COMPETITIVE LEAST SQUARES PROBLEM WITH BOUNDED DATA UNCERTAINTIES

Nargiz Kalantarova, Mehmet A. Donmez, Suleyman S. Kozat, Senior Member, IEEE

Koc University, Istanbul, Turkey
Email: {nkalantarova,medonmez,skozat}@ku.edu.tr

ABSTRACT

We study robust least squares problem with bounded data uncertainties in a competitive algorithm framework. We propose a competitive least squares (LS) approach that minimizes the worst case "regret" which is the difference between the squared data error and the smallest attainable squared data error of an LS estimator. We illustrate that the robust least squares problem can be put in an SDP form for both structured and unstructured data matrices and uncertainties. Through numerical examples we demonstrate the potential merit of the proposed approaches.

***Index Terms*— Robust least squares, deterministic, min-max, regret.**

I. INTRODUCTION

We investigate robust least squares (LS) problem with bounded data uncertainties. We consider least-squares problem, where we estimate a deterministic signal observed through a data matrix. However, the observed data matrix and the output vector are not exactly known, but an estimate and an uncertainty bound are provided for both the data matrix and the output vector. Since the observed data matrix and the output vector are not exactly known, one may not obtain the optimal solution with the standard LS method. In order to find a robust solution for the transmitted signal, one may optimize the worst case data error as in [1]. However, the min-max approach studied in [1] may yield overly conservative solutions since the cost function is minimized for the worst case perturbations. In order to counterbalance the conservative character of [1] we propose a competitive LS approach minimizing the worst case "regret" which is the difference between the squared data error and the smallest attainable squared data error with an LS estimator.

Research related to the least squares problem has been performed extensively in the signal processing literature [2], [3]. In many applications, the observed output vector and the data matrix in a least squares problem may not match to the "true" data matrix and output vector. These parameters may be subject to errors due to high energy noise in measurements or estimation procedure in which incorrect model assumptions may be are made. One appealing approach to find robust solutions is the robust LS method [1], where the uncertainties in the data matrix and the output vector are incorporated into optimization via min-max formulation approach. Furthermore, in many linear regression problems, the data matrix has a special structure, e.g., Toeplitz. Integrating this prior knowledge in the problem formulation improves the performance of LS approaches [1]. However, we emphasize that the robust LS approach is a pessimistic approach since the data error is minimized for the worst perturbations under uncertainty bounds. To alleviate the pessimistic nature of the robust LS approach, we use the min-max regret approach [4], [5].

Here, we consider a competitive approach to estimate the input signal where the coefficient matrix is subject to deterministic perturbations and seek a linear estimator whose performance is as close as possible to that of the optimal estimator for all possible values of the perturbations on the coefficient matrix and output vector. We emphasize that the competitive method studied in this paper significantly differs from [1], [4], [5]. Note that, the cost function studied here is different than [1], where the regret term is appended in the cost function, and the solutions for the competitive LS problem are provided in the SDP form for both unstructured perturbations and structured perturbations. Although in [4], [5] a similar regret notion is used, the cost function as well as the constraints are substantially different here.

The paper is organized as follows. In Section II, we provide the problem framework. In section III and IV, we introduce the proposed unstructured and structured competitive LS approaches and provide the SDP formulations. In section V, the numerical examples are given and the paper concludes in section VI.

II. SYSTEM DESCRIPTION

Consider the problem of estimating a deterministic vector $\mathbf{x} \in \mathbb{R}^n$ which is observed through a deterministic data matrix. However, the actual coefficient matrix and the output vector are not known but their estimates $\mathbf{A} \in \mathbb{R}^{mxn}$ (where $m \geq n$) and $\mathbf{y} \in \mathbb{R}^m$ and uncertainty bound on the estimates are provided. Here, we assume that $\mathbf{A}$ is of full rank. Our aim is to solve the least squares problem $\mathbf{Ax} \approx \mathbf{y}$, such that $(\mathbf{A} + \Delta\mathbf{A})\mathbf{x} = \mathbf{y} + \Delta\mathbf{y}$ for some deterministic perturbations $\Delta\mathbf{A} \in \mathbb{R}^{mxn}$, $\Delta\mathbf{y} \in \mathbb{R}^m$ which are unknown but bounds on each of the perturbation are provided, i.e., $\|\Delta\mathbf{A}\| \leq \rho_A$ and

978-1-4673-0046-9/12/$26.00 ©2012 IEEE

$\|\Delta\mathbf{y}\| \le \rho_Y$, where $\rho_A, \rho_Y \ge 0$. [1]

In order to estimate $\mathbf{x}$ one may substitute the estimates $\mathbf{A}$ and $\mathbf{y}$ into the LS estimator and obtain the following LS solution

$$\hat{\mathbf{x}} = (\mathbf{I} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T)\mathbf{y}. \quad (1)$$

However, this approach yields inferior results when the errors in the estimates are relatively high [1], [4], [5]. Hence, in order to find a robust solution one may use the worst case residual approach [1]

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \max_{\|\Delta\mathbf{A}\|\le\rho_A, \|\Delta\mathbf{y}\|\le\rho_Y,} \|(\mathbf{A}+\Delta\mathbf{A})\mathbf{x} - (\mathbf{y}+\Delta\mathbf{y})\|^2, \quad (2)$$

where $\mathbf{x}$ is found by minimizing the worst case error under the uncertainty bounds. Since in this min-max approach, the worst case residual is minimized, the solution may be highly conservative [1], [4], [5]. In order to compensate the conservative nature of this solution and to preserve robustness, we introduce a competitive LS approach which provides a trade off between the performance and robustness [4], [5]. We define the regret for not using the optimal LS method as the difference between the squared data error with an estimate of the input vector and the squared data error with the LS estimator as

$$\begin{aligned} r(\mathbf{A},\mathbf{y}) &\triangleq \|(\mathbf{A}+\Delta\mathbf{A})\mathbf{x} - (\mathbf{y}+\Delta\mathbf{y})\|^2 \\ &\quad - \min_{\mathbf{v}} \|(\mathbf{A}+\Delta\mathbf{A})\mathbf{v} - (\mathbf{y}+\Delta\mathbf{y})\|^2, \end{aligned}$$

where $\|\Delta\mathbf{A}\| \le \rho_A$, and $\|\Delta\mathbf{y}\| \le \rho_Y$, $\rho_A, \rho_Y \ge 0$. In the next section, the formulation of our approach is provided.

III. UNSTRUCTURED ROBUST LEAST SQUARES

In this section, we consider a competitive LS approach in a certain min-max framework. Given $\mathbf{A} \in \mathbb{R}^{mxn}$, $m \ge n$, $\mathbf{y} \in \mathbb{R}^m$, and $\rho_A, \rho_Y \in \mathbb{R}^+$, we seek to find $\mathbf{x}$ that solves

$$\begin{aligned} &\min_{\mathbf{x}} \max_{\|\Delta\mathbf{A}\|\le\rho_A, \|\Delta\mathbf{y}\|\le\rho_Y} \big\{\|(\mathbf{A}+\Delta\mathbf{A})\mathbf{x} - (\mathbf{y}+\Delta\mathbf{y})\|^2 \\ &- \min_{\mathbf{v}} \|(\mathbf{A}+\Delta\mathbf{A})\mathbf{v} - (\mathbf{y}+\Delta\mathbf{y})\|^2\big\}. \end{aligned} \quad (3)$$

Defining $\tilde{\mathbf{A}} = \mathbf{A}+\Delta\mathbf{A}$, $\tilde{\mathbf{y}} = \mathbf{y}+\Delta\mathbf{y}$ and inserting the LS solution in (3) yields

$$\min_{\mathbf{x}} \max_{\|\Delta\mathbf{A}\|\le\rho_A, \|\Delta\mathbf{y}\|\le\rho_Y} \big\{\|\tilde{\mathbf{A}}\mathbf{x} - \tilde{\mathbf{y}}\|^2 - \|(\mathbf{I} - \tilde{\mathbf{A}}\tilde{\mathbf{A}}^+)\tilde{\mathbf{y}}\|^2\big\}, \quad (4)$$

where $\tilde{\mathbf{A}}^+ \triangleq (\tilde{\mathbf{A}}^T\tilde{\mathbf{A}})^{-1}\tilde{\mathbf{A}}^T$ is the pseudo inverse of $\tilde{\mathbf{A}}$ and $\mathbf{P}_{\alpha}^{\perp} \triangleq (\mathbf{I} - \tilde{\mathbf{A}}\tilde{\mathbf{A}}^+)$ is the projection matrix of the space perpendicular to the range space of $\tilde{\mathbf{A}}$. Here, we assume that $\tilde{\mathbf{A}}$ is of full rank. If we define $f(\tilde{\mathbf{A}},\tilde{\mathbf{y}}) \triangleq \|(\mathbf{I} - \tilde{\mathbf{A}}\tilde{\mathbf{A}}^+)\tilde{\mathbf{y}}\|^2$, and use the first order Taylor series approximation [6] for $f(\tilde{\mathbf{A}},\tilde{\mathbf{y}})$, then we get

$$\begin{aligned} f(\tilde{\mathbf{A}},\tilde{\mathbf{y}}) &= f(\mathbf{A},\mathbf{y}) + 2\mathrm{Tr}\left\{\nabla f(\tilde{\mathbf{A}},\tilde{\mathbf{y}})|^T_{\tilde{\mathbf{A}}=\mathbf{A},\tilde{\mathbf{y}}=\mathbf{y}} \begin{bmatrix}\Delta\mathbf{A} & \Delta\mathbf{y}\end{bmatrix}\right\} \\ &\quad + O\left(\|\begin{bmatrix}\Delta\mathbf{A} & \Delta\mathbf{y}\end{bmatrix}\|^2\right). \end{aligned} \quad (5)$$

Based on this approximation, the regret can be written as

$$\begin{aligned} r(\mathbf{A},\mathbf{y}) &\approx \|(\mathbf{A}+\Delta\mathbf{A})\mathbf{x} - (\mathbf{y}+\Delta\mathbf{y})\|^2 - f(\mathbf{A},\mathbf{y}) \\ &\quad - 2\mathrm{Tr}\left\{\nabla f(\tilde{\mathbf{A}},\tilde{\mathbf{y}})|^T_{\tilde{\mathbf{A}}=\mathbf{A},\tilde{\mathbf{y}}=\mathbf{y}} \begin{bmatrix}\Delta\mathbf{A} & \Delta\mathbf{y}\end{bmatrix}\right\}. \end{aligned}$$

To calculate $\nabla f(\tilde{\mathbf{A}},\tilde{\mathbf{y}})$ we introduce the following lemma.

Lemma 1: Let $\mathbf{A} \in \mathbb{R}^{mxn}$, $\mathbf{y} \in \mathbb{R}^n$ and define $f(\mathbf{A},\mathbf{y}) \triangleq \mathbf{y}^T(\mathbf{I} - \mathbf{A}\mathbf{A}^+)\mathbf{y}$, then

$$\frac{\partial f(\mathbf{A},\mathbf{y})}{\partial\mathbf{A}} = -2\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}\mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1} - 2\mathbf{y}\mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}. \quad (6)$$

Proof: Based on [6], we take the partial derivative of $f(\mathbf{A},\mathbf{y})$ with respect to $\mathbf{A}_{kl}$.

$$\begin{aligned} \frac{\partial f(\mathbf{A},\mathbf{y})}{\partial\mathbf{A}_{kl}} &= -\mathbf{y}^T\big[2\mathbf{e}_k\mathbf{e}_l^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \\ &- 2\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{e}_k\mathbf{e}_l^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T + \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{e}_l\mathbf{e}_k^T\big]\mathbf{y} \\ &= \mathbf{e}_l^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}\mathbf{y}^T\mathbf{e}_k + \mathbf{e}_k^T\mathbf{y}\mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{e}_l \\ &- 2\mathbf{e}_l^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}\mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{e}_k. \end{aligned} \quad (7)$$

Since the transpose of the term in (7) is by definition the kl-th entry of the matrix
$-2\left[\mathbf{y}\mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1} - \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}\mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\right]$
the result in (6) follows. □
Also, from [6]

$$\frac{\partial f(\mathbf{A},\mathbf{y})}{\partial\mathbf{y}} = 2(\mathbf{I} - \mathbf{A}\mathbf{A}^+)\mathbf{y} = 2\mathbf{P}_{\alpha}^{\perp}\mathbf{y}. \quad (8)$$

By using Lemma 1 and (8) in (5), we get

$$\begin{aligned} f(\tilde{\mathbf{A}},\tilde{\mathbf{y}}) &\approx \eta - 2(\mathbf{y}^T\Delta\mathbf{A}\mathbf{b} + \mathbf{y}^T\mathbf{A}\mathbf{A}^+\Delta\mathbf{A}\mathbf{b}) + 2\mathbf{y}^T\mathbf{P}_{\alpha}^{\perp}\Delta\mathbf{y} \\ &= \eta - 2[\mathbf{y}^T\mathbf{B}\mathbf{a} + \mathbf{y}^T\mathbf{A}^+\mathbf{B}\mathbf{a} - \mathbf{y}^T\mathbf{P}_{\alpha}^{\perp}\Delta\mathbf{y}] \quad (9) \\ &= \eta + \mathbf{c}^T\mathbf{a} + \mathbf{a}^T\mathbf{c} + \Delta\mathbf{y}^T\mathbf{P}_{\alpha}^{\perp}\mathbf{y} + \mathbf{y}^T\mathbf{P}_{\alpha}^{\perp}\Delta\mathbf{y}, \quad (10) \end{aligned}$$

where $\eta \triangleq \mathbf{y}^T(\mathbf{I} - \mathbf{A}\mathbf{A}^+)\mathbf{y}$, $\mathbf{b} = \mathbf{A}^+\mathbf{y}$, and $\mathbf{c} = -(\mathbf{y}^T\mathbf{B} + \mathbf{y}^T\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{B})^T$. Equation (9) follows since $\Delta\mathbf{A}\mathbf{b} = \mathbf{B}\mathbf{a}$, where, $\mathbf{a} = \mathrm{vec}\left(\Delta\mathbf{A}^T\right)$ and $\mathbf{B}$ is an $mxmn$ matrix constructed by $\mathbf{b}$ as

$$\mathbf{B} \triangleq \begin{bmatrix} \mathbf{b}^T & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{b}^T & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{b}^T \end{bmatrix}.$$

[1] Throughout the paper, all vectors are column vectors and represented by boldface lowercase letters. Matrices are represented by boldface uppercase letters. Given a vector $\mathbf{x}$, $\|\mathbf{x}\| = \sqrt{\mathbf{x}^T\mathbf{x}}$ is the l_2-norm, $\mathbf{x}^T$ is the transpose. For a matrix $\mathbf{A}$, $\|\mathbf{A}\|$ implies the Frobenius norm. For a square matrix $\mathbf{M}$, $\mathrm{Tr}(\mathbf{M})$ is the trace. Contingent upon the context, $\mathbf{0}$ denotes a vector or matrix with all zero elements,where the dimension can be deduced from the context. The $\mathrm{vec}(.)$ operator stacks the columns of a matrix of dimension mxn into a mnx1 by 1 column vector [6].

In the following theorem, we demonstrate how to convert the min-max regret problem (4) into an SDP formulation.

Theorem 1: Let $\mathbf{A} \in \mathbb{R}^{mxn}$, $m \geq n$, $\mathbf{y} \in \mathbb{R}^m$, and $\rho_A, \rho_Y \in \mathbb{R}^+$, then

$$\min_{\mathbf{x}} \max_{\|\Delta\mathbf{A}\|\leq\rho_A, \|\Delta\mathbf{y}\|\leq\rho_Y} \Big\{ \|(\mathbf{A}+\Delta\mathbf{A})\mathbf{x} - (\mathbf{y}+\Delta\mathbf{y})\|^2 - \eta - 2\mathrm{Tr}\{\nabla f(\tilde{\mathbf{A}}, \tilde{\mathbf{y}})|^T_{\tilde{\mathbf{A}}=\mathbf{A}, \tilde{\mathbf{y}}=\mathbf{y}} \begin{bmatrix} \Delta\mathbf{A} & \Delta\mathbf{y} \end{bmatrix} \} \Big\} \quad (11)$$

is equivalent to solving

$$\min \lambda \quad \text{subject to}$$

$$\begin{bmatrix} \lambda+\eta-\tau-\theta & (\mathbf{A}\mathbf{x}-\mathbf{y})^T & \rho_A \mathbf{c}^T & \rho_Y \mathbf{y}^T \mathbf{P}_\alpha^\perp \\ (\mathbf{A}\mathbf{x}-\mathbf{y}) & \mathbf{I} & \rho_A \mathbf{X} & -\rho_Y \mathbf{I} \\ \rho_A \mathbf{c} & \rho_A \mathbf{X}^T & \tau \mathbf{I} & \mathbf{0} \\ \rho_Y \mathbf{P}_\alpha^\perp \mathbf{y} & -\rho_Y \mathbf{I} & \mathbf{0} & \theta \mathbf{I} \end{bmatrix} \geq 0, \quad (12)$$

where $\eta \triangleq \mathbf{y}^T(\mathbf{I}-\mathbf{A}\mathbf{A}^+)\mathbf{y}$, $\tau, \theta \geq 0$ and $\mathbf{X}$ is an $mxmn$ matrix constructed by $\mathbf{x}$ as

$$\mathbf{X} \triangleq \begin{bmatrix} \mathbf{x}^T & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{x}^T & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \mathbf{x}^T \end{bmatrix}.$$

Proof: By using (10) and applying *S*-procedure [7] to (11), it follows that (11) is equivalent to

$$\min \lambda \quad \text{subject to}$$

$$\begin{bmatrix} \lambda+\eta+\Delta\mathbf{y}^T\mathbf{P}_\alpha^\perp\mathbf{y}+\mathbf{y}^T\mathbf{P}_\alpha^\perp\Delta\mathbf{y} & (\mathbf{A}\mathbf{x}-\tilde{\mathbf{y}})^T \\ (\mathbf{A}\mathbf{x}-\tilde{\mathbf{y}}) & \mathbf{I} \end{bmatrix} \geq -\begin{bmatrix} \mathbf{c}^T \\ \mathbf{X} \end{bmatrix} \mathbf{a} \begin{bmatrix} 1 & \mathbf{0} \end{bmatrix} - \begin{bmatrix} 1 \\ \mathbf{0} \end{bmatrix} \mathbf{a}^T \begin{bmatrix} \mathbf{c} & \mathbf{X}^T \end{bmatrix}, \quad (13)$$

where $\mathbf{a} = \mathrm{vec}(\Delta\mathbf{A}^T)$, and $\Delta\mathbf{A}\mathbf{x} = \mathbf{X}\mathbf{a}$. By applying Proposition 2 of [4] two times to (13) results (12). □

IV. STRUCTURED ROBUST LEAST SQUARES

In the previous section, the perturbations on the data matrix and the output vector do not have a special structure. However, in many applications the data matrix has a special structure, e.g., Toeplitz, hence the perturbations on them. A solution to competitive LS problem with this prior knowledge may improve the performance of the min-max regret approach. Therefore, in this section, we consider a special case of the problem (3). Here, the associated perturbations for $\mathbf{A}$ and $\mathbf{y}$ are structured. We define the structure on the perturbations as $\Delta\mathbf{A} = \sum_{i=1}^n \alpha_i \mathbf{A}_i$ and $\Delta\mathbf{y} = \sum_{i=1}^n \beta_i \mathbf{y}_i$, where $\mathbf{A}_i$ and $\mathbf{y}_i$ are known and but $\alpha_i, \beta_i \in \mathbb{R}$, $i = 1, \dots, n$ are not known. However, bounds on the norm of $\boldsymbol{\alpha} \triangleq \begin{bmatrix} \alpha_1 & \alpha_2 & \dots & \alpha_n \end{bmatrix}^T$ and $\boldsymbol{\beta} \triangleq \begin{bmatrix} \beta_1 & \beta_2 & \dots & \beta_n \end{bmatrix}^T$ are provided as $\|\boldsymbol{\alpha}\| \leq \rho_A$, $\|\boldsymbol{\beta}\| \leq \rho_B$, $\rho_A, \rho_Y \geq 0$. We seek to solve the following optimization problem:

$$\min_{\mathbf{x}} \max_{\|\boldsymbol{\alpha}\|\leq\rho_A, \|\boldsymbol{\beta}\|\leq\rho_B} \left[\|\mathbf{A}(\boldsymbol{\alpha})\mathbf{x} - \mathbf{y}(\boldsymbol{\beta})\|^2 - \min_{\mathbf{v}} \|\mathbf{A}(\boldsymbol{\alpha})\mathbf{v} - \mathbf{y}(\boldsymbol{\beta})\|^2 \right], \quad (14)$$

where $\mathbf{A}(\boldsymbol{\alpha}) = \mathbf{A} + \sum_{i=1}^n \alpha_i \mathbf{A}_i$, $\mathbf{y}(\boldsymbol{\beta}) = \mathbf{y} + \sum_{i=1}^n \beta_i \mathbf{y}_i$, and $\rho_A, \rho_B \geq 0$. Substituting the LS solution to (14) yields

$$\min_{\mathbf{x}} \max_{\|\boldsymbol{\alpha}\|\leq\rho_A, \|\boldsymbol{\beta}\|\leq\rho_B} \{ \|\mathbf{A}(\boldsymbol{\alpha})\mathbf{x} - \mathbf{y}(\boldsymbol{\beta})\|^2 - \|(\mathbf{I} - \mathbf{A}(\boldsymbol{\alpha})\mathbf{A}(\boldsymbol{\alpha})^+)\mathbf{y}(\boldsymbol{\beta})\|^2 \}, \quad (15)$$

where $\mathbf{A}(\boldsymbol{\alpha})^+ \triangleq (\mathbf{A}(\boldsymbol{\alpha})^T\mathbf{A}(\boldsymbol{\alpha}))^{-1}\mathbf{A}(\boldsymbol{\alpha})^T$ is the pseudo inverse of $\mathbf{A}(\boldsymbol{\alpha})$ and $\mathbf{P}_\alpha^\perp \triangleq (\mathbf{I} - \mathbf{A}(\boldsymbol{\alpha})\mathbf{A}(\boldsymbol{\alpha})^+)$ is the projection matrix of the space perpendicular to the range space of $\mathbf{A}(\boldsymbol{\alpha})$. We use the first order Taylor series approximation to express the term $\mathbf{y}(\boldsymbol{\beta})^T(\mathbf{I} - \mathbf{A}(\boldsymbol{\alpha})\mathbf{A}(\boldsymbol{\alpha})^+)\mathbf{y}(\boldsymbol{\beta})$ as

$$\mathbf{y}(\boldsymbol{\beta})^T(\mathbf{I} - \mathbf{A}(\boldsymbol{\alpha})\mathbf{A}(\boldsymbol{\alpha})^+)\mathbf{y}(\boldsymbol{\beta}) = \mathbf{y}(\mathbf{0})^T(\mathbf{I} - \mathbf{A}(0)\mathbf{A}(0)^+)\mathbf{y}(\mathbf{0}) + 2\mathrm{Tr}\{\nabla_{[\boldsymbol{\alpha}\ \boldsymbol{\beta}]}\|\mathbf{P}_\alpha^\perp\mathbf{y}\|^2\|^T_{[\boldsymbol{\alpha}\ \boldsymbol{\beta}]=\mathbf{0}}[\boldsymbol{\alpha}\ \boldsymbol{\beta}]\} + O\left(\|[\boldsymbol{\alpha}\ \boldsymbol{\beta}]\|^2\right). \quad (16)$$

If we denote the regret term as

$$r(\mathbf{A}, \mathbf{y}) = \|\mathbf{A}(\boldsymbol{\alpha})\mathbf{x} - \mathbf{y}(\boldsymbol{\beta})\|^2 - \|(\mathbf{I} - \mathbf{A}(\boldsymbol{\alpha})\mathbf{A}(\boldsymbol{\alpha})^+)\mathbf{y}(\boldsymbol{\beta})\|^2,$$

then based on (16) the regret can be written as

$$r(\mathbf{A}, \mathbf{y}) \approx \|\mathbf{A}(\boldsymbol{\alpha})\mathbf{x} - \mathbf{y}(\boldsymbol{\beta})\|^2 - \mathbf{y}(\mathbf{0})^T(\mathbf{I} - \mathbf{A}(0)\mathbf{A}(0)^+)\mathbf{y}(\mathbf{0}) - 2\mathrm{Tr}\{\nabla_{[\boldsymbol{\alpha}\ \boldsymbol{\beta}]}\|\mathbf{P}_\alpha^\perp\mathbf{y}\|^2\|^T_{[\boldsymbol{\alpha}\ \boldsymbol{\beta}]=\mathbf{0}}[\boldsymbol{\alpha}\ \boldsymbol{\beta}]\}. \quad (17)$$

In order to compute the last term in (17), we introduce the following lemma.

Lemma 2: Let $\mathbf{A}, \mathbf{A}_1, \dots, \mathbf{A}_n \in \mathbb{R}^{mxn}$, $\mathbf{y} \in \mathbb{R}^m$ and define $f(\mathbf{A}(\boldsymbol{\alpha}), \mathbf{y}) \triangleq \mathbf{y}^T(\mathbf{I} - \mathbf{A}(\boldsymbol{\alpha})\mathbf{A}(\boldsymbol{\alpha})^+)\mathbf{y}$ where $\mathbf{A}(\boldsymbol{\alpha}) = \mathbf{A} + \sum_{i=1}^n \alpha_i \mathbf{A}_i$, then $\frac{\partial f(\mathbf{A}(\boldsymbol{\alpha}), \mathbf{y})}{\partial \alpha_i} = -2\mathbf{y}^T\mathbf{P}_\alpha^\perp\mathbf{A}_i\mathbf{A}(\boldsymbol{\alpha})^+\mathbf{y}$.

Proof: Using the result of Lemma 1,

$$\frac{\partial f(\mathbf{A}(\boldsymbol{\alpha}), \mathbf{y})}{\partial \alpha_i} = \mathrm{Tr}\left\{ \frac{\partial f(\mathbf{A})}{\partial \mathbf{A}}^T \frac{\partial \mathbf{A}}{\partial \alpha_i} \right\} = -2\mathbf{y}^T\mathbf{P}_\alpha^\perp\mathbf{A}_i\mathbf{A}^+\mathbf{y},$$

where $\mathbf{P}_{\boldsymbol{\alpha}} \triangleq \mathbf{A}(\boldsymbol{\alpha})\mathbf{A}(\boldsymbol{\alpha})^+$ is the projection matrix into the range space of $\mathbf{A}(\boldsymbol{\alpha})$. □

From Lemma 2 it follows that $\frac{\partial}{\partial \alpha_i}(\mathbf{y}(\boldsymbol{\beta})^T\mathbf{P}_\alpha^\perp\mathbf{y}(\boldsymbol{\beta})) = -2\mathbf{y}(\boldsymbol{\beta})^T\mathbf{P}_\alpha^\perp\mathbf{A}_i\mathbf{A}(\boldsymbol{\alpha})^+\mathbf{y}(\boldsymbol{\beta})$. Also, $\frac{\partial}{\partial \beta_i}(\mathbf{y}(\boldsymbol{\beta})^T\mathbf{P}_\alpha^\perp\mathbf{y}(\boldsymbol{\beta})) = 2\mathbf{y}(\boldsymbol{\beta})^T\mathbf{P}_\alpha^\perp\mathbf{y}_i$. If we denote $\eta \triangleq \mathbf{y}(\mathbf{0})^T(\mathbf{I} - \mathbf{A}(\mathbf{0})\mathbf{A}(\mathbf{0})^+)\mathbf{y}(\mathbf{0})$, $\mathbf{b} \triangleq \frac{1}{2}\nabla_{\boldsymbol{\alpha}}\|\mathbf{P}_\alpha^\perp\mathbf{y}(\boldsymbol{\beta})\|^2|_{\boldsymbol{\alpha}=\mathbf{0},\boldsymbol{\beta}=\mathbf{0}}$, and $\mathbf{c} \triangleq \frac{1}{2}\nabla_{\boldsymbol{\beta}}\|\mathbf{P}_\alpha^\perp\mathbf{y}(\boldsymbol{\beta})\|^2|_{\boldsymbol{\alpha}=\mathbf{0},\boldsymbol{\beta}=\mathbf{0}}$, then (16) is equal to

$$\mathbf{y}(\boldsymbol{\beta})^T(\mathbf{I} - \mathbf{A}(\boldsymbol{\alpha})\mathbf{A}(\boldsymbol{\alpha})^+)\mathbf{y}(\boldsymbol{\beta}) = \eta + \mathbf{b}^T\boldsymbol{\alpha} + \boldsymbol{\alpha}^T\mathbf{b} + \boldsymbol{\beta}^T\mathbf{c} + \mathbf{c}^T\boldsymbol{\beta}. \quad (18)$$

In the following theorem, by using the result (18) in (15), we show that the problem (15) can be cast as an SDP problem.

Theorem 2: Let $\mathbf{A}, \mathbf{A}_1, \dots, \mathbf{A}_n \in \mathbb{R}^{mxn}$, where $m \geq n$. Also, let $\mathbf{y}, \mathbf{y}_1, \dots, \mathbf{y}_n \in \mathbb{R}^m$, and $\rho_A, \rho_B \in \mathbb{R}^+$, then

$$\min_{\mathbf{x}} \max_{\|\boldsymbol{\alpha}\|\leq\rho_A, \|\boldsymbol{\beta}\|\leq\rho_B} \{ \|\mathbf{A}(\boldsymbol{\alpha})\mathbf{x} - \mathbf{y}(\boldsymbol{\beta})\|^2 - \eta - 2\mathrm{Tr}\{\nabla_{[\boldsymbol{\alpha}\ \boldsymbol{\beta}]}\|\mathbf{P}_\alpha^\perp\mathbf{y}\|^2|^T_{[\boldsymbol{\alpha}\ \boldsymbol{\beta}]=\mathbf{0}}[\boldsymbol{\alpha}\ \boldsymbol{\beta}]\}\}, \quad (19)$$

is equivalent to solving the following SDP problem

$$\min \lambda \quad \text{subject to}$$

$$
\begin{bmatrix} \lambda+\eta-\tau-\theta & (\mathbf{Ax}-\mathbf{y})^T & \rho_A\mathbf{b}^T & \rho_B\mathbf{c}^T \\ \mathbf{Ax}-\mathbf{y} & \mathbf{I} & \rho\mathbf{M} & -\rho_B\mathbf{Q} \\ \rho_A\mathbf{b} & \rho\mathbf{M}^T & \tau\mathbf{I} & \mathbf{0} \\ \rho_B\mathbf{c} & -\rho_B\mathbf{Q}^T & \mathbf{0} & \theta\mathbf{I} \end{bmatrix} \geq 0,
$$

where $\eta \triangleq \mathbf{y}(\mathbf{0})^T(\mathbf{I} - \mathbf{A}(0)\mathbf{A}(0)^+)\mathbf{y}(\mathbf{0})$, $\boldsymbol{\alpha} \triangleq [\alpha_1 \quad \alpha_2 \quad \dots \quad \alpha_n]^T$, $\mathbf{A}(\boldsymbol{\alpha}) = \mathbf{A} + \sum_{i=1}^n \alpha_i\mathbf{A}_i$, $\boldsymbol{\beta} \triangleq [\beta_1 \quad \beta_2 \quad \dots \quad \beta_n]^T$, $\mathbf{y}(\boldsymbol{\beta}) = \mathbf{y} + \sum_{i=1}^n \beta_i\mathbf{y}_i$, $\mathbf{M} \triangleq [\mathbf{A}_1\mathbf{x} \quad \mathbf{A}_2\mathbf{x} \quad \dots \quad \mathbf{A}_n\mathbf{x}]$ and $\mathbf{Q} = [\mathbf{y}_1 \quad \dots \quad \mathbf{y}_n]$.

Proof: The proof follows similar lines to the proof of Theorem 1. □

For reference, based on the proof of Theorem 1, it can be shown that the problem of

$$
\min_{\mathbf{x}} \max_{\|\boldsymbol{\alpha}\|\leq\rho} \|\mathbf{A}(\boldsymbol{\alpha})\mathbf{x} - \mathbf{y}(\boldsymbol{\alpha})\|^2
$$

is equivalent to

$$
\min \lambda \quad \text{subject to}
$$

$$
\begin{bmatrix} \lambda-\tau & (\mathbf{Ax}-\mathbf{y})^T & \mathbf{0} \\ \mathbf{Ax}-\mathbf{y} & \mathbf{I} & \rho\mathbf{M} \\ \mathbf{0} & \rho\mathbf{M}^T & \tau\mathbf{I} \end{bmatrix} \geq 0,
$$

where $\mathbf{A}(\boldsymbol{\alpha}) = \mathbf{A} + \sum_{i=1}^n \alpha_i\mathbf{A}_i$, $\mathbf{y}(\boldsymbol{\alpha}) = \mathbf{y} + \sum_{i=1}^n \alpha_i\mathbf{y}_i$, $\rho \geq 0$, and $\mathbf{M} \triangleq [\mathbf{A}_1\mathbf{x}-\mathbf{y}_1 \quad \mathbf{A}_2\mathbf{x}-\mathbf{y}_2 \quad \dots \quad \mathbf{A}_n\mathbf{x}-\mathbf{y}_n]$.

V. NUMERICAL EXAMPLES

In this section, we demonstrate the performance of the introduced algorithms through numerical examples. In Fig. 1, we present error results for the algorithm in Theorem 1 as "*C-LS*", for robust LS algorithm [1] tuned to the worst perturbations as "*R-LS*" and finally for the least squares method tuned to the estimates $\mathbf{A} + \Delta\mathbf{A}$ and $\mathbf{y} + \Delta\mathbf{y}$ as "*LS*". We randomly generate a data matrix $\mathbf{A}$ of size mxn, an output vector $\mathbf{y}$ of size $mx1$, and normalize them, i.e., $\|\mathbf{A}\| = 1$, and $\|\mathbf{y}\| = 1$. Then, we randomly generate 200 random perturbations $\Delta\mathbf{A}$, $\Delta\mathbf{y}$, where $\|dA\| \leq \rho_A$ and $\|\Delta\mathbf{y}\| \leq \rho_Y$, where $m = 3$, $n = 2$, $\rho_A = \rho_Y = 0.4$, and plot the corresponding errors sorted in ascending order. The largest error for $\rho_A = \rho_Y = 0.4$ and given the random $\mathbf{A}$ are: 1.634 for the LS algorithm, 1.069 for the C-LS algorithm and 1.017 for the R-LS algorithm. We observe that since the R-LS algorithm optimizes the worst case squared error with respect to worst possible disturbance, it yields the smallest worst case squared error among all algorithms for these simulations. Nevertheless, due to this highly conservative nature, the overall performance of the R-LS algorithm is significantly inferior to the LS and the C-LS algorithms. Furthermore, we observe that the C-LS algorithm provides superior average performance compared to the R-LS and the LS algorithms, and superior worst case error compared to the "LS" algorithm for these simulations. From Fig. 1, we observe that the R-LS algorithm yields the smallest worst case error, since this algorithm optimizes the data error with respect to the worst perturbations. In addition, the LS algorithm yields the highest worst case error. Although the worst case error of the C-LS algorithm is larger than the worst case error of the R-LS algorithm, the C-LS approach provides a superior performance on the average with respect to both te R-LS and the LS algorithms.

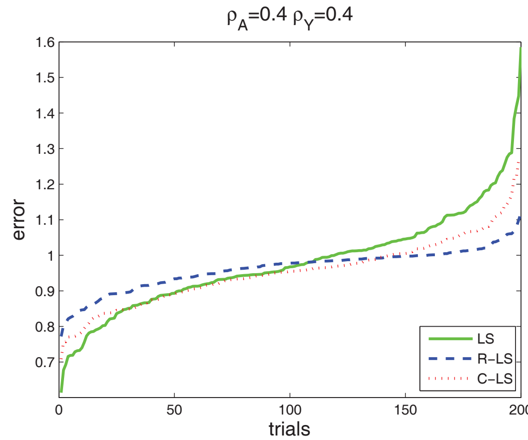

Fig. 1. Sorted erros for "R-LS", "C-LS" and "LS" algorithms over 200 trials when $\rho_A = \rho_Y = 0.4$.

VI. CONCLUSION

We introduced a competitive LS approach when the data parameters are subject to uncertainties. We investigated robust LS problem for both unstructured and structured perturbations. We demonstrated that finding the estimators that minimize the worst case regret formulations can be cast as SDP problems. Through numerical examples, we demonstrate the performance of the proposed algorithms.

VII. REFERENCES

[1] L. El Ghaoui and H. Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM Journal on Matrix Analysis and Applications*, vol. 18, no. 4, pp. 1035–1064, 1997.

[2] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares esprit algorithm," *IEEE Trans. on Signal Processing*, vol. 39, no. 5, pp. 1122–1135, 2002.

[3] K. Koh, S. Kim, S. Boyd, and Y. Lin, "An interior-point method for large-scale l1-regularized logistic regression," *Journal of Machine Learning Research*, vol. 2007, 2007.

[4] Y.C. Eldar and N. Merhav, "A competitive minimax approach to robust estimation and random parameters," *IEEE Trans. on Signal Processing*, vol. 52, pp. 1931–1946, 2004.

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

[6] A. Graham, *Kronecker Products and Matrix Calculus: with Applications*, John Wiley and Sons, 1981.

[7] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*, Studies in Applied Mathematics, 1994.