbf{y}$

$$\begin{aligned} f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) & \approx \kappa + 2 \operatorname{Re} \{ \operatorname{Tr}(\nabla f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) \Big|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}}^H [\Delta \mathbf{H} \Delta \mathbf{y}]) \}, \\ & = \kappa + \mathbf{d}^H \Delta \mathbf{h} + \Delta \mathbf{h}^H \mathbf{d} + \mathbf{b}^H \Delta \mathbf{y} + \Delta \mathbf{y}^H \mathbf{b}, \end{aligned}$$

where $\mathbf{d} \triangleq \operatorname{vec}(\mathbf{D}^H)$, $\Delta \mathbf{h} \triangleq \operatorname{vec}(\Delta \mathbf{H})$,

$$\begin{aligned} \mathbf{D} & \triangleq \left. \frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \tilde{\mathbf{H}}} \right|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}} \\ & = -\mathbf{P}^{-1} \mathbf{y} \mathbf{y}^H \mathbf{P}^{-1} \mathbf{H}, \end{aligned} \quad (20)$$

and

$$\begin{aligned} \mathbf{b} & \triangleq \left. \frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \tilde{\mathbf{y}}} \right|_{\tilde{\mathbf{H}}=\mathbf{H}, \tilde{\mathbf{y}}=\mathbf{y}} \\ & = \mathbf{P}^{-1} \mathbf{y}, \end{aligned}$$

where the last line follows since \mathbf{P} is symmetric. Hence we can approximate the regret in (19) as follows

$$\begin{aligned} \mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}) & \approx \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{x}\|^2 + \mu \|\mathbf{x}\|^2 - (\kappa + \mathbf{d}^H \Delta \mathbf{h} \\ & \quad + \Delta \mathbf{h}^H \mathbf{d} + \mathbf{b}^H \Delta \mathbf{y} + \Delta \mathbf{y}^H \mathbf{b}), \end{aligned} \quad (21)$$

similar to (8). In the following theorem, we illustrate how the optimization problem in (21) can be put in an SDP form.

Theorem 2. Let $\mathbf{H} \in \mathbb{C}^{m \times n}$ and $\mathbf{y} \in \mathbb{C}^m$ be the estimates of the data matrix and the output vector, respectively, both having deterministic additive perturbations $\Delta \mathbf{H} \leq \delta_H$ and $\Delta \mathbf{y} \leq \delta_Y$, respectively, i.e., $\tilde{\mathbf{H}} = \mathbf{H} + \Delta \mathbf{H}$ and $\tilde{\mathbf{y}} = \mathbf{y} + \Delta \mathbf{y}$, where $\tilde{\mathbf{H}}$ is the full rank data matrix, $\tilde{\mathbf{y}}$ is the output vector, and $m \geq n$. Then the problem

$$\min_{\mathbf{x} \in \mathbb{C}^n} \max_{\|\Delta \mathbf{H}\| \leq \delta_H, \|\Delta \mathbf{y}\| \leq \delta_Y} \mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}), \quad (22)$$

where $\mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y})$ is defined as in (21), is equivalent to solving the following SDP problem

$$\begin{aligned} & \min \gamma \\ & \text{subject to} \\ & \tau_1 \geq 0, \tau_2 \geq 0, \quad \text{and} \\ & \begin{bmatrix} \gamma + \kappa - \tau_1 - \tau_2 & (\mathbf{y} - \mathbf{H}\mathbf{x})^H & \mathbf{x}^H & \delta_Y \mathbf{b}^H & \delta_H \mathbf{d}^H \\ \mathbf{y} - \mathbf{H}\mathbf{x} & \mathbf{I} & \mathbf{0} & -\delta_Y \mathbf{I} & \delta_H \mathbf{X} \\ \mathbf{x} & \mathbf{0} & \mu \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \delta_Y \mathbf{b} & -\delta_Y \mathbf{I} & \mathbf{0} & \tau_1 \mathbf{I} & \mathbf{0} \\ \delta_H \mathbf{d} & \delta_H \mathbf{X}^H & \mathbf{0} & \mathbf{0} & \tau_2 \mathbf{I} \end{bmatrix} \geq 0. \end{aligned} \quad (23)$$

Proof of Theorem 2. The proof of Theorem 2 follows similar lines to the proof of Theorem 1, hence is omitted here. \square

Remark 3. Under the framework introduced in this section, one can straightforwardly obtain the corollaries similar to Corollaries 1 and 2 by considering cases in which the uncertainty is either only on the data matrix or only on the output vector, i.e., $\delta_Y = 0$ and $\delta_H = 0$ cases, respectively. The derivations follow similar lines to Corollaries 1, 2 and Theorem 2, hence is omitted. However, similar results can be readily derived from the result in Theorem 2 with suitable changes in the SDP formulations.

3.3. Structured robust least squares estimation

There are various communication systems where the data matrix and the perturbation on it have a special structure such as Toeplitz, Hankel, or Vandermonde [9,15]. Incorporating this prior knowledge into the estimation framework could improve the performance of the regret based minimax LS estimation approach [9, 15]. Hence, in this section, we investigate a special case of the problem in (2), where the associated perturbations for the data matrix \mathbf{H} and the output vector \mathbf{y} have special structures. The structure on the perturbations is defined as follows

$$\Delta \mathbf{H} = \sum_{i=1}^p \alpha_i \mathbf{H}_i, \quad (24)$$

and

$$\Delta \mathbf{y} = \sum_{i=1}^p \beta_i \mathbf{y}_i, \quad (25)$$

where $\mathbf{H}_i \in \mathbb{C}^{m \times n}$, $\mathbf{y}_i \in \mathbb{C}^m$, and p are known but $\alpha_i, \beta_i \in \mathbb{C}$, $i = 1, \dots, p$, are unknown. However, the bounds on the norm of $\boldsymbol{\alpha} \triangleq [\alpha_1, \dots, \alpha_p]^H$ and $\boldsymbol{\beta} \triangleq [\beta_1, \dots, \beta_p]^H$ are provided as $\|\boldsymbol{\alpha}\| \leq \delta_\alpha$ and $\|\boldsymbol{\beta}\| \leq \delta_\beta$, where $\delta_\alpha, \delta_\beta \geq 0$. We emphasize that this formulation can represent a wide range of constraints on the structure of perturbations of the data matrix and the output vector such as

Toeplitz and Hankel [9,10]. Our aim is to solve the following optimization problem

$$\min_{\mathbf{x} \in \mathbb{C}^n} \max_{\|\boldsymbol{\alpha}\| \leq \delta_\alpha, \|\boldsymbol{\beta}\| \leq \delta_\beta} \mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}),$$

where

$$\mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}) \triangleq \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{x}\|^2 - \min_{\mathbf{w} \in \mathbb{C}^n} \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{w}\|^2, \quad (26)$$

$$\tilde{\mathbf{H}} \triangleq \mathbf{H} + \Delta \mathbf{H} = \mathbf{H} + \sum_{i=1}^p \alpha_i \mathbf{H}_i, \quad (27)$$

$$\tilde{\mathbf{y}} \triangleq \mathbf{y} + \Delta \mathbf{y} = \mathbf{y} + \sum_{i=1}^p \beta_i \mathbf{y}_i. \quad (28)$$

After following similar lines to Section 3.1, and introducing the first order Taylor approximation to $f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})$ around $\boldsymbol{\alpha} = \mathbf{0}$ and $\boldsymbol{\beta} = \mathbf{0}$, we obtain

$$f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) \approx \kappa + 2 \operatorname{Re} \left\{ \operatorname{Tr} \left(\nabla f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) \Big|_{\boldsymbol{\alpha}=\mathbf{0}, \boldsymbol{\beta}=\mathbf{0}} [\boldsymbol{\alpha} \ \boldsymbol{\beta}] \right) \right\}, \quad (29)$$

where $f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) = \tilde{\mathbf{y}}^H \tilde{\mathbf{P}} \tilde{\mathbf{y}}$ and $\tilde{\mathbf{P}} = \mathbf{I} - \tilde{\mathbf{H}}\tilde{\mathbf{H}}^+$. We next introduce the following lemma to calculate the first order Taylor approximation in (29) in a closed form.

Lemma 2. Let $\tilde{\mathbf{H}} = \mathbf{H} + \Delta \mathbf{H}$ be a full rank matrix and $\tilde{\mathbf{y}} = \mathbf{y} + \Delta \mathbf{y}$, where $\tilde{\mathbf{H}} \in \mathbb{C}^{m \times n}$, $\tilde{\mathbf{y}} \in \mathbb{C}^m$, $\Delta \mathbf{H}$ and $\Delta \mathbf{y}$ are defined as in (24) and (25), respectively. Then denoting $f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) \triangleq \tilde{\mathbf{y}}^H \tilde{\mathbf{P}} \tilde{\mathbf{y}}$, where $\tilde{\mathbf{P}} \triangleq \mathbf{I} - \tilde{\mathbf{H}}\tilde{\mathbf{H}}^+$, we have

$$\frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \boldsymbol{\alpha}} \Big|_{\boldsymbol{\alpha}=\mathbf{0}, \boldsymbol{\beta}=\mathbf{0}} = [-\mathbf{y}^H \mathbf{P}^H \mathbf{H}_1 \mathbf{H}^+ \mathbf{y}, \dots, -\mathbf{y}^H \mathbf{P}^H \mathbf{H}_p \mathbf{H}^+ \mathbf{y}]^H, \quad (30)$$

and

$$\frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\alpha}=\mathbf{0}, \boldsymbol{\beta}=\mathbf{0}} = [\mathbf{y}^H \mathbf{P} \mathbf{y}_1, \dots, \mathbf{y}^H \mathbf{P} \mathbf{y}_p]^H, \quad (31)$$

where $\mathbf{P} \triangleq \mathbf{I} - \mathbf{H}\mathbf{H}^+$.

Proof of Lemma 2. Note that the derivative of $f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})$ is taken with respect to $[\boldsymbol{\alpha} \ \boldsymbol{\beta}]$, hence we can use the Chain Rule to calculate the derivatives by using the results we have obtained in Lemma 1.

First, we consider the derivative of $f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})$ with respect to α_i , $i = 1, \dots, p$, i.e.,

$$\begin{aligned} d_i &\triangleq \frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \alpha_i} \Big|_{\boldsymbol{\alpha}=\mathbf{0}, \boldsymbol{\beta}=\mathbf{0}} \\ &= \operatorname{Tr} \left(\left(\frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \tilde{\mathbf{H}}} \right)^H \frac{\partial \tilde{\mathbf{H}}}{\partial \alpha_i} \Big|_{\boldsymbol{\alpha}=\mathbf{0}, \boldsymbol{\beta}=\mathbf{0}} \right) \\ &= \operatorname{Tr} (-\mathbf{H}^+ \mathbf{y} \mathbf{y}^H \mathbf{P}^H \mathbf{H}_i) \\ &= -\mathbf{y}^H \mathbf{P}^H \mathbf{H}_i \mathbf{H}^+ \mathbf{y}, \end{aligned}$$

where the last line follows from the cyclic property of the trace operator.

Similarly, we next consider the derivative of $f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})$ with respect to β_i , $i = 1, \dots, p$, i.e.,

$$\begin{aligned} b_i &\triangleq \frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \beta_i} \Big|_{\boldsymbol{\alpha}=\mathbf{0}, \boldsymbol{\beta}=\mathbf{0}} \\ &= \operatorname{Tr} \left(\left(\frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \tilde{\mathbf{y}}} \right)^H \frac{\partial \tilde{\mathbf{y}}}{\partial \beta_i} \Big|_{\boldsymbol{\alpha}=\mathbf{0}, \boldsymbol{\beta}=\mathbf{0}} \right) \\ &= \mathbf{y}^H \mathbf{P} \mathbf{y}_i. \end{aligned}$$

This concludes the proof of Lemma 2. \square

Now turning our attention back to (29), we denote

$$\mathbf{d} \triangleq \frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \boldsymbol{\alpha}} \Big|_{\boldsymbol{\alpha}=\mathbf{0}, \boldsymbol{\beta}=\mathbf{0}},$$

and

$$\mathbf{b} \triangleq \frac{\partial f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}})}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\alpha}=\mathbf{0}, \boldsymbol{\beta}=\mathbf{0}},$$

where we emphasize that the closed form definitions of \mathbf{d} and \mathbf{b} can be obtained from Lemma 2. We then approximate (29) and obtain the first order Taylor approximation as follows

$$f(\tilde{\mathbf{H}}, \tilde{\mathbf{y}}) \approx \kappa + \mathbf{d}^H \boldsymbol{\alpha} + \boldsymbol{\alpha}^H \mathbf{d} + \mathbf{b}^H \boldsymbol{\beta} + \boldsymbol{\beta}^H \mathbf{b}.$$

Therefore, we can approximate the regret in (26) as follows

$$\begin{aligned} \mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}) &\approx \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}}\mathbf{x}\|^2 \\ &\quad - (\kappa + \mathbf{d}^H \boldsymbol{\alpha} + \boldsymbol{\alpha}^H \mathbf{d} + \mathbf{b}^H \boldsymbol{\beta} + \boldsymbol{\beta}^H \mathbf{b}). \end{aligned} \quad (32)$$

In the following theorem, we illustrate how the optimization problem in (32) can be put in an SDP form.

Theorem 3. Let $\mathbf{H}, \mathbf{H}_1, \dots, \mathbf{H}_p \in \mathbb{C}^{m \times n}$, $\mathbf{y}, \mathbf{y}_1, \dots, \mathbf{y}_p \in \mathbb{C}^m$, $\delta_H, \delta_Y \geq 0$, $m \geq n$, where $\tilde{\mathbf{H}}$ is the full rank data matrix defined as in (27), $\tilde{\mathbf{y}}$ is the output vector defined as in (28), with the corresponding estimates \mathbf{H} and \mathbf{y} , respectively. Then the problem

$$\min_{\mathbf{x} \in \mathbb{C}^n} \max_{\|\boldsymbol{\alpha}\| \leq \delta_\alpha, \|\boldsymbol{\beta}\| \leq \delta_\beta} \mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}), \quad (33)$$

where $\mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y})$ is defined as in (32), is equivalent to solving the following SDP problem

$\min \gamma$

subject to

$$\begin{aligned} \tau_1 \geq 0, \tau_2 \geq 0, \quad \text{and} \\ \left[\begin{array}{cccc} \gamma + \kappa - \tau_1 - \tau_2 & (\mathbf{y} - \mathbf{H}\mathbf{x})^H & \delta_\alpha \mathbf{d}^H & \delta_\beta \mathbf{b}^H \\ \mathbf{y} - \mathbf{H}\mathbf{x} & \mathbf{I} & -\delta_\alpha \mathbf{G} & \delta_\beta \mathbf{Q} \\ \delta_\alpha \mathbf{d} & -\delta_\alpha \mathbf{G}^H & \tau_1 \mathbf{I} & \mathbf{0} \\ \delta_\beta \mathbf{b} & \delta_\beta \mathbf{Q}^H & \mathbf{0} & \tau_2 \mathbf{I} \end{array} \right] \geq 0, \end{aligned} \quad (34)$$

where $\mathbf{G} \triangleq [\mathbf{H}_1 \mathbf{x}, \dots, \mathbf{H}_p \mathbf{x}]$ and $\mathbf{Q} \triangleq [\mathbf{y}_1, \dots, \mathbf{y}_p]$.

Proof of Theorem 3. The proof of Theorem 3 follows similar lines to the proof of Theorem 1, hence is omitted here. \square

Remark 4. Under the framework introduced in this section, one can straightforwardly obtain the corollaries similar to Corollaries 1 and 2 by considering cases in which the uncertainty is either only on the data matrix or only on the output vector, i.e., $\delta_\beta = 0$ and $\delta_\alpha = 0$ cases, respectively. The derivations follow similar lines to Corollaries 1, 2 and Theorem 3, hence is omitted. However, similar results can be readily derived from the result in Theorem 3 with suitable changes in the SDP formulations.

Remark 5. The proofs of Theorem 2 and Theorem 3 follow from the results of Theorem 1, which relies on the lossless S-procedure. However, S-procedure is lossless with two constraints when the corresponding two quadratic (Hermitian) forms on the complex linear space [34]. However, classical S-procedure for quadratic forms is, in general, lossy with two constraints in the real case [35]. Hence, Theorem 2 and Theorem 3 cannot be extended for real linear space. On the other hand, under the frameworks described in Remark 3 and Remark 4, one can safely extend the same conclusions for the real case also, since S-procedure is lossless for quadratic forms with one constraint both in complex and real spaces [36,37].

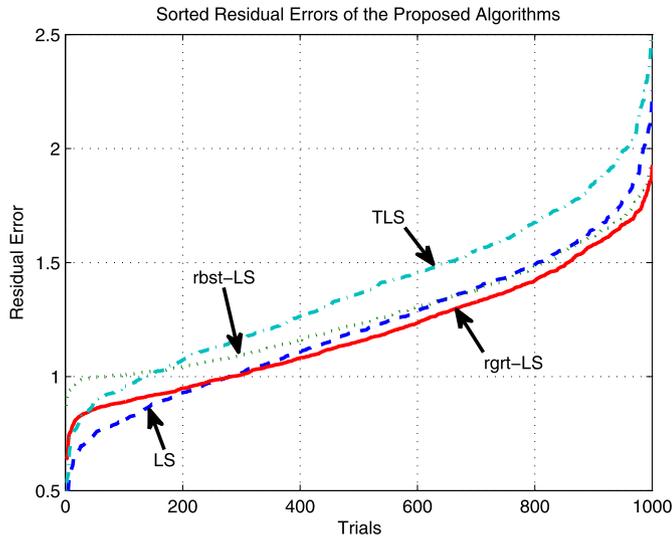


Fig. 1. Sorted residual errors for the rgrt-LS, rbst-LS, LS, and TLS estimators over 1000 trials when $\delta_H = \delta_Y = 1.2$, $m = 5$, and $n = 3$.

4. Simulations

We provide numerical examples in different scenarios in order to illustrate the merits of the proposed algorithms. In the first set of the experiments, we randomly generate a data matrix of size $m \times n$, and an output vector of size $m \times 1$, which are normalized to have unit norms. Then, we generate 1000 random perturbations $\Delta\mathbf{H}$, $\Delta\mathbf{y}$, where $\|\Delta\mathbf{H}\| \leq \delta_H$, $\|\Delta\mathbf{y}\| \leq \delta_Y$, $m = 5$, $n = 3$, and $\delta_H = \delta_Y = 1.2$. Here, we label the algorithm in Theorem 1 as “rgrt-LS”, the robust LS algorithm of [9] as “rbst-LS”, the total LS algorithm [9] as “TLS”, and finally the LS algorithm tuned to the estimates of the data matrix and the output vector as “LS”, where we directly use $\hat{\mathbf{x}} = \mathbf{H}^+ \mathbf{y}$.

For each algorithm and for each random perturbation, we find the corresponding $\hat{\mathbf{x}}$ and calculate the error $\|\mathbf{H}\hat{\mathbf{x}} - \mathbf{y}\|^2$. After we calculate the errors for each algorithm and for all random perturbations, we plot the corresponding sorted errors in ascending order in Fig. 1 for 1000 perturbations. Since the rbst-LS algorithm optimizes the worst-case residual error with respect to worst possible disturbance, it usually yields the smaller worst-case residual error among all algorithms for these simulations. On the other hand, since the LS algorithm directly uses the estimates, it usually yields the smaller residual error when the perturbations on the data matrix and the output vector are significantly small.

These results can be observed in Fig. 1, where in one extreme, the largest residual errors are observed as 2.9762 for the TLS estimator, 2.2557 for the LS estimator, 1.9275 for the rbst-LS estimator, and 1.9325 for the rgrt-LS estimator. In the other extreme, i.e., when there is almost no perturbation, the smallest estimation errors are observed as 0.3035 for the LS estimator, 0.4036 for the TLS estimator, 0.8727 for the rbst-LS estimator, and 0.6387 for the rgrt-LS estimator. While the LS estimator can be preferable when there is relatively smaller perturbations and the rbst-LS estimator can be preferable when there is significantly higher perturbations, the introduced algorithm provides a tradeoff between these algorithms and achieve a significantly smaller average error performance. The average residual error of the rgrt-LS estimator is observed as 1.1928, whereas this value is 1.2180 for the LS estimator, 1.2708 for the rbst-LS estimator, and 1.3826 for the TLS estimator. Hence, the rgrt-LS estimator is not only robust but also efficient in terms of the average error performance compared to its well-known alternatives. Owing to the competitive formulation of our estimators, we achieve such average performance gains especially when the perturbations are moderate.

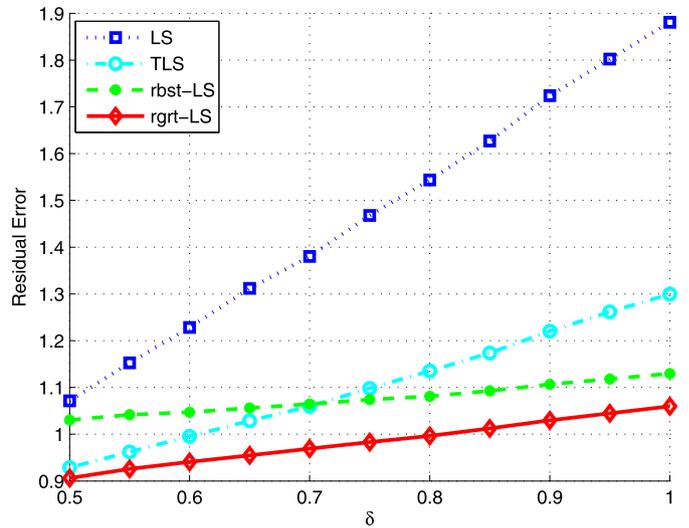


Fig. 2. Averaged residual errors for the rgrt-LS, rbst-LS, LS, and TLS estimators over 2000 trials for $m = 5$ and $n = 3$, when $\delta = \delta_H = \delta_Y \in [0.5, 1]$.

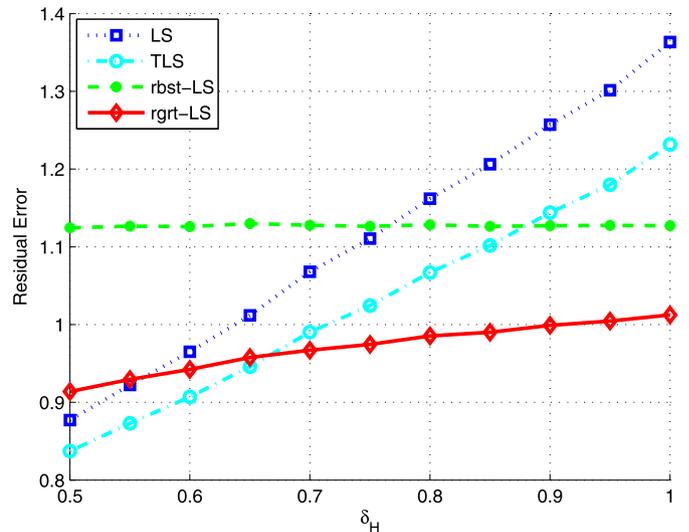


Fig. 3. Averaged residual errors for the rgrt-LS, rbst-LS, LS, and TLS estimators over 2000 trials for $m = 5$ and $n = 3$, when $\delta_H \in [0.5, 1]$ and $\delta_Y = 1$.

In the second set of experiments, we illustrate the performances of the proposed algorithms under various δ_H and δ_Y values. For these experiments, we generate 2000 random perturbations $\Delta\mathbf{H}$, $\Delta\mathbf{y}$, where $\|\Delta\mathbf{H}\| \leq \delta_H$, $\|\Delta\mathbf{y}\| \leq \delta_Y$, $m = 5$, $n = 3$ for different perturbation bounds and compute the averaged error over 2000 trials for the rgrt-LS, LS, rbst-LS, and TLS algorithms. In Fig. 2, we present the averaged residual errors of these algorithms for different values of perturbation bounds, i.e., $\delta = \delta_H = \delta_Y \in [0.5, 1]$. We observe that the proposed rgrt-LS algorithm has the best average residual error performance over different perturbation bounds compared to the LS, the rbst-LS and the TLS algorithms. Furthermore, in Fig. 3 and Fig. 4, we present the averaged residual errors of these algorithms for different perturbation bounds, i.e., when $\delta_H \neq \delta_Y$. Particularly, in Fig. 3, we set $\delta_H \in [0.5, 1]$, $\delta_Y = 1$ and in Fig. 4, we set $\delta_H = 1$, $\delta_Y \in [0.5, 1]$.

As can be observed from Fig. 2, as the perturbation bounds increase, the performances of the LS and the TLS estimators significantly deteriorate, whereas the rgrt-LS estimator provides an excellent performance. The residual error of the rbst-LS estimator, on the other hand, slightly increases as the perturbation bounds increase, i.e., it is the most robust algorithm against the perturba-

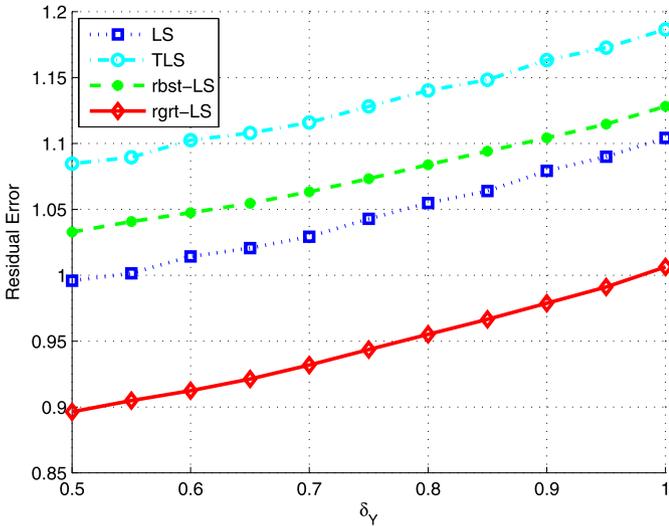


Fig. 4. Averaged residual errors for the rgrt-LS, rbst-LS, LS, and TLS estimators over 2000 trials for $m = 5$ and $n = 3$, when $\delta_H = 1$ and $\delta_\gamma \in [0.5, 1]$.

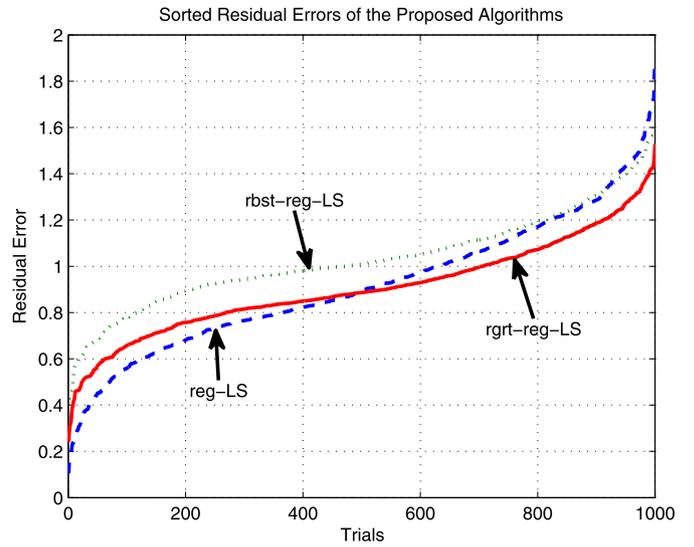


Fig. 6. Sorted residual errors for the rgrt-reg-LS, rbst-reg-LS, and LS estimators over 1000 trials when $\delta_H = \delta_\gamma = 0.65$, $\mu = 0.5$, $m = 3$, and $n = 2$.

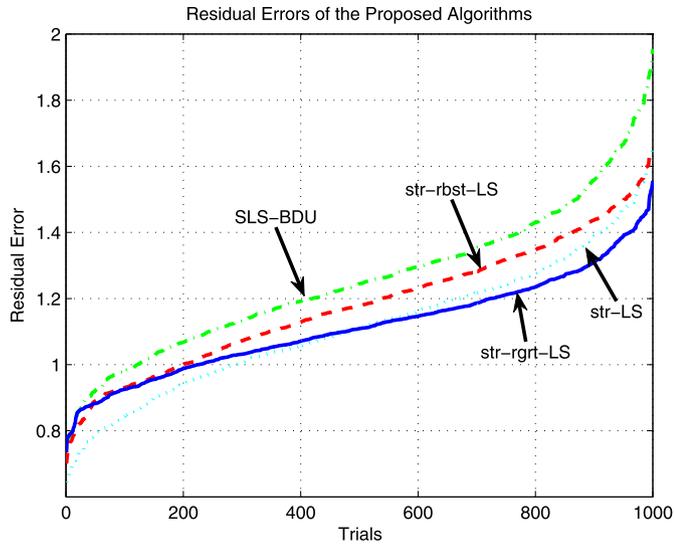


Fig. 5. Sorted residual errors for the str-rgrt-LS, str-rbst-LS, SLS-BDU, and LS estimators over 1000 trials when $\delta_H = \delta_\gamma = 0.75$, $m = 5$, and $n = 3$.

tions due to its highly conservative nature. Yet, the performance of this estimator is significantly inferior to the rgrt-LS estimator. Furthermore, the rgrt-LS estimator provides the best performance under different δ_H and δ_γ values. Particularly, in Fig. 3, we observe a similar behavior to the one in Fig. 2, where our algorithm provides a robust performance while also providing the smallest residual error (especially for high δ_H). On the other hand, in Fig. 4, we observe that the performance of rgrt-LS estimator is less sensitive to the changes in δ_γ compared to the rbst-LS, LS, and TLS estimators.

In the next experiment, we examine a system identification problem [15], which can be formulated as $\mathbf{H}_0 \mathbf{x} = \mathbf{y}_0$, where $\mathbf{H} = \mathbf{H}_0 + \mathbf{W}$ is the observed noisy Toeplitz matrix and $\mathbf{y} = \mathbf{y}_0 + \mathbf{w}$ is the observed noisy output vector. Here, the convolution matrix \mathbf{H} (which is Toeplitz) constructed from \mathbf{h} which is selected as a random sequence of ± 1 's. We then generate 1000 random structured perturbations for \mathbf{H}_0 and \mathbf{y}_0 , where $\|\alpha\| \leq 0.75\|\mathbf{H}_0\|$, and plotted the sorted estimation errors in ascending order in Fig. 5.

The average residual errors are observed as 1.1155 for the structured regret LS estimator “str-rgrt-LS” of Remark 4, 1.1807 for the structured robust LS algorithm “str-rbst-LS”, 1.1138 for the LS es-

timator, and 1.2576 for the structured least squares bounded data uncertainties estimator “SLS-BDU” of [15]. Therefore, we observe that the str-rgrt-LS algorithm yields a smaller average residual error with respect to other robust estimators and achieves the average performance of the LS estimator. In addition, we observe that the maximum residual errors are observed as 1.5554 for the str-rgrt-LS estimator, whereas it is 1.6659 for the LS estimator. Hence, the introduced algorithm can be used to obtain robustness without significant losses in the average estimation performance unlike the conventional robust estimation methods. Nevertheless, we emphasize that for a structured system, the performance of these algorithms are highly sensitive to the structures of the matrices and the vectors. If the perturbation bound is quite high, the robustness may not be preserved under large perturbations.

In the fourth experiment, i.e., in Fig. 6, we provide errors sorted in ascending order for the algorithm in Theorem 2 as “rgrt-reg-LS”, for the robust regularized LS algorithm in [16] as “rbst-reg-LS” and finally for the regularized LS algorithm as “reg-LS” [10], where the experiment setup is the same as in the first experiment except the perturbation bounds are set to 0.65 and the regularization parameter is chosen as $\mu = 0.5$. In Fig. 6, we observe that the robustness and the performance tradeoff (between the rbst-reg-LS and the reg-LS algorithms) of the introduced rgrt-reg-LS algorithm.

When there is small perturbations on the data matrix and the output vector, i.e., in the best-case scenario, the residual error of the reg-LS estimator is 0.1045, whereas it is 0.2416 for the rgrt-reg-LS estimator and 0.4282 for the rbst-reg-LS estimator. As can be observed from Fig. 6, for higher perturbations, the performance of the reg-LS estimator significantly deteriorates, whereas the rgrt-reg-LS and rbst-reg-LS algorithms provide a robust performance. On the other hand, the rgrt-reg-LS estimator significantly outperforms the rbst-reg-LS estimator in terms of the average error performance and achieves even a more desirable error performance compared to the reg-LS estimator. The average residual errors are calculated as 0.9059 for the rgrt-reg-LS estimator, 0.9177 for the reg-LS estimator, and 1.0316 for the rbst-reg-LS estimator. This experiment illustrates the sensitivity of the reg-LS estimator to the perturbations. On the other hand, the rgrt-reg-LS and rbst-reg-LS estimators provides more robust performances compared to the reg-LS estimator. Yet, the highly pessimistic nature of the rbst-reg-LS estimator deteriorates its estimation performance and yields an unacceptable performance. Our algorithm, on the other hand, not only yields a robust performance compared to the reg-LS estima-

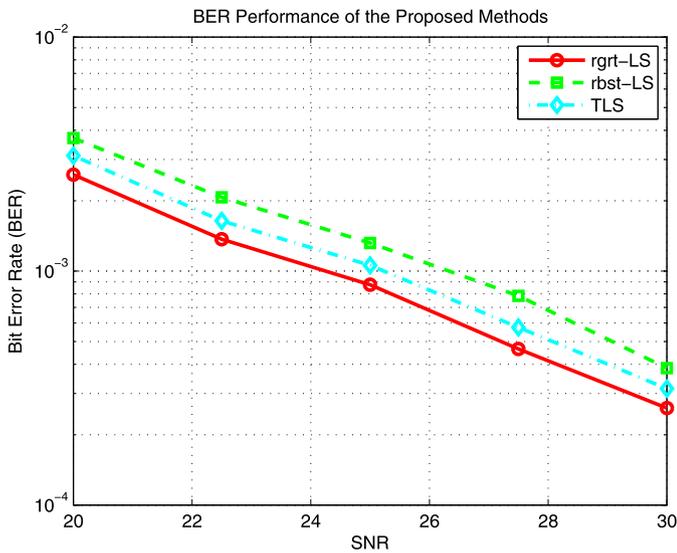


Fig. 7. BER performances of the rgrt-LS, rbst-LS, and TLS estimators (equalizers) over 1 000 000 trials under various SNRs, when $m = 3$ and $n = 2$.

tor but also does not cause any average performance degradations unlike the conventional robust estimation methods.

Finally, we illustrate the possible applications of our algorithm into different frameworks. Particularly, we consider the channel equalization problem and illustrate the bit error rate (BER) performance of our algorithm with respect to its well-known alternatives in the literature as follows.

In these simulations, we define the signal-to-noise ratio (SNR) as follows

$$\text{SNR} = 20 \log \left(\frac{\|\mathbf{x}\|}{\delta} \right),$$

where $\|\mathbf{H}\| = 1$ and $\log(\cdot)$ is the common (i.e., base 10) logarithm. For a given SNR, we generate 1 000 000 symbol vectors of \mathbf{x} (having length 2) from a binary alphabet and 1 000 000 estimates of the (MIMO) channel matrix \mathbf{H} (sized 3×2) both having unit norms, randomly. For every symbol vector and channel estimate couple, we randomly generate perturbations $\Delta\mathbf{H}$ and $\Delta\mathbf{y}$, calculate the corresponding perturbed output vector, and feed this information to the algorithms. We quantize the estimate of the symbol vector $\hat{\mathbf{x}}$ and consider the number of incorrect bits as the BER (i.e., we consider the BER rather than the symbol error rate).

In Fig. 7, we provide the BERs for various SNRs. We observe that the proposed algorithm outperforms its competitors in terms of equalization performance and successfully reconstructs the transmitted bits. While Fig. 7 illustrates the BER of the proposed algorithms averaged over a huge number of channel uses, we also illustrate the robustness of our algorithm over small number of channel uses in Fig. 8 and Fig. 9. In these experiments, we perform 100 independent trials in each of which 10 000 symbol vectors and channel matrix estimates are generated and sent over the channel as in the previous experiment for $\text{SNR} = 20$ and $\text{SNR} = 25$, respectively.

In Fig. 8 and Fig. 9, we observe that our algorithm not only provides a superior averaged performance with respect to its well-known alternatives but also provides a robust performance. The conventional robust LS estimators provide unsatisfactory results since these algorithms adapt themselves to the worst-case scenario. However, the rgrt-LS estimator has a significantly smaller BER compared to the rbst-LS and TLS estimators, since our algorithm does not tune itself to the worst possible perturbation, but considers the worst possible regret. Particularly, when the perturbation on the estimates are relatively small, our algorithm provides

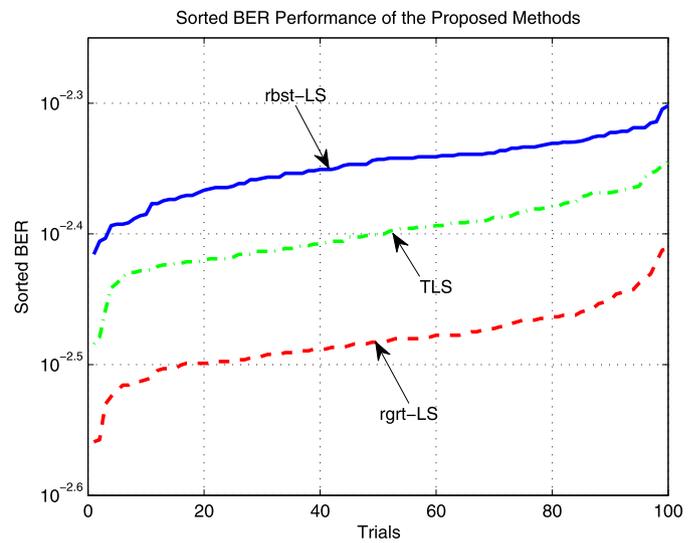


Fig. 8. Sorted BERs for the rgrt-LS, rbst-LS, and TLS estimators (equalizers) over 100 trials, where in each trial 10 000 symbol vectors are sent for $\text{SNR} = 20$, $m = 3$, and $n = 2$.

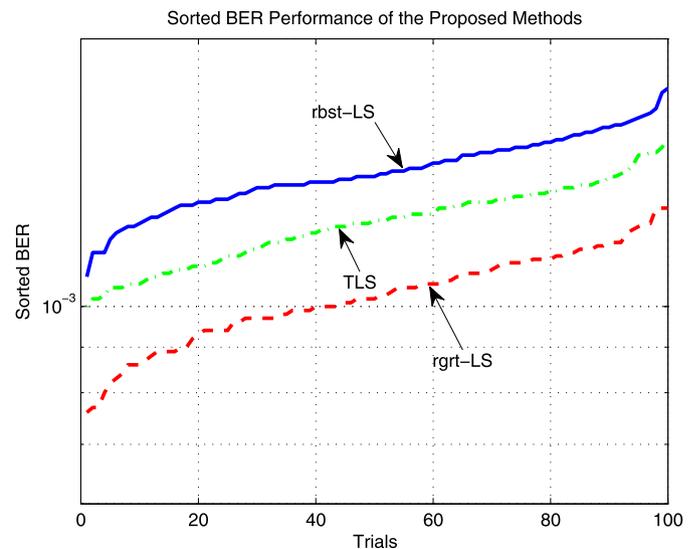


Fig. 9. Sorted BERs for the rgrt-LS, rbst-LS, and TLS estimators (equalizers) over 100 trials, where in each trial 10 000 symbol vectors are sent for $\text{SNR} = 25$, $m = 3$, and $n = 2$.

significant performance improvements compared to the conventional methods as can be seen in Fig. 8 and Fig. 9.

5. Conclusion

In this paper, we introduce a robust approach to LS estimation problems under bounded data uncertainties based on a novel regret formulation. We study the robust LS estimation problems in the presence of unstructured and structured perturbations under residual and regularized residual error criteria. In all cases, the data vectors that minimize the worst-case regrets are found by solving certain SDP problems. In our simulations, we observed that the proposed estimation methods provide an efficient tradeoff between the performance and robustness. Owing to the regret based formulation of the proposed method, we obtain significant improvements in terms of the average estimation performance with respect to the conventional robust minimax estimation methods, while maintaining the robustness as shown in our experiments.

Appendix A. Proof of Theorem 1

Before we introduce the proof of Theorem 1, we need the following proposition that follows Proposition 2 of [18].

Proposition 1. Given matrices $\mathbf{P}_1, \mathbf{Q}_1, \mathbf{P}_2, \mathbf{Q}_2, \mathbf{N}$, where \mathbf{N} is a Hermitian matrix, i.e., $\mathbf{N} = \mathbf{N}^H$,

$$\mathbf{N} \geq \mathbf{P}_1^H \mathbf{Z}_1 \mathbf{Q}_1 + \mathbf{Q}_1^H \mathbf{Z}_1^H \mathbf{P}_1 + \mathbf{P}_2^H \mathbf{Z}_2 \mathbf{Q}_2 + \mathbf{Q}_2^H \mathbf{Z}_2^H \mathbf{P}_2,$$

$\forall \mathbf{Z}_1, \mathbf{Z}_2 : \|\mathbf{Z}_1\| \leq \delta_1, \|\mathbf{Z}_2\| \leq \delta_2$, if and only if there exist $\tau_1, \tau_2 \geq 0$ such that

$$\begin{bmatrix} \mathbf{N} - \tau_1 \mathbf{Q}_1^H \mathbf{Q}_1 - \tau_2 \mathbf{Q}_2^H \mathbf{Q}_2 & -\delta_1 \mathbf{P}_1^H & -\delta_2 \mathbf{P}_2^H \\ -\delta_1 \mathbf{P}_1 & \tau_1 \mathbf{I} & \mathbf{0} \\ -\delta_2 \mathbf{P}_2 & \mathbf{0} & \tau_2 \mathbf{I} \end{bmatrix} \geq 0. \quad (\text{A.1})$$

Proof of Proposition 1. Following similar lines to [18], we first note that

$$\mathbf{N} \geq \mathbf{P}_1^H \mathbf{Z}_1 \mathbf{Q}_1 + \mathbf{Q}_1^H \mathbf{Z}_1^H \mathbf{P}_1 + \mathbf{P}_2^H \mathbf{Z}_2 \mathbf{Q}_2 + \mathbf{Q}_2^H \mathbf{Z}_2^H \mathbf{P}_2,$$

$\forall \mathbf{Z}_1, \mathbf{Z}_2 : \|\mathbf{Z}_1\| \leq \delta_1, \|\mathbf{Z}_2\| \leq \delta_2$, if and only if for every \mathbf{u} we have

$$\begin{aligned} \mathbf{u}^H \mathbf{N} \mathbf{u} &\geq \max_{\|\mathbf{Z}_1\| \leq \delta_1, \|\mathbf{Z}_2\| \leq \delta_2} \{ \mathbf{u}^H \mathbf{P}_1^H \mathbf{Z}_1 \mathbf{Q}_1 \mathbf{u} + \mathbf{u}^H \mathbf{Q}_1^H \mathbf{Z}_1^H \mathbf{P}_1 \mathbf{u} \\ &\quad + \mathbf{u}^H \mathbf{P}_2^H \mathbf{Z}_2 \mathbf{Q}_2 \mathbf{u} + \mathbf{u}^H \mathbf{Q}_2^H \mathbf{Z}_2^H \mathbf{P}_2 \mathbf{u} \} \\ &= 2\delta_1 \|\mathbf{P}_1 \mathbf{u}\| \|\mathbf{Q}_1 \mathbf{u}\| + 2\delta_2 \|\mathbf{P}_2 \mathbf{u}\| \|\mathbf{Q}_2 \mathbf{u}\|, \end{aligned} \quad (\text{A.2})$$

where the last line follows from the Cauchy–Schwartz inequality by choosing

$$\mathbf{Z}_1 = \frac{\delta_1 \mathbf{P}_1 \mathbf{u} \mathbf{u}^H \mathbf{Q}_1^H}{\|\mathbf{P}_1 \mathbf{u}\| \|\mathbf{Q}_1 \mathbf{u}\|},$$

and

$$\mathbf{Z}_2 = \frac{\delta_2 \mathbf{P}_2 \mathbf{u} \mathbf{u}^H \mathbf{Q}_2^H}{\|\mathbf{P}_2 \mathbf{u}\| \|\mathbf{Q}_2 \mathbf{u}\|}.$$

Furthermore, from the Cauchy–Schwartz inequality, (A.2) can be written as

$$\mathbf{u}^H \mathbf{N} \mathbf{u} - 2(\delta_1 \mathbf{y}_1^H \mathbf{P}_1 \mathbf{u} + \delta_2 \mathbf{y}_2^H \mathbf{P}_2 \mathbf{u}) \geq 0, \quad (\text{A.3})$$

$\forall \mathbf{u}, \mathbf{y}_1, \mathbf{y}_2 : \|\mathbf{y}_1\| \leq \|\mathbf{Q}_1 \mathbf{u}\|, \|\mathbf{y}_2\| \leq \|\mathbf{Q}_2 \mathbf{u}\|$. Note that the constraint $\|\mathbf{y}_1\| \leq \|\mathbf{Q}_1 \mathbf{u}\|$ is equivalent to

$$\mathbf{u}^H \mathbf{Q}_1^H \mathbf{Q}_1 \mathbf{u} - \mathbf{y}_1^H \mathbf{y}_1 \geq 0,$$

and similarly, $\|\mathbf{y}_2\| \leq \|\mathbf{Q}_2 \mathbf{u}\|$ is equivalent to

$$\mathbf{u}^H \mathbf{Q}_2^H \mathbf{Q}_2 \mathbf{u} - \mathbf{y}_2^H \mathbf{y}_2 \geq 0.$$

Hence, after some algebra we obtain (A.3) as follows

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}^H \underbrace{\begin{bmatrix} \mathbf{N} & -\delta_1 \mathbf{P}_1^H & -\delta_2 \mathbf{P}_2^H \\ -\delta_1 \mathbf{P}_1 & \mathbf{0} & \mathbf{0} \\ -\delta_2 \mathbf{P}_2 & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\triangleq \mathbf{F}_0} \underbrace{\begin{bmatrix} \mathbf{u} \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}}_{\triangleq \mathbf{y}} \geq 0,$$

$\forall \mathbf{y}$ such that

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}^H \underbrace{\begin{bmatrix} \mathbf{Q}_1^H \mathbf{Q}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}}_{\triangleq \mathbf{F}_1} \begin{bmatrix} \mathbf{u} \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \geq 0,$$

and

$$\begin{bmatrix} \mathbf{u} \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix}^H \underbrace{\begin{bmatrix} \mathbf{Q}_2^H \mathbf{Q}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & -\mathbf{I} \end{bmatrix}}_{\triangleq \mathbf{F}_2} \begin{bmatrix} \mathbf{u} \\ \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} \geq 0.$$

Then applying S-procedure [31], we have

$$\mathbf{y}^H \mathbf{F}_0 \mathbf{y} \geq 0,$$

$$\forall \mathbf{y}: \mathbf{y}^H \mathbf{F}_1 \mathbf{y} \geq 0, \quad \mathbf{y}^H \mathbf{F}_2 \mathbf{y} \geq 0,$$

$$\text{where } \exists \mathbf{y}_0 : \mathbf{y}_0^H \mathbf{F}_1 \mathbf{y}_0 > 0, \mathbf{y}_0^H \mathbf{F}_2 \mathbf{y}_0 > 0. \quad (\text{A.4})$$

Note that due to the structures of \mathbf{F}_1 and \mathbf{F}_2 , the regularity conditions can be easily verified. Since $\mathbf{F}_0, \mathbf{F}_1$, and \mathbf{F}_2 are Hermitian matrices, i.e., $\mathbf{F}_i = \mathbf{F}_i^H, i = 0, 1, 2$, by Theorem 1.1 in [34], (A.4) is satisfied if and only if $\exists \tau_1, \tau_2 \geq 0$ such that

$$\mathbf{F}_0 - \tau_1 \mathbf{F}_1 - \tau_2 \mathbf{F}_2 \geq 0.$$

That is

$$\begin{bmatrix} \mathbf{N} - \tau_1 \mathbf{Q}_1^H \mathbf{Q}_1 - \tau_2 \mathbf{Q}_2^H \mathbf{Q}_2 & -\delta_1 \mathbf{P}_1^H & -\delta_2 \mathbf{P}_2^H \\ -\delta_1 \mathbf{P}_1 & \tau_1 \mathbf{I} & \mathbf{0} \\ -\delta_2 \mathbf{P}_2 & \mathbf{0} & \tau_2 \mathbf{I} \end{bmatrix} \geq 0.$$

This concludes the proof of Proposition 1. \square

Now we consider the minimax problem defined in (9), and reformulate it as follows

$$\min_{\mathbf{x} \in \mathbb{C}^n} \max_{\|\Delta \mathbf{H}\| \leq \delta_H, \|\Delta \mathbf{y}\| \leq \delta_Y} \mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}) = \min_{\mathbf{x}, \gamma} \gamma,$$

subject to

$$\mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}) \leq \gamma, \quad \forall \Delta \mathbf{H}, \Delta \mathbf{y} : \|\Delta \mathbf{H}\| \leq \delta_H, \|\Delta \mathbf{y}\| \leq \delta_Y, \quad (\text{A.5})$$

where

$$\begin{aligned} \mathcal{R}(\mathbf{x}; \Delta \mathbf{H}, \Delta \mathbf{y}) &= \|\tilde{\mathbf{y}} - \tilde{\mathbf{H}} \mathbf{x}\|^2 - (\kappa + \mathbf{d}^H \Delta \mathbf{h} + \Delta \mathbf{h}^H \mathbf{d} \\ &\quad + \mathbf{b}^H \Delta \mathbf{y} + \Delta \mathbf{y}^H \mathbf{b}), \end{aligned} \quad (\text{A.6})$$

and $\kappa \triangleq f(\mathbf{H}, \mathbf{y})$. By applying the Schur complement to the constraints in (A.5), we can compactly denote (A.5) in the matrix form as follows

$$\begin{bmatrix} \gamma + \kappa + \mathbf{d}^H \Delta \mathbf{h} + \Delta \mathbf{h}^H \mathbf{d} + \mathbf{b}^H \Delta \mathbf{y} + \Delta \mathbf{y}^H \mathbf{b} & (\tilde{\mathbf{y}} - \tilde{\mathbf{H}} \mathbf{x})^H \\ \tilde{\mathbf{y}} - \tilde{\mathbf{H}} \mathbf{x} & \mathbf{I} \end{bmatrix} \geq 0, \quad (\text{A.7})$$

$\forall \Delta \mathbf{H}, \Delta \mathbf{y} : \|\Delta \mathbf{H}\| \leq \delta_H, \|\Delta \mathbf{y}\| \leq \delta_Y$. Rearranging terms in (A.7), we obtain

$$\begin{aligned} &\begin{bmatrix} \gamma + \kappa & (\mathbf{y} - \mathbf{H} \mathbf{x})^H \\ \mathbf{y} - \mathbf{H} \mathbf{x} & \mathbf{I} \end{bmatrix} \\ &\geq - \begin{bmatrix} \mathbf{d}^H \\ \mathbf{X} \end{bmatrix} \Delta \mathbf{h} \begin{bmatrix} 1 & \mathbf{0} \end{bmatrix} - \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} \Delta \mathbf{h}^H \begin{bmatrix} \mathbf{d} & \mathbf{X}^H \end{bmatrix} \\ &\quad - \begin{bmatrix} \mathbf{b}^H \\ -\mathbf{I} \end{bmatrix} \Delta \mathbf{y} \begin{bmatrix} 1 & \mathbf{0} \end{bmatrix} - \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} \Delta \mathbf{y}^H \begin{bmatrix} \mathbf{b} & -\mathbf{I} \end{bmatrix}, \end{aligned} \quad (\text{A.8})$$

$\forall \Delta \mathbf{H}, \Delta \mathbf{y} : \|\Delta \mathbf{H}\| \leq \delta_H, \|\Delta \mathbf{y}\| \leq \delta_Y$, where we used $\Delta \mathbf{H} \mathbf{x} = \mathbf{X} \Delta \mathbf{h}$, $\Delta \mathbf{h} = \text{vec}(\Delta \mathbf{H})$, and $\mathbf{X} \triangleq \mathbf{x}^H \otimes \mathbf{I}$. By applying Proposition 1 to (A.8), it follows that (9) is equivalent to

min γ

subject to

$$\begin{aligned} &\tau_1 \geq 0, \tau_2 \geq 0, \quad \text{and} \\ &\begin{bmatrix} \gamma + \kappa - \tau_1 - \tau_2 & (\mathbf{y} - \mathbf{H} \mathbf{x})^H & \delta_Y \mathbf{b}^H & \delta_H \mathbf{d}^H \\ \mathbf{y} - \mathbf{H} \mathbf{x} & \mathbf{I} & -\delta_Y \mathbf{I} & \delta_H \mathbf{X} \\ \delta_Y \mathbf{b} & -\delta_Y \mathbf{I} & \tau_1 \mathbf{I} & \mathbf{0} \\ \delta_H \mathbf{d} & \delta_H \mathbf{X}^H & \mathbf{0} & \tau_2 \mathbf{I} \end{bmatrix} \geq 0, \end{aligned}$$

hence the desired result. Therefore, this concludes the proof of Theorem 1. \square

References

- [1] E.E. Kuruoglu, Nonlinear least lp-norm filters for nonlinear autoregressive α -stable processes, *Digit. Signal Process.* 12 (1) (2002) 119–142.
- [2] M.A. Donmez, H.A. Inan, S.S. Kozat, Robust estimation in flat fading channels under bounded channel uncertainties, *Digit. Signal Process.* 23 (5) (2013) 1592–1601.
- [3] B. Zhang, S. Makram-Ebeid, R. Prevost, G. Pizaine, Fast solver for some computational imaging problems: a regularized weighted least-squares approach, *Digit. Signal Process.* 27 (2014) 107–118.
- [4] R. Caballero-Aguila, A. Hermoso-Carazo, J. Linares-Perez, Least-squares linear estimators using measurements transmitted by different sensors with packet dropouts, *Digit. Signal Process.* 22 (6) (2012) 1118–1125.
- [5] J. Zhang, Y. Pang, Pipelined robust M-estimate adaptive second-order Volterra filter against impulsive noise, *Digit. Signal Process.* 26 (2014) 71–80.
- [6] M.B. Malik, M. Salman, State-space least mean square, *Digit. Signal Process.* 18 (3) (2008) 334–345.
- [7] E.E. Kuruoglu, P.J.W. Rayner, W.J. Fitzgerald, Impulsive noise elimination using polynomial iteratively reweighted least squares, in: *IEEE Digital Signal Processing Workshop Proceedings*, 1996, pp. 347–350.
- [8] N. Kalantarova, M.A. Donmez, S.S. Kozat, Competitive least squares problem with bounded data uncertainties, in: *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP*, 2012, pp. 3841–3844.
- [9] L.E. Ghaoui, H. Le Bret, Robust solutions to least-squares problems with uncertain data, *SIAM J. Matrix Anal. Appl.* 18 (4) (1997) 1035–1064.
- [10] T. Kailath, A.H. Sayed, B. Hassibi, *Linear Estimation*, Prentice–Hall, 2000.
- [11] S. Kay, Y.C. Eldar, Rethinking biased estimation, *IEEE Signal Process. Mag.* (2008) 133–136.
- [12] A.H. Sayed, S. Chandrasekaran, Parameter estimation with multiple sources and levels of uncertainties, *IEEE Trans. Signal Process.* 48 (3) (2000) 680–692.
- [13] A.H. Sayed, S. Chandrasekaran, Estimation in the presence of multiple sources and levels of uncertainties with applications, in: *Proc. Asimolar Conference on Signals, Systems, and Computers*, vol. 2, 1998, pp. 1811–1815.
- [14] A.H. Sayed, V.H. Nascimento, S. Chandrasekaran, Estimation and control with bounded data uncertainties, *Linear Algebra Appl.* 284 (1998) 259–306.
- [15] M. Pilanci, O. Arikan, M. Pinar, Structured least squares problems and robust estimators, *IEEE Trans. Signal Process.* 58 (5) (2010) 2453–2465.
- [16] A.H. Sayed, V.H. Nascimento, F.A.M. Cipparrone, Structured least squares problems and robust estimators, *SIAM J. Matrix Anal. Appl.* 23 (4) (2002) 1120–1142.
- [17] A.H. Sayed, A framework for state-space estimation with uncertain models, *IEEE Trans. Autom. Control* 46 (7) (2001) 998–1013.
- [18] Y.C. Eldar, N. Merhav, A competitive minimax approach to robust estimation of random parameters, *IEEE Trans. Signal Process.* 52 (7) (2004) 1931–1946.
- [19] Y.C. Eldar, A. Ben-Tal, A. Nemirovski, Robust mean-squared error estimation in the presence of model uncertainties, *IEEE Trans. Signal Process.* 53 (1) (2005) 168–181.
- [20] J. Zhang, L. Peng, X. Zhao, E.E. Kuruoglu, Robust data clustering by learning multi-metric lq-norm distances, *Expert Syst. Appl.* 39 (1) (2012) 335–349.
- [21] K.E. Schubert, A new look at robust estimation and identification, Ph.D. Dissertation, Univ. of California, Santa Barbara, CA, September 2003.
- [22] P.J. Huber, Robust estimation of a local parameter, *Ann. Math. Stat.* 35 (1964) 73–101.
- [23] S. Chandrasekaran, G.H. Golub, M. Gu, A.H. Sayed, Parameter estimation in the presence of bounded data uncertainties, *SIAM J. Matrix Anal. Appl.* 19 (1) (1998) 235–252.
- [24] S. Chandrasekaran, G.H. Golub, M. Gu, A.H. Sayed, An efficient algorithm for a bounded errors-in-variables model, *SIAM J. Matrix Anal. Appl.* 20 (4) (1998) 839–859.
- [25] S. Chandrasekaran, G.H. Golub, M. Gu, A.H. Sayed, Parameter estimation in the presence of bounded modeling errors, *IEEE Signal Process. Lett.* 4 (1997) 195–197.
- [26] S. Chandrasekaran, G.H. Golub, M. Gu, A.H. Sayed, Worst-case parameter estimation with bounded model uncertainties, in: *Proc. American Control Conference*, 1997, pp. 171–175.
- [27] S. Chandrasekaran, G.H. Golub, M. Gu, A.H. Sayed, Best-fit parameter estimation for a bounded errors-in-variables model, in: *Proc. American Control Conference*, 1997, pp. 166–170.
- [28] S.S. Kozat, A.T. Erdogan, Competitive linear estimation under model uncertainties, *IEEE Trans. Signal Process.* 58 (4) (2010) 2388–2393.
- [29] Y.C. Eldar, A. Ben-Tal, A. Nemirovski, Linear minimax regret estimation of deterministic parameters with bounded data uncertainties, *IEEE Trans. Signal Process.* 52 (8) (2004) 2177–2188.
- [30] Y.C. Eldar, N. Merhav, Minimax MSE-ratio estimation with signal covariance uncertainties, *IEEE Trans. Signal Process.* 53 (4) (2005) 1335–1347.
- [31] S. Boyd, L.E. Ghaoui, E. Feron, V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, Studies in Applied Mathematics, 1994.
- [32] A.H. Sayed, V.H. Nascimento, F.A.M. Cipparrone, A regularized robust design criterion for uncertain data, *SIAM J. Matrix Anal. Appl.* 23 (4) (2001) 1120–1142.
- [33] A. Graham, *Kronecker Products and Matrix Calculus: With Applications*, John Wiley & Sons, 1981.
- [34] A.L. Fradkov, V.A. Yakubovich, The S-procedure and duality relations in nonconvex problems of quadratic programming, *Vestn. Leningr. Univ., Math.* 1 (1973) 101–109.
- [35] A.L. Fradkov, Duality theorems for certain nonconvex extremal problems, *Sib. Math. J.* 14 (2) (1973) 247–264.
- [36] Y. Huang, D.P. Palomar, Rank-constrained separable semidefinite programming with applications to optimal beamforming, *IEEE Trans. Signal Process.* 58 (2) (2010) 664–678.
- [37] Y. Huang, S. Zhang, Complex matrix decomposition and quadratic programming, *Math. Oper. Res.* 32 (3) (2007) 758–768.

N. Denizcan Vanli was born in Nigde, Turkey, in 1990. He received the B.S. degree with high honors in electrical and electronics engineering from Bilkent University, Ankara, Turkey, in 2013.

He is currently working toward the M.S. degree in the Department of Electrical and Electronics Engineering at Bilkent University. His research interests include sequential learning, adaptive filtering, machine learning, and statistical signal processing.

Mehmet A. Donmez received the B.S. degrees from both the Department of Electrical and Electronics Engineering and the Department of Mathematics, Bogazici University, Turkey, with honors and M.S. degree from the Department of Electrical and Electronics Engineering, Koc University, Turkey.

He is currently a Ph.D. candidate at the Electrical and Computer Engineering Department, University of Illinois at Urbana-Champaign, under the supervision of Prof. Andrew C. Singer. He is working as part of the Systems on Nanoscale Information fabriCs (SONIC) Center, which is a multi-university research center focusing on the design of robust, energy efficient, and intelligent computing platforms using emerging nanoscale devices that are inspired by the information processing principles found in communication and biological systems. His current research interests include adaptive signal processing, intelligent systems, online machine learning, signal processing for communications, and machine learning for signal processing. He has served as a Reviewer for the IEEE Transactions on Signal Processing and Digital Signal Processing.

Suleyman Serdar Kozat received the B.S. degree with full scholarship and high honors from Bilkent University, Turkey. He received the M.S. and Ph.D. degrees in electrical and computer engineering from University of Illinois at Urbana Champaign, Urbana, IL. Dr. Kozat is a graduate of Ankara Fen Lisesi.

After graduation, Dr. Kozat joined IBM Research, T.J. Watson Research Lab, Yorktown, New York, US as a Research Staff Member in the Pervasive Speech Technologies Group. While doing his Ph.D., he was also working as a Research Associate at Microsoft Research, Redmond, Washington, US in the Cryptography and Anti-Piracy Group. He holds several patent inventions due to his research accomplishments at IBM Research and Microsoft Research. After serving as an Assistant Professor at Koc University, Dr. Kozat is currently an Assistant Professor (with the Associate Professor degree) at the electrical and electronics department of Bilkent University.

Dr. Kozat is a Senior Member of the IEEE and the President of the IEEE Signal Processing Society, Turkey Chapter. He has been elected to the IEEE Signal Processing Theory and Methods Technical Committee and IEEE Machine Learning for Signal Processing Technical Committee, 2013. He has been awarded IBM Faculty Award by IBM Research in 2011, Outstanding Faculty Award by Koc University in 2011 (granted the first time in 16 years), Outstanding Young Researcher Award by the Turkish National Academy of Sciences in 2010, ODTU Prof. Dr. Mustafa N. Parlar Research Encouragement Award in 2011, Outstanding Faculty Award by Bilim Kahramanlari, 2013 and holds Career Award by the Scientific Research Council of Turkey, 2009.