Parameter Estimation in the Presence of Bounded Modeling Errors

S. Chandrasekaran, G. H. Golub, *Honorary Member, IEEE,* M. Gu, and A. H. Sayed, *Member, IEEE*

***Abstract*—We formulate and solve a new parameter estimation problem in the presence of bounded data uncertainties. The new method is suitable when *a priori* bounds on the uncertain data are available; its solution guarantees that the effect of the uncertainties will never be unnecessarily overestimated beyond what is reasonably assumed by the *a priori* bounds.**

I. 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, 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 (LS) criterion, which enjoys widespread popularity as a result of its attractive computational and statistical properties. But many alternative optimization criteria have been proposed over the years in order to improve the performance of standard LS estimators in the presence of data uncertainties (e.g., [1]–[4]). Among these we may mention regularized LS, ridge regression, total LS, and robust (or H^∞) estimation. They all allow, in one way or another, for the incorporation of some *a priori* information about the unknown parameter into the problem statement. They are also more effective in the presence of data errors and incomplete statistical information about the exogenous signals (or measurement errors). Nevertheless, these variations can still unnecessarily overemphasize the effect of noise and uncertainties and can, therefore, lead to overly conservative designs.

In this paper, we formulate and solve a new parameter estimation problem with prior bounds on the size of the allowable corrections to the data. A detailed analysis of the new problem, and comparisons with earlier approaches, can be found in the extended paper [5]. Variations and recursive solutions can be found in [6] and [7]. Here, we only wish to report the main ideas and results.

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 $x \in \mathbf{R}^n$, $b = Ax + v$, where $v \in \mathbf{R}^m$ explains the mismatch between Ax 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

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

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

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 the following setting:

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

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 $a+\delta a$. 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 right-most 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 furthest away from b is the one obtained by drawing a line from b through the center of the right-most circle. The intersection of this line with the circle defines a residual vector r_3 whose norm is the largest among all possible residual vectors in the set $\{(a+\delta a)\hat{x}\}$.

Manuscript received April 10, 1996. This work was supported in part by the National Science Foundation under Award MIP-9409319. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. T. S. Durrani.

S. Chandrasekaran is with the Department of Electrical and Computer Engineering, University of California, Santa Barbara, Santa Barbara, CA 93106 USA.

G. H. Golub is with the Department of Computer Science, Stanford University, Stanford, CA 94305 USA.

M. Gu is with the Department of Mathematics, University of California, Los Angeles, Los Angeles, CA 90095 USA.

A. H. Sayed is with the Department of Electrical Engineering, University of California, Los Angeles, CA 90095 USA (e-mail: sayed@ee.ucla.edu).

Publisher Item Identifier S 1070-9908(97)05232-2.

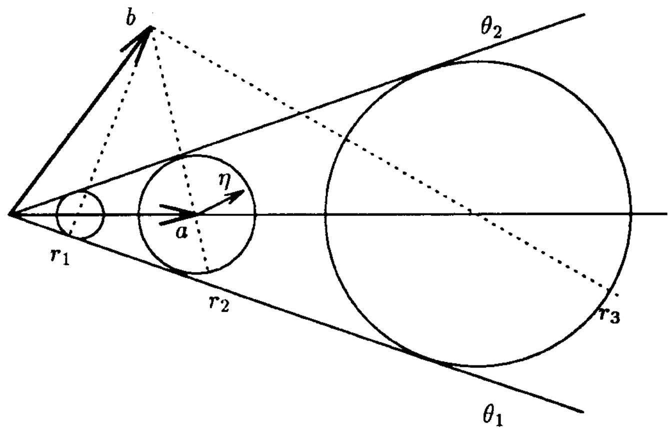

Fig. 1. Geometric construction of the solution for a simple example.

It can be verified that the solution can be obtained geometrically as follows: Drop a perpendicular from b to the lower tangential line θ_1. Pick the point where the perpendicular meets the horizontal line and draw a circle that is tangent to both θ_1 and θ_2. Its radius will be $\eta\hat{x}$, where $\hat{x}$ is the optimal solution. Also, the foot of the perpendicular on θ_1 will be the optimal $\hat{b}$. The segment r_1 denotes the optimum residual (it has the minimum norm among the largest residuals). More details can be found in [5].

II. An Algebraic Solution

It can be shown that problem (1) reduces to the equivalent minimization problem (of a convex cost function) as follows:

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

Note, in particular, that this problem formulation is significantly distinct from a regularized LS formulation, where the *squared* Euclidean norms $\{\|A\hat{x} - b\|_2^2, \|\hat{x}\|_2^2\}$ are used rather than the norms themselves!

The solution to this minimization problem is given as follows [5]. Introduce the SVD of A: $A = U[\Sigma^T\, 0]^T V^T$, partition the vector $U^T b$ into $[b_1^T\, b_2^T]^T = U^T b$, where $b_1 \in \mathbf{R}^n$ and $b_2 \in \mathbf{R}^{m-n}$, and define the function

$$\mathcal{G}(\alpha) = b_1^T(\Sigma^2 - \eta^2 I)(\Sigma^2 + \alpha I)^{-2}\, b_1 - \frac{\eta^2}{\alpha^2}\, \|b_2\|_2^2. \tag{4}$$

Further define $\tau_1 = \|\Sigma^{-1}b_1\|_2 / \|\Sigma^{-2}b_1\|_2$ and $\tau_2 = \|A^T b\|_2/\|b\|_2 = \|\Sigma b_1\|_2/\|b\|_2$. Then two cases are possible.

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

Note that the expression $\hat{x} = (A^T A + \hat{\alpha} I)^{-1} A^T b$ can be regarded as the exact solution of a regularized LS problem of the form

$$\min_{\hat{x}} (\hat{\alpha}\|\hat{x}\|_2^2 + \|A\hat{x} - b\|_2^2)$$

with squared Euclidean distances. In this sense, the solution to the original problem (3) (with norms only rather than squared norms) can be seen to lead to automatic regularization. That is, the solution first determines a regularization parameter $\hat{\alpha}$ and then uses it to solve a regularized least-squares problem of the above form. The scalar $\hat{\alpha}$ can be determined by employing a bisection-type algorithm to solve the secular equation, thus requiring $O[n \log(\hat{\alpha}/\epsilon)]$, where ϵ is the desired precision.

III. Variations

There are several variations that submit to algebraic solutions. We only mention two examples:

Uncertain Weights: Consider the min–max problem

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

It reduces to

$$\min_{\hat{x}} [\|W(A\hat{x} - b)\|_2 + \eta_w \|A\hat{x} - b\|_2]$$

and the optimal solution can be shown to satisfy

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

where $\hat{\alpha}$ satisfies a secular equation similar in form to $\mathcal{G}$ in (4). The details will be published elsewhere.

Multiplicative Uncertainties: Consider the min–max problem

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

It reduces to

$$\min_{\hat{x}} [\|A\hat{x} - b\|_2 + \eta_a \|A\hat{x}\|_2]$$

$$\hat{\alpha} = \begin{cases} \eta_a \|P^\perp b\| \left(\dfrac{\eta_a \|P^\perp b\| + \|Pb\| \sqrt{1 - \eta_a^2}}{\|Pb\|^2 - \eta_a^2 \|b\|^2} \right) & \text{if } \eta_a \|b\| < \|Pb\| \\ \infty & \text{otherwise} \end{cases} \tag{5}$$

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

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

where $\hat{\alpha}$ is given by (5), shown at the bottom of the previous page, where $P = A(A^T A)^{-1}A^T$ and $P^{\perp} = I - P$.

IV. 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 formulation (1)—see, e.g., [5] and [6]. Also, weighted versions with uncertainties in the weight matrices are useful in several applications, as well as cases with more general multiplicative uncertainties. Recursive solutions are also of interest, and results in this direction appear in [7]. We should also mention related work in [8], where the authors have independently formulated and solved an estimation problem similar to (1) by using (convex) semidefinite programming techniques.

References

[1] G. H. Golub and C. F. Van Loan, *Matrix Computations,* 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1997.
[2] S. V. Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis.* Philadelphia, PA: SIAM, 1991.
[3] P. Khargonekar and K. M. Nagpal, "Filtering and smoothing in an H^{∞}-setting," *IEEE Trans. Automat. Contr.,* vol. 36, pp. 151–166, 1991.
[4] B. Hassibi, A. H. Sayed, and T. Kailath, "Recursive linear estimation in Krein spaces—Part I: Theory," *IEEE Trans. Automat. Contr.,* vol. 41, pp. 18–33, Jan. 1996.
[5] 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.,* 1997, to be published.
[6] ——, "Best-fit parameter estimation for a bounded errors-in-variables model," in *Proc. Amer. Control Conf.,* Albuquerque, NM, June 1997.
[7] A. H. Sayed, A. Garulli, and S. Chandrasekaran, "A fast iterative solution for worst-case parameter estimation with bounded model uncertainties," in *Proc. Amer. Control Conf.,* Albuquerque, NM, June 1997.
[8] L. E. Ghaoui and H. Lebret, "Robust solutions to least squares problems with uncertain data," *SIAM J. Matrix Anal. Appl.,* to be published.