Proceedings of the American Control Conference
Albuquerque, New Mexico June 1997
0-7803-3832-4/97/$10.00 © 1997 AACC

BEST-FIT PARAMETER ESTIMATION FOR A BOUNDED ERRORS-IN-VARIABLES MODEL*

S. CHANDRASEKARAN, GENE H. GOLUB, MING GU, AND ALI H. SAYED

Abstract

We pose and solve a parameter estimation problem in the presence of bounded data uncertainties. The new formulation involves a minimization problem and admits a closed form solution in terms of the unique positive root of a secular equation. It also has interesting connections with errors-in-variables and H_∞ methods.

1. INTRODUCTION

Parameter estimation in the presence of data uncertainties is a problem of considerable practical importance. Many estimators have been proposed in the literature with the intent of handling modeling errors and measurement noise. Among the most notable is the total least-squares method [1, 2], also known as orthogonal regression or errors-in-variables method in statistics and system identification [3]. In contrast to the standard least-squares problem, the TLS formulation allows for errors in the data matrix. But it still exhibits certain drawbacks that degrade its performance in some practical situations. In particular, it may unnecessarily over-emphasize the effect of noise and uncertainties and can, therefore, lead to overly conservative results (see, e.g., the explanation in Sec. 1.1 of the companion paper [4]). A similar situation arises in H_∞ filtering where designs are usually formulated to be robust with respect to finite-energy, but otherwise unknown, disturbances and modeling errors (see, e.g., [5]-[8]). The solutions are required to meet certain worst-case performance specifications and may again lead to overly conservative schemes.

This discussion motivates us to introduce a new parameter estimation formulation that incorporates an a-priori bound on the size of the data uncertainty. Connections with the TLS and H_∞ formulations are also addressed. In particular, the solution of the new problem turns out to require the minimization of a cost function in an "indefinite" metric space, in a way that seems at first sight to be similar to recent works on robust (or H_∞) estimation and filtering (e.g., [5]-[8]), but which turns out to be significantly different. The significant distinction is that the new cost function involves norms (or distances) rather than *squared* norms (or distances). Its solution turns out to lead to an automatic regularization procedure, where the regularization parameter is determined as the unique positive root of a certain nonlinear equation that we refer to as the *secular equation.* The solution involves an SVD step and its computational complexity amounts to $O(mn^2 + n^3)$, where n is the smaller matrix dimension.

*S. Chandrasekaran is with Dept. of ECE, UCSB, CA 93106. G. H. Golub is with Dept. of CS, Stanford University, CA 94305, M. Gu is with Dept. of Math., UCLA, CA 90095, and A. H. Sayed is with Dept. of EE, UCLA, CA 90095. The work of A. H. Sayed was in part supported by the National Science Foundation under Award No. MIP-9409319.

2. PROBLEM STATEMENT

Let $A \in \mathbf{R}^{m\times n}$ be a given full rank matrix with $m \geq n$ and let $b \in \mathbf{R}^m$ be a given vector. The quantities (A, b) are assumed to be linearly related via an unknown vector of parameters $x \in \mathbf{R}^n$, $b = A \cdot x + v$, where $v \in \mathbf{R}^m$ 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 $\|\delta A\|_2 \leq \eta$. The notation $\|\cdot\|_2$ denotes either the 2−induced norm of its matrix argument or the Euclidean norm of its vector argument.

We pose the problem of finding an estimate $\hat{x}$ that solves:

$$\min_{\hat{x}} \min\left\{ \|(A+\delta A)\cdot\hat{x} - b\|_2 : \|\delta A\|_2 \leq \eta \right\} . \quad (1)$$

Intuitively, this formulation corresponds to "choosing" a perturbation δA, within the bounded region, that would allow us to best predict the right-hand side b from the column span of $(A+\delta A)$. The situation is depicted in Fig. 1. Any particular choice for $\hat{x}$ would lead to many residual norms, $\|(A+\delta A)\cdot\hat{x} - b\|_2$, one for each possible choice of δA. We want to choose an estimate $\hat{x}$ that minimizes the minimum possible residual norm.

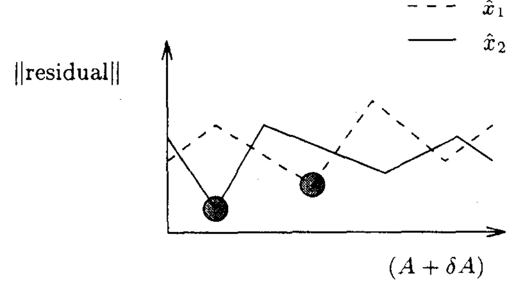

Figure 1: *Two illustrative residual-norm curves.*

It turns out that the existence of a unique solution to this problem will require a fundamental condition on the data (A, b, η), which we describe further ahead. When the condition is violated, the problem will become degenerate. In fact, such existence and uniqueness conditions also arise in other formulations of estimation problems (such as the TLS and H_∞ problems, which will be shown later to have some relation to the above optimization problem). In the H_∞ context, for instance, similar fundamental conditions arise, which when violated indicate that the problem does not have a meaningful solution (see, e.g., [5]-[8]).

2.1. Comparison with TLS

Given (A, b), the TLS solution finds the "smallest" δA (in a Frobenius norm sense) that would allow to estimate b from the column span of $(A + \delta A)$, viz., it solves the following problem [1,2]:

$$\min_{\delta A, \hat{x}} \left\| \begin{bmatrix} \delta A & (A+\delta A)\hat{x} - b \end{bmatrix} \right\|_F . \tag{2}$$

We therefore see that there is *not* an a priori bound on the size of the allowable perturbation δA. Although small in a certain (Frobenius) sense, the resulting correction δA in TLS need not satisfy an a-priori bound on its size. The problem we formulated above explicitly incorporates a bound on the size of the allowable perturbations.

We may remark in passing that we have posed and solved a related problem in the companion paper [3]; it guarantees optimal performance in a worst-case scenario with bounded data as well.

2.2. A Geometric Interpretation

The optimization problem (1) admits an interesting geometric formulation that highlights some of the issues involved in its solution.

For this purpose, and for the sake of illustration, assume we have a unit-norm vector b, $\|b\|_2 = 1$. Assume further that A is simply a column vector, say a, with $\eta \neq 0$. Now problem (1) becomes

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

The situation is depicted in Fig. 2. 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 $a + \delta a$.

For any $\hat{x}$ that we pick, the set $\{(a + \delta a)\hat{x}\}$ describes a disc of center $a\hat{x}$ and radius $\eta\hat{x}$. This is indicated in the figure by the largest rightmost circle, which corresponds to a choice of a positive $\hat{x}$ that is larger than one. The vector in $\{(a+\delta a)\hat{x}\}$ that is the closest to b is the one obtained by drawing a line from b through the center of the rightmost circle. The intersection of this line with the circle defines a residual vector r_3 whose norm is the smallest among all possible residual vectors in the set $\{(a + \delta a)\hat{x}\}$.

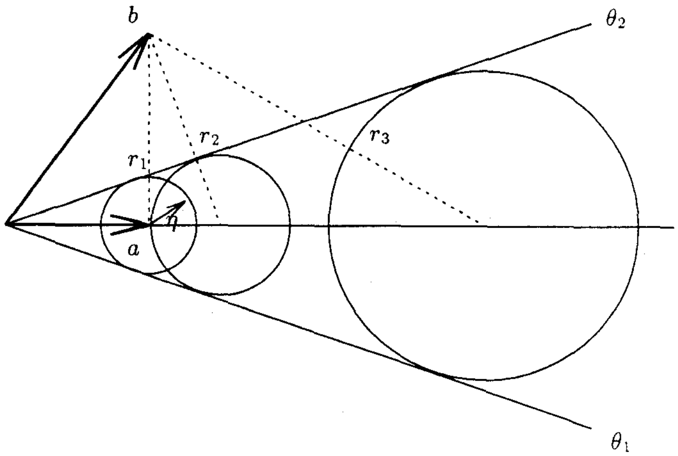

Figure 2: *Geometric construction in a simplified scenario.*

More generally, for any $\hat{x}$ that we pick, it will determine a circle and the corresponding smallest residual is obtained by finding the closest point on the circle to b. This is the point where the line that passes through b and the center of the circle intersects the circle on the side that is closest to b.

We need to pick an $\hat{x}$ that minimizes the smallest residual norm. The claim is that we can proceed as follows: we drop a perpendicular from b to the upper tangent line denoted by θ_2. This perpendicular intersects the horizontal line in a point where we draw a new circle (the middle circle) that is tangent to both θ_1 and θ_2. This circle corresponds to a choice of $\hat{x}$ such that the closest point on it to b is the foot of the perpendicular from b to θ_2. The residual indicated by r_2 is to the desired solution; it has the minimum norm among the smallest residuals.

3. AN EQUIVALENT PROBLEM

We start by showing how to reduce the optimization problem (1) to an equivalent problem. For this purpose, we note that

$$\| (A + \delta A) \cdot \hat{x} - b \|_2 \geq \left| \, \|A \cdot \hat{x} - b\|_2 - \|\delta A \cdot \hat{x}\|_2 \, \right| . \tag{3}$$

The lower bound on the right-hand side of the above inequality is a non-negative quantity and, therefore, the least it can get is zero. This will in turn depend on how big or how small the value of $\|\delta A\|_2$ can be. For example, if it happens that for all vectors $\hat{x}$ we always have

$$\eta \cdot \|\hat{x}\|_2 < \|A \cdot \hat{x} - b\|_2 , \tag{4}$$

then we conclude, using the triangle inequality of norms, that

$$\|\delta A \cdot \hat{x}\|_2 \leq \|\delta A\|_2 \cdot \|\hat{x}\|_2 \leq \eta \cdot \|\hat{x}\|_2 < \|A \cdot \hat{x} - b\|_2 .$$

It then follows from (3) that, under the assumption (4), we obtain

$$\|(A+\delta A)\cdot \hat{x} - b\|_2 \quad \geq \quad \|A\cdot \hat{x} - b\|_2 - \eta \cdot \|\hat{x}\|_2 \; .$$

It turns out that condition (4) is the main (and only) case of interest in this paper, especially since we shall argue later that a degenerate problem arises when it is violated. For this reason, we shall proceed for now with our analysis under the assumption (4) and shall postpone our discussion of what happens when it is violated until later in this section.

Now the lower bound in (3) is in fact achievable by choosing δA as the rank one matrix

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

This leads to a vector $\delta A^o \cdot \hat{x}$ that is collinear with the vector $(A\cdot \hat{x} - b)$ and, hence,

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

We are therefore reduced to the solution of the following equivalent optimization problem.

__Problem__. *Given a matrix $A \in \mathbf{R}^{m\times n}$, with $m \geq n$, a vector $b \in \mathbf{R}^m$, a nonnegative real number η, and assume that for all vectors $\hat{x}$ it holds that*

$$\eta \cdot \|\hat{x}\|_2 \; < \; \|A\cdot \hat{x} - b\|_2 \qquad \textit{(fundamental assumption)}. \quad (5)$$

Determine, if possible, an $\hat{x}$ that solves

$$\min_{\hat{x}} \; (\|A\cdot \hat{x} - b\|_2 - \eta \cdot \|\hat{x}\|_2) \; . \qquad (6)$$

3.1. Connections to TLS and $\mathbf{H}_\infty$

Before solving the above problem, we elaborate on its connections with other formulations in the literature that also attempt, in one way or another, to take into consideration uncertainties and perturbations in the data.

First, cost functions similar to (6) but with squared distances, say

$$\min_{\hat{x}} \; \left(\|A\cdot \hat{x} - b\|_2^2 - \gamma \cdot \|\hat{x}\|_2^2\right) \; , \qquad (7)$$

for some γ, often arise in the study of indefinite quadratic cost functions in robust or H^∞ estimation (see, e.g., the developments in [5, 9]). The major distinction between this cost and the one posed in (6) is that the latter involves *distance* terms and it will be shown to provide an automatic procedure for selecting a "regularization" factor that plays the role of γ in (6).

Likewise, the TLS problem seeks a matrix δA and a vector $\hat{x}$ that solve (2). The solution of the TLS problem is well-known and is given by the following construction [2][p. 36]. Let $\{\sigma_1, \ldots, \sigma_n\}$ denote the singular values of A, with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$. Let also $\{\bar{\sigma}_1, \ldots, \bar{\sigma}_n, \bar{\sigma}_{n+1}\}$ denote the singular values of the extended matrix $\begin{bmatrix} A & b \end{bmatrix}$, with $\bar{\sigma}_i \geq 0$. If $\bar{\sigma}_{n+1} < \sigma_n$, then a unique solution $\hat{x}$ of the TLS problem exists and is given by

$$\hat{x} = (A^T\cdot A - \bar{\sigma}_{n+1}^2 I)^{-1} A^T \cdot b. \qquad (8)$$

For our purposes, it is more interesting to consider the following interpretation of the TLS solution (see, e.g., [9]). Note that the condition $\bar{\sigma}_{n+1} < \sigma_n$ assures that $(A^T \cdot A - \bar{\sigma}_{n+1}^2 I)$ is a positive-definite matrix, since σ_n^2 is the smallest eigenvalue of $A^T A$. Therefore, we can regard (8) as the solution of the following optimization problem, with an indefinite cost function,

$$\min_{\hat{x}} \left(\|A\cdot \hat{x} - b\|_2^2 \; - \; \bar{\sigma}_{n+1}^2 \cdot \|\hat{x}\|_2^2\right) \; .$$

This is a special form of (7) with a particular choice for γ. It again involves squared distances, while (6) involves *distance* terms and it will provide another choice of a γ-like parameter. In particular, compare (8) with the expression (12) derived further ahead for the unique solution of (6). We see that the new problem replaces $\bar{\sigma}_{n+1}^2$ with a new parameter α that will be shown to be the unique positive root of a secular (nonlinear) equation.

3.2. The Degenerate Case

We solve problem (6) in the next section assuming (5) holds. When this condition is violated, it can be shown that the problem becomes degenerate in the following sense. If (5) is violated at some point $\hat{x}^{(1)}$, viz.,

$$\eta \cdot \|\hat{x}^{(1)}\|_2 \; \geq \; \|A\cdot \hat{x}^{(1)} - b\|_2 \; ,$$

then it turns out that $x^{(1)}$ can be taken as a solution to (1). Moreover, once such $x^{(1)}$ has been found, an infinite number of others can be constructed from it. For details, see [10].

We shall not treat the degenerate case in this paper (as well as the case when (5) is violated only with equality). We shall instead assume throughout that the fundamental condition (5) holds. Under this assumption, the problem will turn out to always have a unique solution.

3.3. The Fundamental Assumption

The fundamental condition (5) needs to be satisfied for all vectors $\hat{x}$. This can be restated in terms of conditions on the data (A, b, η) alone. To see this, note that (5) implies, by squaring, that we must have, for all $\hat{x}$,

$$J(\hat{x}) \triangleq \hat{x}^T \cdot (\eta^2 \cdot I - A^T \cdot A)\cdot \hat{x} + 2\hat{x}^T \cdot A^T \cdot b - b^T \cdot b \; < 0 \; .$$

Since $J(\hat{x})$ is quadratic in $\hat{x}$, this is only possible if

(i) $J(\hat{x})$ has a maximum with respect to $\hat{x}$, and

(ii) the value of $J(\hat{x})$ at its maximum is negative.

The necessary condition for the existence of a unique maximum (since we have a quadratic cost function) is $(\eta^2 \cdot I - A^T \cdot A) < 0$, which means that η should satisfy $\eta < \sigma_{min}(A)$. Under this condition, the expression for the maximum point $\hat{x}_{max}$ of $J(\hat{x})$ is

$$\hat{x}_{max} = (A^T \cdot A - \eta^2 \cdot I)^{-1} \cdot A^T \cdot b \,.$$

Evaluating $J(\hat{x})$ at $\hat{x} = \hat{x}_{max}$ we obtain

$$J(\hat{x}_{max}) = b^T \cdot \left[A \cdot (A^T \cdot A - \eta^2 \cdot I)^{-1} \cdot A^T \; - \; I\right] \cdot b \,.$$

Therefore, the requirement that $J(\hat{x}_{max})$ be negative corresponds to

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

In summary, we are led to the following result.

<u>Lemma</u>. *The necessary and sufficient conditions in terms of (A, b, η) for the fundamental relation (5) to hold are:*

$$(\eta^2 \cdot I - A^T \cdot A) \; < \; 0 \qquad \Longleftrightarrow \qquad \eta \; < \; \sigma_{min}(A) \,, \qquad (9)$$

$$b^T \cdot \left[I \; - \; A \cdot (A^T \cdot A - \eta^2 \cdot I)^{-1} \cdot A^T\right] \cdot b \; > 0 \,. \qquad (10)$$

Note that for a well-defined problem of the form (1) we need to assume $\eta > 0$ which, in view of (9), means that A should be full rank.

Note also that the fundamental condition (5) rules out the condition $A \cdot \hat{x} = b$, i.e., it rules out the case when b lies in the range space of A.

4. ALGEBRAIC SOLUTION

To solve (6), 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 (5). 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, $\nabla\mathcal{L}(\hat{x})$, 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 (5).

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

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

where we have introduced the positive real number

$$\alpha = \frac{\eta \cdot ||A \cdot \hat{x} - b||_2}{||\hat{x}||_2} \,. \qquad (11)$$

In view of the fundamental condition (5) we see that the value of α is necessarily larger than η^2, $\alpha > \eta^2$.

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

$$\hat{x} = \left(A^T \cdot A - \alpha I\right)^{-1} \cdot A^T \cdot b \,. \qquad (12)$$

Expressions (11)-(12) define a system of equations with two unknowns $\{\hat{x}, \alpha\}$. If we replace (12) into (11) we obtain a nonlinear equation in α, which will further lead to what we shall refer to as the *secular equation.* We shall show later that the secular equation has only one positive root. Hence, determining its positive root uniquely determines α, which in turn uniquely determines $\hat{x}$.

Since we are also interested in the numerical reliability of the resulting computational procedure, we find it useful to perform these substitutions and calculations by invoking the SVD of A, say

$$A = U \cdot \begin{bmatrix} \Sigma \\ 0 \end{bmatrix} \cdot V^T \,, \qquad (13)$$

where $U \in \mathbf{R}^{m \times m}$ and $V \in \mathbf{R}^{n \times n}$ are orthogonal, and $\Sigma = \mathrm{diag}(\sigma_1, \cdots, \sigma_n)$ is diagonal, with $\sigma_1 \geq \cdots \geq \sigma_n > 0$. We further partition the vector $U^T b$ into $U^T b = \mathrm{col}\{b_1, b_2\}$, where $b_1 \in \mathbf{R}^n$ and $b_2 \in \mathbf{R}^{m-n}$.

In this case, the expression (11) reduces to

$$\alpha = \frac{\eta \cdot \sqrt{||b_2||_2^2 + \alpha^2 \cdot ||\left(\Sigma^2 - \alpha I\right)^{-1} \cdot b_1||_2^2}}{||\Sigma \cdot \left(\Sigma^2 - \alpha I\right)^{-1} \cdot b_1||_2} \,. \qquad (14)$$

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

4.1. The Secular Equation

Define the nonlinear function in α,

$$\mathcal{G}(\alpha) = b_1^T \cdot (\Sigma^2 - \eta^2 I) \cdot \left(\Sigma^2 - \alpha I\right)^{-2} \cdot b_1 - \frac{\eta^2}{\alpha^2} \cdot ||b_2||_2^2 \,. \quad (15)$$

It follows that α is a positive solution to (14) if, and only if, it is a positive root of $\mathcal{G}(\alpha)$. Following [1], we refer to the equation $\mathcal{G}(\alpha) = 0$ as a *secular* equation.

It can be shown that, under the the fundamental condition (5), the function $\mathcal{G}(\alpha)$ has a unique positive root and that this root lies in the interval (η^2, σ_n^2) (proofs are omitted for brevity).

Figure 3 provides a sketch of the behaviour of the secular equation $\mathcal{G}(\alpha)$ as a function of α. In particular, note that $\lim_{\alpha \to +\infty} \mathcal{G}(\alpha) = 0$.

4.2. Finding the Global Minimum

Our original motivation has been to solve (6), and hence minimize the cost function

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

Since $\mathcal{L}(\hat{x}) \geq 0$, and it is continuous in $\hat{x}$, it necessarily has a minimum (possibly at infinity). The candidates for

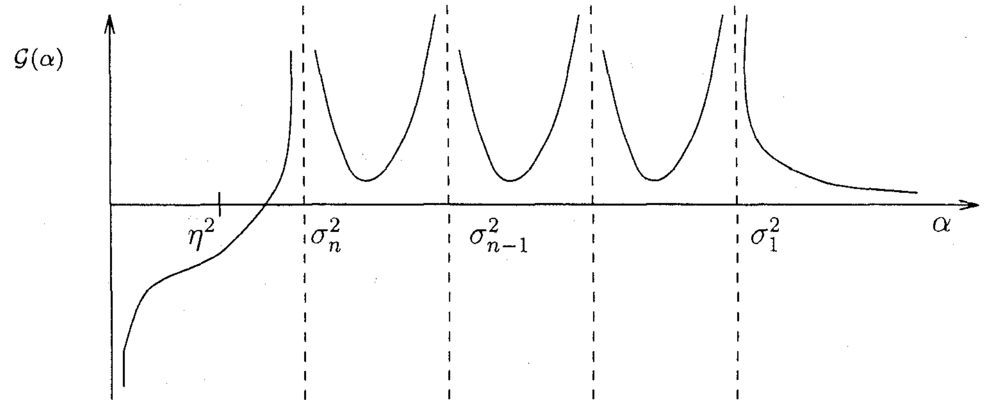

Figure 3: *A plot of the secular function for $\alpha \geq 0$.*

a minimizing $\hat{x}$ are either the points at which the gradient of $\mathcal{L}(\hat{x})$ vanishes or the points $\hat{x} = 0$ and $\hat{x} = \infty$.

It is clear that we can rule out $\hat{x} = \infty$ since $\lim_{\|\hat{x}\|_2 \to \infty} \mathcal{L}(\hat{x}) = \infty$. Likewise, we can rule out the choice $\hat{x} = 0$ since we can show that $\mathcal{L}(0) > \mathcal{L}((A^T \cdot A - \alpha I)^{-1} \cdot A^T \cdot b)$.

__Theorem__. *Given $A \in \mathbf{R}^{m \times n}$, with $m \geq n$ and A full rank, $b \in \mathbf{R}^m$, and a nonnegative real number $\eta < \sigma_{min}(A)$. Assume further that the conditions (9)–(10) are satisfied. Then problem (1) has a unique solution that can be constructed as follows.*

- *Introduce the SVD of A as in (13).*
- *Partition the vector $U^T b$ into $U^T b = col\{b_1, b_2\}$.*
- *Introduce the secular function $\mathcal{G}(\alpha)$ as in (15).*
- *Determine the unique positive root $\hat{\alpha}$ of $\mathcal{G}(\alpha)$ – it lies in the interval (η^2, σ_n^2). This can be achieved via a bisection method with cost $O\left(n \log \frac{\sigma_n^2}{\epsilon}\right)$, where ϵ is the desired precision.*
- *Then $\hat{x} = (A^T \cdot A - \hat{\alpha} \cdot I)^{-1} A^T \cdot b$.*

5. FURTHER REMARKS

The discussion can be further extended to handle cases when only selected columns of A are uncertain (see [4,10]).

The solution proposed herein requires the computation of the SVD of the data matrix and the determination of the unique positive root of the nonlinear secular equation. In the companion paper [11], we pursue an iterative solution that exploits a certain contraction mapping relation.

Moreover, in the second companion paper [4] we consider an alternative formulation of the estimation problem that is suitable for worst-case identification with bounded data uncertainties.

References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, third ed., 1997.

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

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

[4] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, *Worst-case parameter estimation with bounded model uncertainties*, in Proceedings of this Conference (ACC'97).

[5] B. Hassibi, A. H. Sayed, and T. Kailath, *Recursive linear estimation in Krein spaces - Part I: Theory*, IEEE Trans. Automatic Control, vol. 41, no. 1, pp. 18–33, Jan. 1996.

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

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

[8] J. C. Doyle, K. Glover, P. Khargonekar, and B. Francis, *State-space solutions to standard H_2 and H_∞ control problems*, IEEE Transactions on Automatic Control, 34 (1989), pp. 831–847.

[9] A. H. Sayed, B. Hassibi, and T. Kailath, *Fundamental inertia conditions for the minimization of quadratic forms in indefinite metric spaces*, Operator Theory: Advances and Applications, Birkhauser, 1996.

[10] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, *An efficient algorithm for a bounded errors-in-variables model*, submitted for publication.

[11] A. H. Sayed, A. Garulli, and S. Chandrasekaran, *A fast iterative solution for worst-case parameter estimation with bounded model uncertainties*, in Proceedings of this Conference (ACC'97).