In *Recent advances in TLS Techniques and Errors-in-Variables Modeling*, edited by S. V. Huffel, SIAM, PA, pp. 171–180, 1997

Efficient Algorithms for Least Squares Type Problems with Bounded Uncertainties

Shivkumar Chandrasekaran* Gene H. Golub† Ming Gu‡ Ali H. Sayed§

Abstract

We formulate and solve new least squares type problems for parameter estimation in the presence of bounded data uncertainties. The new methods are suitable when a priori bounds on the uncertain data are available, and their solutions lead to more meaningful results especially when compared with other methods such as total least squares and robust estimation. Their superior performance is due to the fact that the new methods guarantee that the effect of the uncertainties will never be unnecessarily over-estimated, beyond what is reasonably assumed by the a priori bounds. Geometric interpretations of the solutions are provided, along with closed-form expressions for them.

1 Introduction

The central problem in estimation is to recover, to good accuracy, a set of unobservable parameters from corrupted data. Several optimization criteria have been used for estimation purposes over the years, but the most important, at least in the sense of having had the most applications, are criteria that are based on quadratic cost functions. The most striking among these is the linear least squares criterion, which was perhaps first developed by Gauss (ca. 1795) in his work on celestial mechanics. Since then, it has enjoyed widespread popularity in many diverse areas as a result of its attractive computational and statistical properties (see, e.g., [5], [9], [11], [14], [15]). Among these attractive properties, the most notable are the facts that least squares solutions can be explicitly evaluated in closed forms, they can be recursively updated as more input data is made available, and they are also maximum likelihood estimators in the presence of normally distributed measurement noise.

But alternative optimization criteria have been proposed over the years including, among others, regularized least squares [5], ridge regression [5], [11], total least squares [3], [4], [5], [8], and robust estimation [7], [10], [13], [16]. These different formulations allow, in one way or another, to incorporate further a priori information about the unknown parameter into the problem statement. They are also more effective in the presence of data

*Dept. Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106.

†Computer Science Dept., Stanford University, Palo Alto, CA 94305.

‡Department of Mathematics and Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720.

§Dept. of Electrical Engineering, University of California, Los Angeles, CA 90095. The work of A. H. Sayed was supported in part by the National Science Foundation under Award No. MIP-9409319.

errors and incomplete statistical information about the exogenous signals (or measurement errors).

Among the most notable variations is the total least squares method, also known as orthogonal regression or errors-in-variables method in statistics and system identification [12]. The method is usually more effective than standard least squares techniques in the presence of data errors. But it still exhibits certain drawbacks that degrade its performance in practical situations. In particular, it may unnecessarily over-emphasize the effect of noise and uncertainties and can, therefore, lead to overly conservative results.

More specifically, assume $A \in \mathcal{R}^{m\times n}$ is a given full rank matrix with $m \geq n$, $b \in \mathcal{R}^m$ is a given vector, and consider the problem of solving the inconsistent linear system $A \cdot \hat{x} \approx b$ in the least squares sense. The TLS solution assumes data uncertainties in A and proceeds to correct A and b by replacing them by their projections, $\hat{A}$ and $\hat{b}$, onto a specific subspace and by solving the consistent linear system of equations $\hat{A}\hat{x} = \hat{b}$. The spectral norm of the correction $(A - \hat{A})$ in the TLS solution is bounded by the smallest singular value of $\begin{bmatrix} A & b \end{bmatrix}$. While this norm might be small for vectors b that are close enough to the range space of A, it need not always be so. In other words, the TLS solution may lead to situations in which the correction term is unnecessarily large.

Consider, for example, a situation in which the uncertainties in A are very small, say A is almost known exactly. Assume further that b is far from the range space of A. In this case, it is not difficult to visualize that the TLS solution will need to rotate (A, b) into $(\hat{A}, \hat{b})$ and may therefore end up with an overly corrected approximant for A, despite the fact that A is almost exact.

These facts motivate us to introduce new parameter estimation formulations with prior bounds on the size of the allowable corrections to the data. More specifically, we formulate and solve new least squares type problems that are more suitable for parameter estimation scenarios in which a priori bounds on the uncertain data are known. The solutions lead to more meaningful results in the sense that they guarantee that the effect of the uncertainties will never be unnecessarily over-estimated, beyond what is reasonably assumed by the a priori bounds.

For brevity, we only report here the main ideas and results. The details can be found in [1], [2].

2 A New Least Squares Formulation

The first statement involves a min-max optimization problem and it leads to a regularized least squares solution with automatic selection of the regularization parameter.

2.1 Problem Statement

Let $A \in \mathcal{R}^{m\times n}$ be a given full rank matrix with $m \geq n$ and let $b \in \mathcal{R}^m$ be a given vector. The quantities (A, b) are assumed to be linearly related via an unknown vector of parameters $x \in \mathcal{R}^n$,

$$b = A \cdot x + v$$

where $v \in \mathcal{R}^m$ explains the mismatch between $A \cdot x$ and b. We assume that the "true" coefficient matrix is $A + \delta A$, and that we only know an upper bound on the perturbation δA, say $\|\delta A\|_2 \leq \eta$. Likewise, we assume that the "true" observation vector is $b + \delta b$, and that we know an upper bound η_b on the perturbation δb, say $\|\delta b\|_2 \leq \eta_b$. We pose the problem of finding an estimate that performs "well" for any allowed perturbation $(\delta A, \delta b)$.

That is, we would like to determine, if possible, an $\hat{x}$ that solves

$$\min_{\hat{x}} \left(\max_{\|\delta A\|_2 \leq \eta,\ \|\delta b\|_2 \leq \eta_b} \| (A + \delta A) \cdot \hat{x} - (b + \delta b) \|_2 \right) . \tag{1}$$

Any value that we pick for $\hat{x}$ would lead to many residual norms, $\| (A + \delta A) \cdot \hat{x} - (b + \delta b) \|_2$, one for each possible $(\delta A, \delta b)$. We want to determine the particular value(s) for $\hat{x}$ whose maximum residual is the least possible. It turns out that this problem always has a unique solution except in a special degenerate case in which the solution is nonunique.

2.2 A Geometric Formulation

The problem also admits an interesting geometric formulation. For this purpose, and for the sake of illustration, assume we have a unit-norm vector b, $\|b\|_2 = 1$, with no uncertainties in it ($\eta_b = 0$; it turns out that the solution does not depend on η_b). Assume further that A is simply a column vector, say a, with $\eta \neq 0$, and consider (1) in this setting:

$$\min_{\hat{x}} \left(\max_{\|\delta a\|_2 \leq \eta} \| (a + \delta a) \cdot \hat{x} - b \|_2 \right) . \tag{2}$$

The situation is depicted in Fig. 1. The vectors a and b are indicated in thick black lines. The vector a is shown in the horizontal direction and a circle of radius η around its vertex indicates the set of all possible vertices for $a + \delta a$. It can be verified that the solution can be obtained by drawing a perpendicular from b to the lower tangential line θ_1. The segment r_1 denotes the optimum residual. More details can be found in [1].

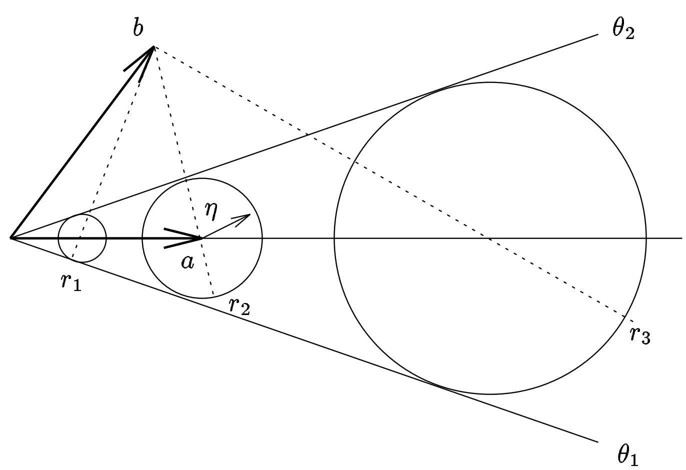

FIG. 1. Geometric construction of the solution for a simple example.

2.3 Reducing the Min-Max Problem to a Minimization Problem

The constrained min-max problem (1) can be reduced to a standard minimization problem as follows. First note that

$$\| (A + \delta A) \cdot \hat{x} - (b + \delta b) \|_2 \quad \leq \quad \|A \cdot \hat{x} - b\|_2 + \eta \cdot \|\hat{x}\|_2 \ + \ \eta_b \ ,$$

which provides an upper bound for $\| (A+\delta A)\cdot \hat{x} - (b+\delta b)\|_2$. The upper bound can in fact be achieved by choosing $(\delta A, \delta b)$ as

$$\delta A^o = \frac{(A\cdot \hat{x} - b)}{\|A\cdot \hat{x} - b\|_2} \cdot \frac{\hat{x}^T}{\|\hat{x}\|_2}\cdot \eta\ , \quad \delta b^o = -\frac{(A\cdot \hat{x} - b)}{\|A\cdot \hat{x} - b\|_2} \cdot \eta_b\ .$$

For these choices of perturbations in A and b, it follows that the vectors

$$\{(A\cdot \hat{x} - b),\ \delta A^o \cdot \hat{x},\ \delta b^o\}$$

are collinear. Hence,

$$\| (A+\delta A^o)\cdot \hat{x} - (b+\delta b^o)\|_2 \quad = \quad \|A\cdot \hat{x} - b\|_2 + \eta \cdot \|\hat{x}\|_2 + \eta_b\ ,$$

which is the desired upper bound. We therefore conclude that the constrained min-max problem (1) is equivalent to the following minimization problem. Given $A \in \mathcal{R}^{m\times n}$, with $m \geq n$, $b \in \mathcal{R}^m$, and nonnegative real numbers (η, η_b). Determine, if possible, an $\hat{x}$ that solves

(3) $$\min_{\hat{x}} \left(\|A\cdot \hat{x} - b\|_2 + \eta \cdot \|\hat{x}\|_2 + \eta_b\right)\ .$$

Note that the cost function in (3) is not of the same form as a regularized cost function. In particular, only the Euclidean norms of $(A\cdot\hat{x} - b)$ and $\hat{x}$ appear in (3) rather than the squared Euclidean norms of these quantities.

2.4 Solving the Minimization Problem

To solve (3), we define the cost function

$$\mathcal{L}(\hat{x}) = \|A\cdot \hat{x} - b\|_2 + \eta \cdot \|\hat{x}\|_2 + \eta_b\ .$$

It is easy to check that $\mathcal{L}(\hat{x})$ is a convex continuous function in $\hat{x}$ and hence any local minimum of $\mathcal{L}(\hat{x})$ is also a global minimum. But at any local minimum of $\mathcal{L}(\hat{x})$, it either holds that $\mathcal{L}(\hat{x})$ is not differentiable or its gradient $\bigtriangledown\mathcal{L}(\hat{x})$ is 0. In particular, note that $\mathcal{L}(\hat{x})$ is not differentiable only at $\hat{x} = 0$ and at any $\hat{x}$ that satisfies $A\cdot\hat{x} - b = 0$.

We first consider the case in which $\mathcal{L}(\hat{x})$ is differentiable and, hence, the gradient of $\mathcal{L}(\hat{x})$ exists and is given by

$$\bigtriangledown\mathcal{L}(\hat{x}) \quad = \quad \frac{1}{\|A\cdot \hat{x} - b\|_2}\cdot \left(\left(A^T\cdot A + \alpha I\right)\cdot \hat{x} - A^T \cdot b\right)\ ,$$

where we have introduced the positive real number

(4) $$\alpha = \frac{\eta \cdot \|A\cdot \hat{x} - b\|_2}{\|\hat{x}\|_2}\ .$$

By setting $\bigtriangledown\mathcal{L}(\hat{x}) = 0$ we obtain that any stationary solution $\hat{x}$ of $\mathcal{L}(\hat{x})$ is given by

(5) $$\hat{x} = \left(A^T\cdot A + \alpha I\right)^{-1}\cdot A^T\cdot b\ .$$

We still need to determine the parameter α that corresponds to $\hat{x}$, and which is defined in (4). To solve for α, we introduce the singular value decomposition (SVD) of A:

(6) $$A = U \cdot \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} \cdot V^T\ ,$$

where $U \in \mathcal{R}^{m\times m}$ and $V \in \mathcal{R}^{n\times n}$ are orthogonal, and $\Sigma = \text{diag}(\sigma_1, \cdots, \sigma_n)$ is diagonal, with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$ being the singular values of A. We further partition the vector $U^T b$ into

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = U^T \cdot b \tag{7}$$

where $b_1 \in \mathcal{R}^n$ and $b_2 \in \mathcal{R}^{m-n}$. In this case, it can be verified that equation (4) for α reduces to the following nonlinear equation that is only a function of α and the given data (A, b, η),

$$\alpha = \frac{\eta \cdot \sqrt{\|b_2\|_2^2 + \alpha^2 \cdot \| (\Sigma^2 + \alpha I)^{-1} \cdot b_1\|_2^2}}{\|\Sigma \cdot (\Sigma^2 + \alpha I)^{-1} \cdot b_1\|_2} . \tag{8}$$

Note that only the norm of b_2, and not b_2 itself, is needed in the above expression.

Remark. We have assumed in the derivation so far that A is full rank. If this were not the case, i.e., if A (and hence Σ) were singular, then equation (8) can be reduced to an equation of the same form but with a non-singular Σ of smaller dimension. Indeed, if we partition

$$\Sigma = \begin{bmatrix} \hat{\Sigma} & 0 \\ 0 & 0 \end{bmatrix} ,$$

where $\hat{\Sigma} \in \mathcal{R}^{k\times k}$ is non-singular, and let $\hat{b}_1 \in \mathcal{R}^k$ be the first k components of b_1; $\tilde{b}_1 \in \mathcal{R}^{n-k}$ be the last $n-k$ components of b_1; and let

$$\|\hat{b}_2\|_2^2 = \|b_2\|_2^2 + \|\tilde{b}_1\|_2^2 .$$

Then equation (8) reduces to

$$\alpha = \frac{\eta \cdot \sqrt{\|\hat{b}_2\|_2^2 + \alpha^2 \cdot \| \left(\hat{\Sigma}^2 + \alpha I\right)^{-1} \cdot \hat{b}_1\|_2^2}}{\|\hat{\Sigma} \cdot \left(\hat{\Sigma}^2 + \alpha I\right)^{-1} \cdot \hat{b}_1\|_2} , \tag{9}$$

the same form as (8). From now on, we assume that A is full rank and, hence, Σ is invertible.

2.5 The Secular Equation

Define the nonlinear function in α,

$$\mathcal{G}(\alpha) = b_1^T \cdot \left(\Sigma^2 - \eta^2 I\right) \cdot \left(\Sigma^2 + \alpha I\right)^{-2} \cdot b_1 - \frac{\eta^2}{\alpha^2} \cdot \|b_2\|_2^2 . \tag{10}$$

It follows that α is a positive solution to (8) if, and only if, it is a positive root of $\mathcal{G}(\alpha)$. Following [5], we refer to the equation

$$\mathcal{G}(\alpha) = 0 \tag{11}$$

as a *secular* equation.

The function $\mathcal{G}(\alpha)$ has several useful properties that will allow us to provide conditions for the existence of a unique positive root α. The following results are proved in [1].

LEMMA 2.1. *The function $\mathcal{G}(\alpha)$ in (10) can have at most one positive root. In addition, if $\hat{\alpha} > 0$ is a root of $\mathcal{G}(\alpha)$, then $\hat{\alpha}$ is a simple root and $\mathcal{G}'(\hat{\alpha}) > 0$.*

LEMMA 2.2. *Assume $\eta > 0$ (a standing assumption) and $b_2 \neq 0$, i.e., b does not belong to the column span of A. Then the function $\mathcal{G}(\alpha)$ in (10) has a unique positive root if, and only if,*

$$\eta < \frac{\|A^T \cdot b\|_2}{\|b\|_2} \,. \tag{12}$$

LEMMA 2.3. *Assume $\eta > 0$ (a standing assumption) and $b_2 = 0$, i.e., b lies in the column span of A. Then the function $\mathcal{G}(\alpha)$ in (10) has a positive root if, and only, if*

$$\tau_1 < \eta < \tau_2 \,, \tag{13}$$

where

$$\tau_1 = \frac{\|\Sigma^{-1} \cdot b_1\|_2}{\|\Sigma^{-2} \cdot b_1\|_2} \quad \textit{and} \quad \tau_2 = \frac{\|\Sigma \cdot b_1\|_2}{\|b_1\|_2} \,.$$

It can be further shown [1] that whenever $\mathcal{G}(\alpha)$ has a positive root $\hat{\alpha}$, the corresponding vector $\hat{x}$ in (5) must be the global minimizer of $\mathcal{L}(\hat{x})$.

LEMMA 2.4. *Let $\hat{\alpha}$ be a positive root of $\mathcal{G}(\alpha)$ and let $\hat{x}$ be defined by (5) for $\alpha = \hat{\alpha}$. Then $\hat{x}$ is the global minimum of $\mathcal{L}(\hat{x})$.*

We still need to consider the points at which $\mathcal{L}(\hat{x})$ is not differentiable. These include $\hat{x} = 0$ and any solution of $A \cdot \hat{x} = b$. We omit the details here and refer to [1].

2.6 Solution of the Constrained Min-Max Problem

We collect in the following statement the full algebraic solution to (1).

THEOREM 2.1. *Given $A \in \mathcal{R}^{m \times n}$, with $m \geq n$ and A full rank, $b \in \mathcal{R}^m$, and nonnegative real numbers (η, η_b). The optimization problem (1) always has a solution $\hat{x}$. The solution(s) can be constructed as follows:*

- *Introduce the SVD of A as in (6).*
- *Partition the vector $U^T b$ as in (7).*
- *Introduce the secular function (10).*
- *Define*

$$\tau_1 = \frac{\|\Sigma^{-1} b_1\|_2}{\|\Sigma^{-2} b_1\|_2} \quad \textit{and} \quad \tau_2 = \frac{\|A^T b\|_2}{\|b\|_2} \,.$$

First case: b does not belong to the column span of A.

1. *If $\eta \geq \tau_2$ then the unique solution is $\hat{x} = 0$.*
2. *If $\eta < \tau_2$ then the unique solution is $\hat{x} = (A^T \cdot A + \hat{\alpha} \cdot I)^{-1} A^T \cdot b$, where $\hat{\alpha}$ is the unique positive root of the secular equation $\mathcal{G}(\alpha) = 0$.*

Second case: b belongs to the column span of A.

1. *If $\eta \geq \tau_2$ then the unique solution is $\hat{x} = 0$.*

2. *If $\tau_1 < \eta < \tau_2$ then the unique solution is $\hat{x} = (A^T \cdot A + \hat{\alpha} \cdot I)^{-1} A^T \cdot b$, where $\hat{\alpha}$ is the unique positive root of the secular equation $\mathcal{G}(\alpha) = 0$.*

3. *If $\eta \leq \tau_1$ then the unique solution is $\hat{x} = V\Sigma^{-1}b_1 = A^\dagger \cdot b$.*

4. *If $\eta = \tau_1 = \tau_2$ then there are infinitely many solutions that are given by $\hat{x} = \beta \cdot V \cdot \Sigma^{-1} \cdot b_1 = \beta \cdot A^\dagger \cdot b$, for any $0 \leq \beta \leq 1$.*

2.7 The Case of Sparse Data Matrices

The above algorithm is suitable when the computation of the SVD of A is feasible. For large sparse matrices A, it is better to reformulate the secular equation as follows. Squaring both sides of (4), and after some manipulation, we are led to

$$d^T(C + \alpha \cdot I)^{-2} d = \frac{\eta^2}{\alpha^2} \left[b^T \cdot B - d^T (C + \alpha \cdot I)^{-1} d - \alpha \cdot d^T (C + \alpha \cdot I)^{-2} d \right] ,$$

where we have defined $C = A^T \cdot A$ and $d = A^T \cdot b$. Therefore, finding α reduces to finding the positive root of

$$\mathcal{H}(\alpha) \triangleq d^T(C + \alpha \cdot I)^{-2} d - \frac{\eta^2}{\alpha^2} \left[b^T \cdot B - d^T (C + \alpha \cdot I)^{-1} d - \alpha \cdot d^T (C + \alpha \cdot I)^{-2} d \right] .$$

In this form, one can develop techniques similar to those suggested in [6] to find α efficiently.

3 Another New Formulation: TLS with Bounded Uncertainties

We now introduce another optimization problem that turns out to involve the minimization of a cost function in an "indefinite" metric, in a way that is similar to more recent works on robust estimation and filtering (e.g., [7], [10], [13]). However, the cost function considered in this paper is more complex and, contrary to robust estimation where no prior bounds are imposed on the size of the disturbances, the derivation here shows how to solve the resulting optimization problem in the presence of such constraints. A closed form and explicit solution are again obtained in terms of the unique positive root of a secular equation.

3.1 Problem Statement

The formulation now involves a min-min optimization problem. Given $A \in \mathcal{R}^{m \times n}$, with $m \geq n$, $b \in \mathcal{R}^m$, and a nonnegative real number η. Determine, if possible, an $\hat{x}$ that solves

$$\min_{\hat{x}} \left(\min_{\|\delta A\|_2 \leq \eta} \| (A + \delta A) \cdot \hat{x} - b \|_2 \right) . \tag{14}$$

This problem also admits a geometric interpretation that is similar to the min-max case. The details can be found in [2].

3.2 Reducing the Min-Min Problem to a Minimization Problem

It can be shown, under an additional fundamental assumption, that the min-min problem is equivalent to a standard minimization problem. Assume that for all vectors $\hat{x}$ it holds that

$$\eta \cdot \|\hat{x}\|_2 < \|A \cdot \hat{x} - b\|_2 \qquad \text{(fundamental assumption).} \tag{15}$$

The equivalent problem is to determine, if possible, an $\hat{x}$ that solves

$$\min_{\hat{x}} \left(\|A \cdot \hat{x} - b\|_2 - \eta \cdot \|\hat{x}\|_2 \right) . \tag{16}$$

We may remark that cost functions similar to (16) but with squared distances, say

$$\min_{\hat{x}} \left(\|A \cdot \hat{x} - b\|_2^2 - \gamma \cdot \|\hat{x}\|_2^2 \right) ,$$

for some γ, often arise in the study of indefinite quadratic cost functions in robust estimation [7], [13]. The major distinction between both formulations, is that the cost function (16) involves distance terms and it will be shown to provide an automatic procedure for selecting the "regularization" factor γ (viz., the factor $\hat{\alpha}$ that is introduced later).

3.3 The Fundamental Condition for Non-Degeneracy

The fundamental condition (15) needs to be satisfied for all vectors $\hat{x}$ in order to avoid degenerate cases. This can be restated in terms of conditions on the data (A, b, η) alone.

LEMMA 3.1. *Necessary and sufficient conditions in terms of (A, b, η) for the fundamental relation (15) to hold are:*

$$(\eta^2 \cdot I - A^T \cdot A) \ < \ 0 \quad \Longleftrightarrow \quad \eta \ < \ \sigma_{min}(A) , \tag{17}$$

and

$$b^T \cdot \left[I \ - \ A \cdot (A^T \cdot A - \eta^2 \cdot I)^{-1} \cdot A^T \right] \cdot b \ > 0 . \tag{18}$$

The notation $\sigma_{min}(A)$ stands for the minimum singular value of A. Now note that for a well-defined problem of the form (14) we need to assume $\eta > 0$ which, in view of (17), means that A should be full rank so that $\sigma_{min}(A) > 0$. We therefore assume, from now on, that A is full rank.

3.4 Solving the Minimization Problem

To solve (16), we define the (non-convex) cost function

$$\mathcal{L}(\hat{x}) = \|A \cdot \hat{x} - b\|_2 - \eta \cdot \|\hat{x}\|_2 ,$$

which is continuous in $\hat{x}$ and bounded from below by zero in view of (15). A minimum point for $\mathcal{L}(\hat{x})$ can only occur at ∞, at points where $\mathcal{L}(\hat{x})$ is not differentiable, or at points where its gradient, $\bigtriangledown\mathcal{L}(\hat{x})$, is 0. In particular, note that $\mathcal{L}(\hat{x})$ is not differentiable only at $\hat{x} = 0$ and at any $\hat{x}$ that satisfies $A \cdot \hat{x} - b = 0$. But points $\hat{x}$ satisfying $A \cdot \hat{x} - b = 0$ are excluded by the fundamental condition (15). Following arguments similar to what we have done in the min-max case, we are led to the following complete characterization of the solution.

THEOREM 3.1. *Given $A \in \mathcal{R}^{m \times n}$, with $m \geq n$ and A full rank, $b \in \mathcal{R}^m$, and a nonnegative real number $\eta < \sigma_{min}(A)$. Assume further that*

$$b^T \cdot \left[I \ - \ A \cdot (A^T \cdot A - \eta^2 \cdot I)^{-1} \cdot A^T \right] \cdot b \ > 0 .$$

The optimization problem (14) has a unique solution that can be constructed as follows.

- *Introduce the SVD of A as in (6).*

- *Partition the vector $U^T b$ as in (7).*
- *Introduce the new secular function*

$$\mathcal{G}(\alpha) = b_1^T \cdot \left(\Sigma^2 - \eta^2 I\right) \cdot \left(\Sigma^2 - \alpha I\right)^{-2} \cdot b_1 - \frac{\eta^2}{\alpha^2} \cdot \|b_2\|_2^2 \,. \tag{19}$$

- *Determine the unique positive root $\hat{\alpha}$ of $\mathcal{G}(\alpha)$ that lies in the interval (η^2, σ_n^2).*
- *Then*

$$\hat{x} = (A^T \cdot A - \hat{\alpha} \cdot I)^{-1} A^T \cdot b \,.$$

4 Variations

There are several variations to the above min-max and min-min optimization problems that submit to algebraic solutions. We only mention two examples here:

Uncertain Weights. Consider the min-max problem

$$\min_{\hat{x}} \max_{\|\delta W\|_2 \leq \eta_w} \|(W + \delta W) \cdot (A \cdot \hat{x} - b)\|_2 \,.$$

It reduces to

$$\min_{\hat{x}} \left[\, \|W \cdot (A \cdot \hat{x} - b)\|_2 \; + \; \eta_w \cdot \|A \cdot \hat{x} - b\|_2 \,\right] \,,$$

and the optimal solution can be shown to satisfy

$$A^T \cdot (W^T \cdot W + \hat{\alpha} \cdot I) \cdot A \cdot \hat{x} = A^T \cdot (W^T \cdot W + \hat{\alpha} \cdot I) \cdot b \,,$$

where $\hat{\alpha}$ satisfies a secular equation similar in form to $\mathcal{G}$ in (10). The details will be provided elsewhere.

Multiplicative Uncertainties. Consider the min-max problem

$$\min_{\hat{x}} \max_{\|\delta A\|_2 \leq \eta_a} \|(I + \delta A) \cdot A \cdot \hat{x} - b\|_2.$$

It reduces to

$$\min_{\hat{x}} \left[\, \|A \cdot \hat{x} - b\|_2 \; + \; \eta_a \, \|A \cdot \hat{x}\|_2 \right] \,,$$

and the optimal solution can be shown to be a scaled version of the least squares solution, viz.,

$$(1 + \hat{\alpha}) \cdot A^T \cdot A \cdot \hat{x} = A^T \cdot b \,,$$

where $\hat{\alpha}$ is given by

$$\hat{\alpha} = \begin{cases} \eta_a \, \|P^\perp \cdot b\| \left(\frac{\eta_a \cdot \|P^\perp \cdot b\| + \|P \cdot b\| \cdot \sqrt{1 - \eta_a^2}}{\|P \cdot b\|^2 - \eta_a^2 \cdot \|b\|^2} \right) & \text{if } \; \eta_a \cdot \|b\| < \|P \cdot b\| \\ \infty & \text{otherwise} \end{cases}$$

where $P = A(A^T A)^{-1} A^T$ and $P^\perp = I - P$.

5 Concluding Remarks

Several extensions are possible. For example, if only selected columns of the A matrix are uncertain, while the remaining ones are known precisely, the problem can be reduced to the formulations (1) and (14) (as discussed in [1], [2]). Also, weighted versions with uncertainties in the weight matrices are useful in several applications, as well as cases with more general multiplicative uncertainties, and these are currently under investigation.

Several other interesting issues remain to be addressed, and will be pursued elsewhere. Among these, we state the following:

1. A study of the statistical properties of the constrained min-max and min-min problems is valuable for a better understanding of their performances in stochastic settings.

2. The numerical properties of the algorithms proposed in this paper need also be addressed.

3. Extensions of the algorithms to deal with perturbations in submatrices of A are of interest and will be studied elsewhere.

References

[1] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, *Parameter estimation in the presence of bounded data uncertainties*, SIAM J. Matrix Anal. Appl., (1996). Submitted for publication.

[2] ——, *Total least-squares with bounded uncertainties*, SIAM J. Matrix Anal. Appl., (1996). Submitted for publication.

[3] G. H. Golub, *Some modified matrix eigenvalue problems*, SIAM Review, 15 (1973), pp. 318–344.

[4] G. H. Golub and C. F. Van Loan, *An analysis of the total least squares problem*, SIAM J. Numer. Anal., 17 (1980), pp. 883–893.

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 2nd ed., MD, 1989.

[6] G. H. Golub and U. von Matt, *Generalized cross-validation for large scale problems*, in this volume, S. Van Huffel, ed., SIAM, Philadelphia, PA, 1997.

[7] B. Hassibi, A. H. Sayed, and T. Kailath, *Recursive linear estimation in Krein spaces - Part I: Theory*, IEEE Trans. Automatic Control, AC-41 (1996), pp. 18–33.

[8] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, SIAM, Philadelphia, 1991.

[9] S. M. Kay, *Fundamentals of Statistical Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[10] P. Khargonekar and K. M. Nagpal, *Filtering and smoothing in an $H^\infty-$ setting*, IEEE Trans. on Automatic Control, AC-36 (1991), pp. 151–166.

[11] C. L. Lawson and R. J. Hanson, *Solving Least-Squares Problems*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[12] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*, MIT Press, Cambridge, MA, 1983.

[13] A. H. Sayed, B. Hassibi, and T. Kailath, *Fundamental Inertia Conditions for the Minimization of Quadratic Forms in Indefinite Metric Spaces*, in Operator Theory: Advances and Applications, Birkhaüser, 1996.

[14] A. H. Sayed and T. Kailath, *A state-space approach to adaptive RLS filtering*, IEEE Signal Processing Magazine, 11 (1994), pp. 18–60.

[15] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation and Time Series Analysis*, Addison–Wesley Publishing Company, MA, 1991.

[16] U. Shaked and Y. Theodor, *$H^\infty-$optimal estimation: A tutorial*, in Proc. IEEE Conference on Decision and Control, Tucson, AZ, Dec. 1992, pp. 2278–2286.