LINEAR ALGEBRA AND ITS APPLICATIONS, VOL. 284, PP. 259–306, NOV. 1998

ESTIMATION AND CONTROL WITH BOUNDED DATA UNCERTAINTIES*

A. H. SAYED[1], V. NASCIMENTO[1], AND S. CHANDRASEKARAN[2]

[1] ELECTRICAL ENGINEERING DEPARTMENT
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CA 90095

[2] DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING
UNIVERSITY OF CALIFORNIA
SANTA BARBARA, CA 93106

Abstract. The paper describes estimation and control strategies for models with bounded data uncertainties. We shall refer to them as BDU estimation and BDU control methods, for brevity. They are based on constrained game-type formulations that allow the designer to explicitly incorporate into the problem statement a-priori information about bounds on the expected sizes of the uncertainties. In this way, the effect of uncertainties is not unnecessarily over-emphasized beyond what is implied by the a-priori bounds; consequently, overly conservative designs, as well as overly sensitive designs, are avoided. A feature of these new formulations is that geometric insights and recursive techniques, which are widely known and appreciated for classical quadratic-cost designs, can also be pursued in this new framework. Also, algorithms for computing the optimal solutions with the same computational effort as standard least-squares solutions exist, thus making the new formulations attractive for practical use. Moreover, the framework is broad enough to encompass applications across several disciplines, not just estimation and control. Examples will be given of a quadratic control design, an $\mathcal{H}_\infty$ control design, a total-least-squares design, image restoration, image separation, and co-channel interference cancellation.

A major theme in this paper is the emphasis on geometric and linear algebraic arguments. Despite the interesting results that will be discussed, several issues remain open and indicate potential future developments; these will be briefly discussed.

1. INTRODUCTION.

A fundamental problem in estimation is to recover to good accuracy a set of unobservable parameters from corrupted, incomplete, or distorted data. Likewise, a fundamental problem in control is to determine suitable control signals for possibly erroneous models. Examples to both effects abound in the fields of signal processing, system identification, image processing, digital communications, statistics, and others, as can be found in many textbooks – see, *e.g.,* [1]–[11] and the many references therein. In all these fields, several optimization criteria have been proposed over the years for design purposes. Some of the most distinguished criteria are the following.

a) The least-squares (LS) method, which has been one of the most widely used design criteria since its inception by Gauss (around 1795) in his studies on celestial mechanics (*e.g.,* [12, 13, 14]).

*This article is based on an invited lecture delivered by the first author at the ILAS International Symposium on Fast Algorithms for Control, Signal, and Image Processing, and which was held in Manitoba, Canada, June 1997. The works of A. H. Sayed and S. Chandrasekaran were supported in part by the National Science Foundation under Award numbers MIP-9796147 and CCR-9734290, respectively. The work of V. Nascimento was also supported by a fellowship from Conselho Nacional de Pesquisa e Desenvolvimento - CNPq - Brazil, while on leave from Escola Politécnica da Universidade de São Paulo.

b) The regularized least-squares method, which is used to combat much of the ill-conditioning that arises in pure LS problems (*e.g.,* [12, 15, 16]).
c) The total-least-squares (TLS) or errors-in-variables method, which provides a way to deal with uncertainties in the data (*e.g.,* [17, 18]).
d) The $\mathcal{H}_\infty$ approach, which combats uncertainties in the data by designing for the worst possible scenario (*e.g.,* [9, 11]).
e) The l_1 approach for robust identification and control, which exploits linear programming and interpolation techniques (*e.g.,* [19]).
f) The set-membership identification approach, which is based on constructing converging ellipsoids that encircle the unknown parameter (*e.g.,* [20, 21]).

Among the most successful design criteria, which submit to analytical studies and derivations and which have had the most applications in identification, control, signal processing, and communications, the least-squares criterion of C. F. Gauss (1795) stands out unchallenged [14]. It was also independently formulated by A. M. Legendre in 1805, who praised the method in no uncertain terms (e.g., [13]):

" Of all the principles that can be proposed, I think there is none more general, more exact, and more easy of application, than that which consists of rendering the sum of squares of the errors a minimum."

A. M. Legendre (Paris, 1805)

In this paper, we propose and study new design criteria for estimation and control purposes that are based on new cost functions. In order to appreciate the significance of the new formulations, we first provide an overview of some of the existing methods in Secs. 2 and 3. We then motivate and introduce the new cost functions in Sec. 4. In Sec. 5 we study in detail the estimation problem and in Sec. 6 we study the control problem. One major theme in our arguments is the emphasis on geometric and linear algebraic arguments, which lead to useful insights about the nature of the new formulations. Also, throughout the paper, several examples from the fields of image processing, communications, and control are included for illustrative purposes.

We start by reviewing the least-squares problem.

2. THE LEAST-SQUARES CRITERION. The least-squares method forms the backbone of many well-developed theories in estimation and control including Kalman filtering, linear quadratic control, and identification methods. Its popularity is due to several good reasons.

To begin with, the least-squares criterion is extremely simple to state and solve. Given a noisy measurement vector b that is related to an unknown vector x via the linear model

$$b = Ax + v \; , \tag{2.1}$$

for some known matrix A, we estimate x by solving

$$\min_x \; \|Ax - b\| \; , \tag{2.2}$$

where the dimensions of A are taken to be $N \times n$ with $N \geq n$ [we use the capital letter N to denote the larger dimension of A and the letter n to denote the smaller dimension of A]. Here, the notation $\|\cdot\|$ denotes the Euclidean norm of its vector argument (it will also be used to denote the maximum singular value of a matrix argument).

2.1. The Orthogonality Condition. The vector v in the model (2.1) denotes a noise term that explains the mismatch between the measured vector b and the vector Ax. In the absence of v, the vector b would lie in the column span of A, denoted by $\mathcal{R}(A)$. Due to v, the vector b will not in general lie in $\mathcal{R}(A)$. The least-squares problem therefore seeks the vector $\hat{b} = A\hat{x}$ in $\mathcal{R}(A)$ that is closest to b in the Euclidean norm sense. The solution of (2.2) can be obtained by solving the normal equations

$$(A^T A)\hat{x} = A^T b. \tag{2.3}$$

These equations can have multiple solutions $\hat{x}$, depending on whether A has full column rank or not. However, regardless of which solution $\hat{x}$ we pick, the so-called projection of b onto $\mathcal{R}(A)$, given by $\hat{b} = A\hat{x}$ is unique. When A is full rank, this is given by

$$\hat{b} = A(A^T A)^{-1} A^T b \ \stackrel{\Delta}{=}\ \mathcal{P}_A b$$

where we use the symbol $\mathcal{P}_A$ to denote the orthogonal projection matrix onto the column span of A (it satisfies $\mathcal{P}_A^2 = \mathcal{P}_A$ and $\mathcal{P}_A^T = \mathcal{P}_A$.). These are well-known properties of least-squares solutions (*e.g.,* [2, 12, 13]).

The normal equations (2.3) also show that the least-squares solution $\hat{x}$ satisfies an important geometric property, *viz.*, that the residual vector $(A\hat{x} - b)$ is necessarily orthogonal to the data matrix (see Fig. 2.1),

$$A^T(A\hat{x} - b) = 0\ . \tag{2.4}$$

We shall see later in Sec. 5 that this useful geometric property extends to the BDU case.

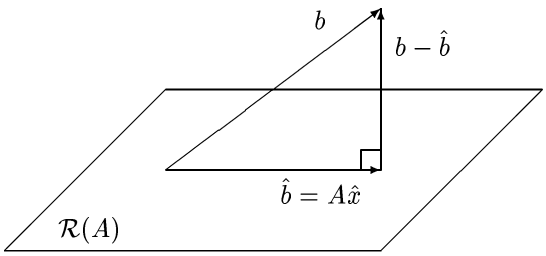

FIG. 2.1. The residual vector is orthogonal to $\mathcal{R}(A)$.

It is further well-known that the solution $\hat{x}$ of least-squares problems can be updated in $O(n^2)$ operations when a new row is added to A and a new entry is added to b. This is achieved via the so-called recursive least-squares (RLS) method (also derived by C. F. Gauss), and by many of its variants that are nowadays widely employed in adaptive filter theory (see, *e.g.,* [6, 8, 22]). We may add that there are also a variety of reliable algorithms and software available for least-squares based designs [12, 13, 16, 23].

2.2. Sensitivity to Data Errors. Given all the above useful properties of least-squares solutions, the natural question is to wonder why we would need to consider alternatives to the least-squares method? One prominent reason that has attracted much attention, especially in the signal processing and control communities, is that least-squares methods are sensitive to errors in the data.

More specifically, a least-squares design that is based on given data (A, b) can perform poorly if the true data happens to be a perturbed version of (A, b), say $(A + \delta A, b)$ for some unknown δA. Indeed, assume that a solution $\hat{x}$ has been determined using (2.3), where b is assumed to have been obtained from a noisy measurement of Ax, as in (2.1). Now if the b that we are using has in fact been obtained not from A but from a perturbed A, say $A + \delta A$,

$$b = (A + \delta A)x + v \ ,$$

then the $\hat{x}$ computed from (2.3) will result in a residual norm that satisfies, in view of the triangle inequality of norms,

$$\text{new residual} \ = \ \|(A + \delta A)\hat{x} - b\| \ \leq \ \underbrace{\|A\hat{x} - b\|}_{\text{LS residual}} \ + \ \underbrace{\|\delta A \hat{x}\|}_{\text{additional term}} \tag{2.5}$$

The first term on the right-hand side is equal to the least-squares residual norm that is associated with $(A, b, \hat{x})$. The second term is the increase in the residual norm due to the perturbation δA in the data.

Perturbation errors in the data are very common in practice and they can be due to several factors including the approximation of complex models by simpler ones, the presence of unavoidable experimental errors when collecting data, or even due to unknown or unmodelled effects. Regardless of their source, Eq. (2.5) shows that they can degrade the performance of least-squares designs. Two simple examples that illustrate this effect in the context of image processing and quadratic control are discussed below.

2.3. Image Restoration Example. Consider a two-dimensional $N \times N$ image (Fig. 2.2(a)) and collect its pixels into an $N^2 \times 1$ vector x. Blurring occurs by applying a matrix A to x, in addition to additive noise, thus leading to a blurred image vector b, say $b = Ax + v$ (see Fig. 2.2(b)). We can recover the original image x from b by using the least-squares solution, say $\hat{x} = (A^T A)^{-1} A^T b$, as shown in Fig. 2.2(c). But what if the blur was not caused by A but by $(A + \delta A)$, for some unknown δA? That is, what if the vector b that we are using came from $b = (A + \delta A)x + v$ and not from $b = Ax + v$? In this case, the $\hat{x}$ constructed above need not recover the original image satisfactorily. The situation is depicted in Figs. 2.2(d) and 2.2(e). Fig. 2.2(d) shows the original image blurred by $(A + \delta A)$, where the relative size of δA to A is about 8.5% (measured in terms of the ratio of their maximum singular values). Figs. 2.2(b) and 2.2(d) are similar, yet Fig. 2.2(e) shows that the least-squares solution fails in the perturbed case. Several regularization methods that are superior to the pure least-squares method have been proposed in the literature for image restoration purposes, some of which are discussed in [24]–[28]. We shall have more to say about regularization in the sequel (see Secs. 3.1 and 5.6).

2.4. Linear Quadratic Regulator Example. Another well-known manifestation of the sensitivity of least-squares-based designs to modeling errors occurs in quadratic control (see, *e.g.,* [9, 11, 30, 33]). In the so-called linear quadratic regulator (LQR) problem, the primary objective is to regulate the state of a linear state-space model to zero while keeping the control cost low.

Consider the simple one-dimensional state-space model,

$$x_{i+1} = f x_i + g u_i \ , \quad f = 0.9, \quad g = 1 \ , \ \ x_0 = 10, \tag{2.6}$$

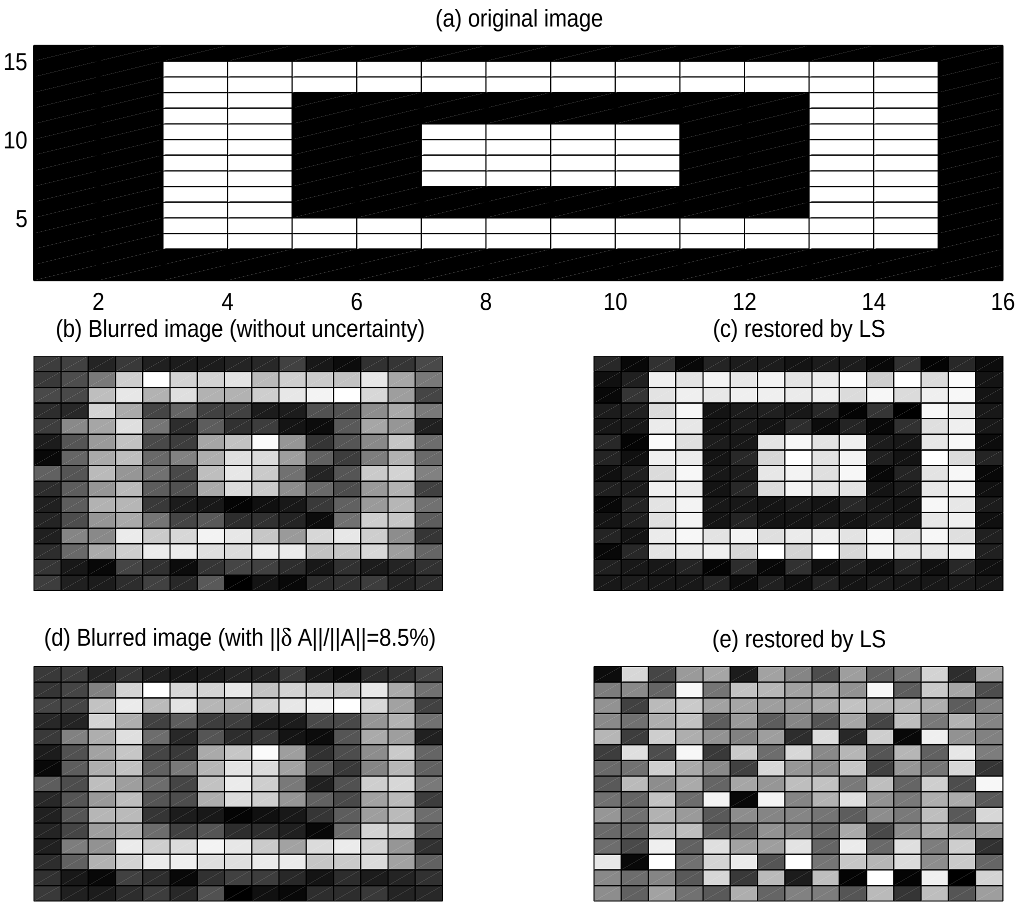

FIG. 2.2. Recovering an image from its blurred versions.

where x_0 denotes the value of the initial state, and the $\{u_i\}$ denote the control (input) sequence. In the LQR problem, we seek a control sequence $\{u_i\}$ that solves

$$\min_{\{u_j\}} \left(p x_{N+1}^2 \ + \ \sum_{j=0}^{N} \left[q u_j^2 + r x_j^2 \right] \right) \ , \quad q > 0, \ \ r \geq 0, \ \ p > 0. \tag{2.7}$$

for some given $\{q, r, p\}$ and over an interval of time $0 \leq j \leq N$. The cost function in (2.7) penalizes the control $\{u_j\}$, the state trajectory $\{x_j\}$, and the final state (at time $N+1$). Hence, intuitively, the LQR solution tries to keep the state trajectory close to zero by employing a low energy control sequence.

It is well known that the LQR problem can be solved recursively as follows. We split the cost function into two terms and write,

$$\min_{\{u_0,\ldots,u_{N-1}\}} \left(\sum_{j=0}^{N-1} \left[q u_j^2 + r x_j^2 \right] \ + \ \min_{u_N} \left[p x_{N+1}^2 \ + \ q u_N^2 + r x_N^2 \right] \right) \ , \tag{2.8}$$

where only the second term, through the state-equation (2.6) for x_{N+1}, is dependent on u_N. Minimizing over u_N then leads to the following state-feedback law,

$$\left\{\begin{array}{l} \hat{u}_N = -k_N x_N \\ \\ k_N = \frac{f g p_{N+1}}{q+g^2 p_{N+1}} \\ \\ p_N = f^2 p_{N+1} - \frac{k_N^2}{q+g^2 p_{N+1}} + r \ , \quad p_{N+1} = p \end{array}\right. \tag{2.9}$$

These equations show that the optimal control at time N is a scaled multiple of the state at the same time instant N. The gain k_N is defined in terms of the given model parameters $\{f, g, q\}$ and in terms of the cost p_{N+1}.

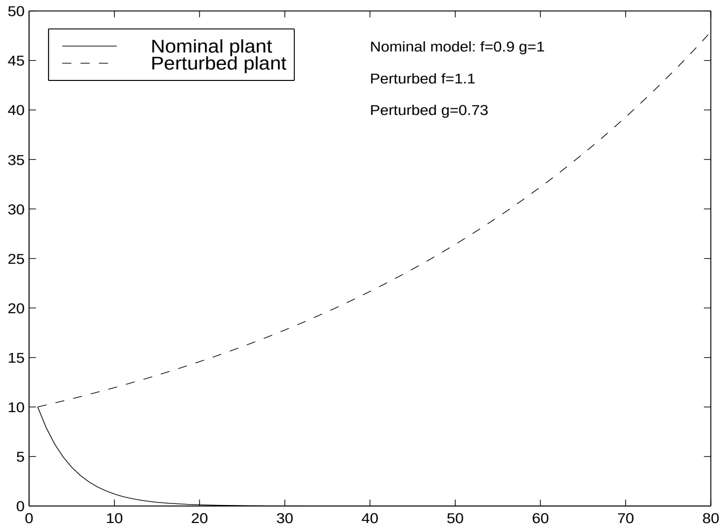

FIG. 2.3. An LQR design with a perturbed model.

More generally, at any particular time instant i, the optimal control signal $\hat{u}_i$ will be a scaled multiple of the state at that time instant, x_i. The gain k_i will be determined in terms of the given quantities $\{f, g, q\}$ and in terms of an intermediate quantity p_{i+1} that is propagated via the Riccati recursion

$$p_i = f^2 p_{i+1} - \frac{k_i^2}{q + g^2 p_{i+1}} + r \ , \quad 0 \leq i \leq N.$$

with boundary condition $p_{N+1} = p$. The state of the controlled (also called closed-loop) system will therefore evolve along the trajectory

$$x_{i+1} = (f - g k_i) x_i \ .$$

The solid line in Fig. 2.3 shows the evolution of the state of the closed-loop nominal system. It decays to zero and the overall cost for $N = 80$ is 13.86. Also, the closed-loop pole in steady-state (*i.e.,* the value of $f - g k_i$ for large enough i) tends to 0.79044. But how does this solution perform when the actual model is not defined by (f, g) but by $(f + \delta f)$ and $(g + \delta g)$, for some unknown $(\delta f, \delta g)$? In this case, the state will evolve along the perturbed trajectory

$$x_{i+1} = [f + \delta f - (g + \delta g) k_i] x_i \ .$$

The dotted line in Fig. 2.3 shows the state evolution of the closed-loop perturbed system for some $\{\delta f, \delta g\}$; it clearly grows unbounded and the overall cost for $N = 80$ is 9025.9. The closed-loop pole now tends to 1.02. Similar issues arise in Kalman filtering design (*e.g.,* [2, 31, 32, 33, 34]).

3. SOME ALTERNATIVE DESIGN METHODS. The alternative design methods that we listed before in Sec. 1 address in their own ways the sensitivity of least-squares solutions to uncertain data. In this section we comment briefly on the regularized least-squares method, the total-least-squares method, and the $\mathcal{H}_\infty$ method.

3.1. Regularized Least-Squares. Regularized least-squares methods have been proposed to combat the sensitivity of least-squares solutions to ill-conditioned data [15], where by ill-conditioning it is meant that small changes in the data may lead to large changes in the result.

Regularization involves choosing in advance a positive parameter γ and then selecting x by solving (*e.g.,* [12, 15, 16])

$$\min_x \left[\gamma\|x\|^2 + \|Ax - b\|^2\right] \; . \tag{3.1}$$

The solution $\hat{x}$ is now unique and given by

$$\hat{x} = [A^T A + \gamma I]^{-1} A^T b \; . \tag{3.2}$$

The uniqueness of $\hat{x}$ is due to the fact that the coefficient matrix $(A^T A + \gamma I)$ is always invertible (in fact, positive-definite and better conditioned than $A^T A$ in the pure least-squares method). Applications of such regularized costs in the image processing context abound and can be found, for example, in [24]–[28].

It will turn out that the BDU methods discussed further ahead in this paper perform automatic regularization. That is, while the above classical regularization method still requires an intelligent selection of the parameter γ by the designer, the BDU methods will select the the parameter γ from the given data without user intervention and in a certain optimal manner (see Sec. 5.6). We shall also compare these approaches with the so-called cross-validation method [12, 13, 36], which is a procedure for the automatic selection of γ but one that is not specifically designed to deal with model uncertainties (as is the case with the BDU methods - see, *e.g.,* the simulations in Sec. 7).

3.2. The Total-Least-Squares Method. The total least-squares method, also known as orthogonal regression or errors-in-variables methods in statistics and system identification, has been proposed to combat uncertainties in the data matrix A. Although orthogonal regression methods have been long studied in statistics, apparently starting in the 1870's with a special case in [37], the name total-least-squares (TLS) was coined in the 1980's [17], and the method has since received much attention (see, *e.g.,* [18]).

The TLS method combats uncertainties in A by assuming an erroneous matrix and by trying to estimate what the true A should have been. It can be explained as follows. Assume $A \in \mathbb{R}^{N\times n}$ is a full rank matrix with $N \geq n$, and $b \in \mathbb{R}^N$. Consider the problem of solving the inconsistent linear system $Ax \approx b$, where the symbol $\approx$ is used to signify that $b \notin \mathcal{R}(A)$. The TLS formulation assumes errors in A and seeks an $\hat{x}$ that solves the consistent linear equations $\hat{A}\hat{x} = \hat{b}$, where $\{\hat{A}, \hat{b}\}$ solve [18]:

$$\min_{\hat{A},\ \hat{b}\in\mathcal{R}(\hat{A})} \left\| \begin{bmatrix} A & b \end{bmatrix} - \begin{bmatrix} \hat{A} & \hat{b} \end{bmatrix} \right\|_{\mathrm{F}}^2 . \tag{3.3}$$

The notation $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm of its argument. That is, the TLS method replaces A and b by estimates $\hat{A}$ and $\hat{b}$ with $\hat{b}$ belonging to the range space of $\hat{A}$. It turns out that the $\hat{A}$ and $\hat{b}$ are obtained by projecting A and b, respectively, onto the subspace that is defined by the n dominant singular vectors of the extended matrix $\begin{bmatrix} A & b \end{bmatrix}$ (*i.e.,* by the singular vectors that correspond to the n largest singular values of the matrix) - see Fig. 3.1.

The spectral norm of the correction $(A-\hat{A})$ is determined by the smallest singular value of $\begin{bmatrix} A & b \end{bmatrix}$. This norm can be large even when A is almost precise, *e.g.,* when b is sufficiently far from the column space of A. In this case, the TLS method may end up overly correcting A and unnecessarily replacing it by an $\hat{A}$ far from it, which may lead to a conservative solution. This is a reflection of the fact that in the TLS formulation (3.3) there is no a priori bound on the size of the allowable correction to A – see also some simulation results in Sec. 7.

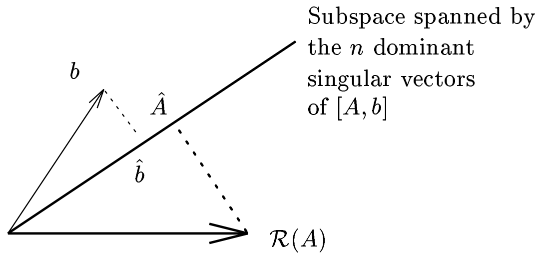

FIG. 3.1. Construction of the TLS solution.

3.3. The $\mathcal{H}_\infty$ or Game-Theoretic Design Method. A design methodology that handles rather successfully the control of a perturbed version of the state-space model (2.6), and which has been receiving considerable attention in the literature, is the $\mathcal{H}_\infty$ or game-theoretic approach (see, *e.g.,* [9, 11, 29] and [38]–[42] and the many references therein). The approach is based on the idea of designing for the worst possible scenario (or model). This is in contrast to the TLS paradigm, where the idea is to first estimate what the true model should have been and then proceed with the design using the estimated model.

We explain the $\mathcal{H}_\infty$ method briefly in the context of the quadratic regulator problem of Sec. 2.4. Detailed treatments can be found in [9, 11]. Here we only wish to highlight the main ideas, and the discussion in this section is in fact not necessary for the understanding of the rest of the paper and can be skipped on a first reading (the reader can go directly to Sec. 4). The results in this section are included for comparison purposes and for readers that might not be familiar with the $\mathcal{H}_\infty$ design methodology.

Returning to the perturbed version of the state-space model (2.6),

$$x_{i+1} = (f + \delta f)x_i + (g + \delta g)u_i \ , \tag{3.4}$$

we first note that the system can be represented in diagram form as shown in Fig. 3.2. The signals w_1 and w_2 denote the perturbations $\{\delta g u_i, \delta f x_i\}$, and the block with z^{-1} indicates a unit time-delay. The block with $K(z)$ indicates the transfer function of a controller that we wish to determine in order to stabilize the closed-loop system (*i.e.*, stabilize the transfer function from col$\{w_1, w_2\}$ to col$\{u_i, x_i\}$ even in the presence of the uncertainties $\{\delta f, \delta g\}$).

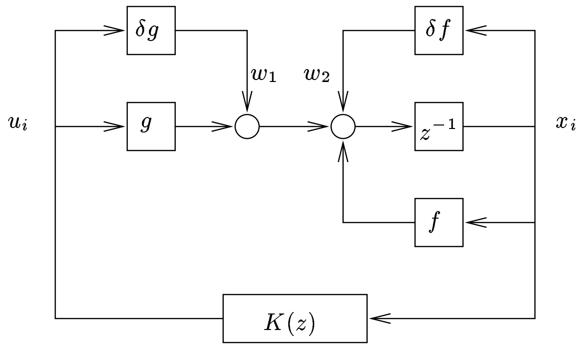

FIG. 3.2. Block diagram representation of the perturbed state-equation.

Define the vector signals

$$w_i \triangleq \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \quad z_i \triangleq \begin{bmatrix} u_i \\ x_i \end{bmatrix}.$$

Here w_i represents the perturbations and z_i contains the signals that we wish to regulate, *viz.*, the state and the control. We can now re-draw the block diagram of Fig. 3.2 in an equivalent form that is standard in the literature on $\mathcal{H}_\infty$ control, as shown in Fig. 3.3. The $P(z)$ denotes the transfer function from the input signals $\{w_i, u_i\}$ to the output signals $\{z_i, y_i\}$, where we are denoting the input of $K(z)$ by y_i (clearly in our problem $y_i = x_i$). The transfer function $\Delta(z)$ represents the mapping that relates z_i to w_i.

It is immediate to verify that in our particular problem, $\Delta(z)$ is diagonal with constant *real* entries and is given by

$$\Delta(z) = \begin{bmatrix} \delta f & 0 \\ 0 & \delta g \end{bmatrix}.$$

Moreover, $P(z)$ has a state-space realization that is given by

$$\begin{cases} x_{i+1} &= f x_i + g u_i + \begin{bmatrix} 1 & 1 \end{bmatrix} w_i \\ z_i &= \begin{bmatrix} 0 \\ 1 \end{bmatrix} x_i + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u_i \\ y_i &= x_i \end{cases}$$

Let $F(z)$ denote the transfer function from the perturbation w_i to the regulated output z_i in the absence of $\Delta(z)$. This transfer function is dependent on $K(z)$ and $P(z)$. The design of a robust stable controller $K(z)$ in an $\mathcal{H}_\infty$ framework is concerned with the problem of determining a stable $K(z)$ that stabilizes the closed-loop system for all possible $\Delta(z)$ of a specified form. [By a stable $K(z)$ we mean one that has poles inside the open unit disc.]

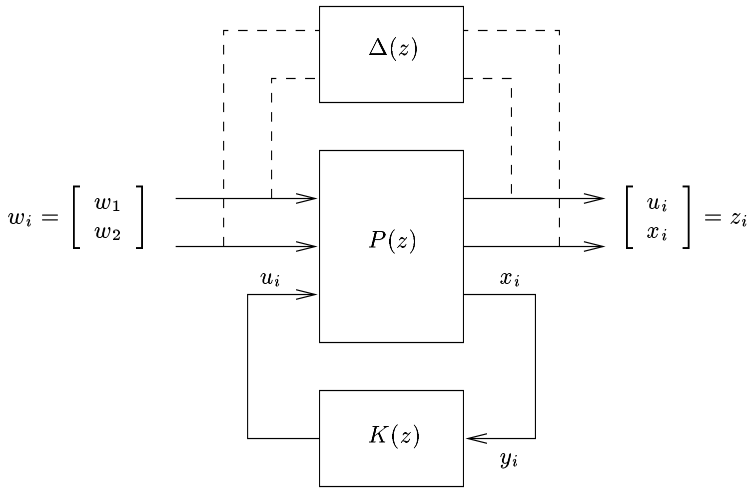

FIG. 3.3. Representation of the perturbed state-equation in standard form.

A powerful tool in this regard is the so-called structured singular value (SSV) of a transfer function [11, 43] (see [44] for a survey and also Ch. 8 of [45] for an overview with several examples). The SSV of the transfer function $F(z)$ is dependent on the structure of $\Delta(z)$. It is denoted by $\mu_\Delta(F)$ and is defined as follows. Let $\|A\|_\infty$ denote the so-called $\mathcal{H}_\infty$ norm of a stable transfer function $A(z)$,

$$\|A\|_\infty \;=\; \sup_{\omega\in[0,2\pi]} \;\; \sigma_{\max}\;\left[A(e^{j\omega})\right] \;,$$

where $\sigma_{\max}$ is the maximum singular value of its argument. To determine $\mu_\Delta(F)$, we find the smallest $\Delta(z)$, say $\Delta^o(z)$, in the allowed class of uncertainties (measured in terms of $\|\Delta\|_\infty$) that makes the closed-loop system unstable. This corresponds to the smallest uncertainty $\Delta(z)$ that makes $\det[I - F(z)\Delta(z)] = 0$. Then

$$\mu_\Delta(F) \;=\; \frac{1}{\|\Delta^o\|_\infty} \;.$$

Using the notion of SSV, a variant of a well-known theorem in system theory, known as the small-gain theorem [11, 45], states that the closed-loop transfer function in Fig. 3.3 is stable for all allowed stable structured $\Delta(z)$ if, and only if, the SSV of $F(z)$ and the $\mathcal{H}_\infty$ norm of $\Delta(z)$ satisfy

$$\mu_\Delta(F)\|\Delta\|_\infty \;<1\;.$$

Hence, a robust control design (also known as $\mu-$synthesis) reduces to determining a controller $K(z)$ that minimizes $\mu_\Delta(F)$ so that the resulting closed-loop system will be stable for the largest class of uncertainties.

It turns out that the computational complexity of computing the SSV of a transfer function $F(z)$ is NP-hard. There is also considerable difference in the effort required when the uncertainty $\Delta(z)$ is real-valued or complex-valued. The former (real-valued case) is considerably more difficult. In the $\mu-$toolbox of Matlab[1], a so-called DK

[1]Matlab©is a registered trademark of The MathWorks Inc.

iteration is used (and a more complex variant for real-valued uncertainties) that minimizes an upper bound for $\mu_\Delta(F)$ rather than minimizing $\mu_\Delta(F)$ itself. Also, most results and algorithms, even those implemented in the Matlab $\mu-$toolbox, are developed almost exclusively for continuous-time systems.

For our perturbed model (3.4), with the nominal values $f = 0.9$ and $g = 1$, and with a diagonal uncertainty $\Delta(z)$ of norm 0.27 (since $\delta f = 0.2$ and $\delta g = -0.27$) we used the $\mu-$toolbox for complex-valued diagonal uncertainties[2] to design a controller for uncertainties as large as $\|\Delta\|_\infty = 0.28$. We found

$$K(z) = -0.6034\,.$$

Simulation results are shown in Fig. 3.4, where the solid line shows the unstable evolution of the state trajectory when LQR control is used. The dotted line shows the evolution of the state vector when $\mathcal{H}_\infty$ control is used. The state of the closed-loop system goes to zero as desired, even in the presence of the uncertainties. The overall cost in this case for $N = 80$ adds up to approximately 71.53.

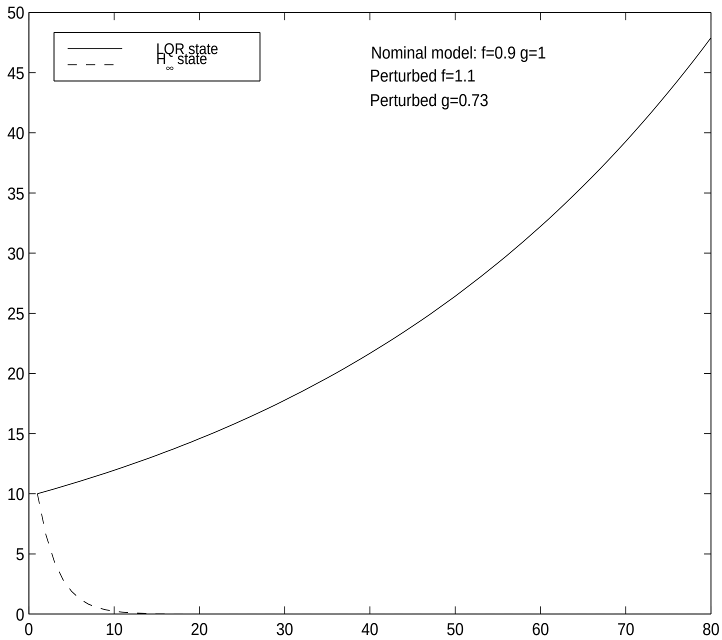

FIG. 3.4. $\mathcal{H}_\infty$ control of the perturbed plant.

The closed-loop pole in steady-state is now located at approximately 0.6595. Using the $\mu-$toolbox, we also determined that the largest $\|\Delta\|_\infty$ for which a stabilizing controller could be found was $\|\Delta\|_\infty = 0.52$ with the corresponding controller being $K(z) = -0.9$ $(= -f/g)$.

We may add that more sophisticated design procedures exist that employ prespecified weighting functions, or even bound certain $\mathcal{H}_\infty$ norms subject to $\mathcal{H}_2$ or quadratic constraints, in order to guarantee some desired levels of performance. These

[2]To the authors' knowledge, a design procedure that deals with the case of real-valued diagonal uncertainties for discrete-time systems is not immediately available in the Matlab $\mu-$toolbox.

schemes are usually more complex, and in some cases not yet fully developed. In this section, we opted to illustrate the $\mathcal{H}_\infty$ procedure in one of its most standard forms without additional constraints.

4. NEW BDU DESIGN CRITERIA. The design techniques reviewed in Sec. 3 combat modeling uncertainties in several ways. In the TLS case, for example, the true model is first estimated, but without any bound on how large the correction $A - \hat{A}$ can be. In the $\mathcal{H}_\infty$ formulation, on the other hand, the design procedure can end up being conservative.

In this section, we study several cost functions for design purposes that explicitly incorporate a-priori bounds on the size of the data uncertainties. In so doing, the resulting solutions guarantee that the effect of the uncertainties will not be unnecessarily over-emphasized beyond what is reasonably assumed by the a-priori bounds. In many cases, we will be able to characterize completely the solutions and provide algorithms for their computation with essentially the same computational effort as standard least-squares solutions, thus making the new formulations attractive for practical use.

We start by reconsidering a problem first formulated in [46, 47, 48] and which was originally fully solved in [47] and via a more costly linear matrix inequality (or convex optimization) technique in [48]. We shall refer to it as a BDU estimation problem (with BDU standing for Bounded Data Uncertainties). In this paper (Sec. 5), we re-solve this problem from a different perspective. Rather than rely on algebraic arguments similar to those in [47, 48], we shall develop a geometric theory for BDU estimation. In particular, we shall extend the famed orthogonality principle of least-squares theory to the context of BDU estimation. In the process of developing the geometric framework, several concepts from linear algebra and matrix theory will play a prominent role.

In addition to the geometric formulation, we shall also motivate and formulate several extensions for BDU estimation and control purposes (see Sec. 4.3 and Secs. 6–7). In these new formulations, we allow for different sources and levels of uncertainties in the model, as well as for more general cost functions. Applications in the context of image processing, co-channel interference cancellation, and quadratic control are discussed.

4.1. The Data Model. We motivate the BDU framework as follows. Let $x \in \mathbb{R}^n$ be a column vector of unknown parameters, $b \in \mathbb{R}^N$ a vector of measurements, and $A \in \mathbb{R}^{N\times n}$, $N \geq n$, a known full rank matrix. The matrix A represents nominal data in the sense that the true matrix that relates b to x is not A itself but rather a perturbed version of A, say

$$b = (A + \delta A)x + v \; . \tag{4.1}$$

The perturbation δA is not known. What is known is a bound on how far the true matrix $(A + \delta A)$ can be from the assumed nominal value A, say

$$\|\delta A\| \leq \eta \; , \tag{4.2}$$

in terms of the 2−induced norm of δA, or equivalently, its maximum singular value. [All the results will apply, and in fact become simpler, if we instead employ the Frobenius norm of δA, say $\|\delta A\|_{\mathrm{F}} \leq \eta$, rather than the 2−induced norm. We shall comment on this point later – see the remark after the proof of Lemma 5.4.]

The standard least-squares criterion (2.2), often used in practical designs, would seek to recover x from b by relying on the available nominal data A, and without taking into account the fact that the true data is not A itself but lies around A within a ball of size η. This is clearly a limitation. On the other hand, the total least-squares criterion (3.3) is aware of possible perturbations in A and tries to replace it with an estimate $\hat{A}$ before seeking to estimate x. It however does not explicitly incorporate the a-priori bound η into its statement. In this way, there is no guarantee that the estimate $\hat{A}$ that it finds will lie within the ball of size η; it may end up being overly corrected.

These difficulties motivate the introduction of new design criteria that explicitly incorporate bounds on the sizes of the perturbations.

4.2. A BDU Cost Function for Estimation. The first cost function we consider is the following [46, 47]:

$$\min_{x} \max_{\|\delta A\| \leq \eta} \| (A + \delta A)\, x - b \| \,. \tag{4.3}$$

This formulation seeks a solution $\hat{x}$ that performs "best" in the worst-possible scenario. It can be regarded as a constrained two-player game problem, with the designer trying to pick an x that minimizes the residual norm while the opponent δA tries to maximize the residual norm. The game problem is constrained since it imposes a limit on how large (or how damaging) the opponent δA can be.

In order to further clarify the meaning of (4.3), note that any value that we pick for x would lead to many residual norms, $\| (A + \delta A)\, x - b\|$, one for each possible δA. We want then to determine the $\hat{x}$ whose maximum residual is the smallest possible. Assume, for illustration purposes, that we only have two choices for x, say x_1 and x_2. In Fig. 4.1 we plot the residual curves as a function of δA, *i.e.,* we plot

$$\|(A + \delta A)x_1 - b\| \quad \text{and} \quad \|(A + \delta A)x_2 - b\| \,,$$

for all δA satisfying $\|\delta A\| \leq \eta$. The dark discs indicate the points at which the residuals attain their maxima. We see from the figure that the maximum residual attained by x_2 is smaller than the maximum residual attained by x_1. Hence, the solution to the BDU estimation problem in this case is $\hat{x} = x_2$.

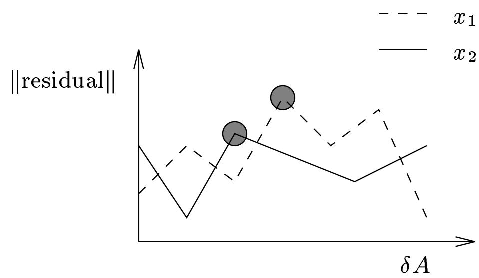

FIG. 4.1. Two illustrative residual-norm curves.

It turns out that the solution of (4.3) has an interesting and powerful geometric interpretation that resembles the orthogonality condition of least-squares problems. Before establishing this result, and before studying its implications, we list in the next section several other cost functions and elaborate on their significance. In a later section, we shall reconsider some of these newer costs and apply them to problems in image processing, co-channel interference cancellation, and control design.

4.3. More General BDU Cost Functions. We start by noting that in some applications, we might be uncertain only about part of the data matrix while the remaining data is known exactly. This motivates us to formulate a BDU problem with partial uncertainties as follows [47]:

$$\min_x \left(\max_{\|\delta A_2\| \leq \eta_2} \left\| \begin{bmatrix} A_1 & A_2 + \delta A_2 \end{bmatrix} x - b \right\| \right) . \tag{4.4}$$

We can also handle situations with different levels of uncertainties in different parts of the data matrix by introducing a BDU problem with multiple uncertainties [49],

$$\min_x \max_{\begin{pmatrix} \|\delta A_j\| \leq \eta_j \\ 1 \leq j \leq K \end{pmatrix}} \left\| \begin{bmatrix} A_1 + \delta A_1 & \ldots & A_K + \delta A_K \end{bmatrix} x - b \right\| . \tag{4.5}$$

Here, the $\{A_j\}$ denote submatrices (column-wise) of A. Such cost functions are useful for multiple-user or multiple-experiment environments and will be applied to image restoration and co-channel interference later in Sec. 7.

Another useful cost function is a BDU formulation for multi-state or discrete-event case, where the uncertainty δA can be only one of a finite number of possibilities, *viz.*,

$$\min_x \left(\max_{\delta A \in \{\delta A_1, \ldots, \delta A_L\}} \|(A + \delta A)x - b)\| \right) . \tag{4.6}$$

This cost is useful for estimation purposes in multi-state environments where only the discrete models $\{A + \delta A_1, A + \delta A_2, \ldots, A + \delta A_L\}$ are possible.

For control purposes, we find it useful to introduce the following two BDU formulations [50]

$$\min_x \left(\max_{\|\delta A\| \leq \eta,\ \|\delta b\| \leq \beta} \|\,(A + \delta A)\,x - (b + \delta b)\|^2 \ + \ \rho \|x\|^2 \right) , \tag{4.7}$$

and

$$\min_x \max_{\|\delta A\| \leq \eta,\ \|\delta b\| \leq \beta} [(A + \delta A)x - (b + \delta b)]^T W [(A + \delta A)x - (b + \delta b)] , \tag{4.8}$$

where we now allow for uncertainties in A and b, and also employ weighting factors W and ρ (with W a matrix and ρ a scalar). We shall demonstrate an application of (4.7) later in Sec. 6.

We can also formulate cost functions that treat data uncertainties multiplicatively rather than additively,

$$\min_x \max_{\|\delta A\| \leq \eta} \|(I + \delta A)Ax - b\| , \tag{4.9}$$

and also treat weight uncertainties,

$$\min_x \max_{\|\delta W\| \leq \eta_w} \|(W + \delta W)(Ax - b)\| . \tag{4.10}$$

This later cost is a variation of the weighted least-squares criterion in which the uncertainty is taken to be in the weight matrix itself. Such situations arise, in more structured forms, in Kalman filtering theory where the noise covariance matrices play the role of weighting matrices. But since these covariance matrices are not always known a priori, they need to be estimated before applying the Kalman filter equations (*e.g.,* [34, 35]). In this way, we may end up employing perturbed weight matrices. The above cost then seeks an $\hat{x}$ that performs best in the face of the worst possible choice for the weight matrix.

We can as well consider BDU formulations with an average (or stochastic) performance index, *e.g.,*

$$\min_x \ \mathrm{avg}_{\|\delta A\| \leq \eta} \ \| (A + \delta A)\, x - b \| \ ,$$

where "avg" denotes symbolically some notion of average, or some alternative cost functions with stochastic assumptions on both x and δA. Such stochastic extensions will be discussed elsewhere.

5. BDU ESTIMATION.

We now consider the BDU formulation (4.3),

$$\min_x \ \max_{\|\delta A\| \leq \eta} \ \| (A + \delta A)\, x - b \| \ ,$$

and study it in some detail. As mentioned earlier, this cost function was studied in [47] and the solution was found there algebraically. Here we shall re-solve the problem from a different perspective. Rather than rely on algebraic arguments, we shall develop a geometric theory for BDU estimation. In particular, we shall extend the orthogonality principle of least-squares theory to this context. Several concepts from linear algebra and linear vector spaces will play a prominent role in our arguments.

For ease of exposition, and in order to avoid degenerate cases, we shall assume in this paper that A is full rank and that b does not lie in the range space of A (and is nonzero),

$$\mathrm{rank}(A) = n \quad \text{and} \quad b \notin \mathcal{R}(A) \ . \tag{5.1}$$

These conditions rule out the case of a square invertible matrix A and therefore require $N > n$. However, if these conditions do not hold, then the solution is only slightly more complex. The geometric arguments given below can still be extended but, for the sake of clarity and in order to emphasize the main ideas, we shall assume that conditions (5.1) hold. We shall pursue the geometry of the degenerate case elsewhere – see though [47] for a statement of the solution in the degenerate case.

5.1. The Uncertainty Set.

In the BDU formulation (4.3), the perturbation δA is restricted to the ball $\|\delta A\| \leq \eta$. It turns out that the solution of the problem will depend on a crucial property of this ball, *viz.,* whether there exists a δA such that the perturbed matrix $(A + \delta A)$ becomes orthogonal to b. The following result establishes when this is possible. Later we shall see why this property is crucial to the solution of (4.3).

LEMMA 5.1 (A bound on η). *The uncertainty set $\{\|\delta A\| \leq \eta\}$ contains a perturbation δA such that $(A + \delta A)^T b = 0$ if, and only if,*

$$\eta \ \geq \frac{\|A^T b\|}{\|b\|} \tag{5.2}$$

Proof: Assume there exists a δA, say $\overline{\delta A}$, such that $(A+\overline{\delta A})^T b = 0$. Then $(\overline{\delta A})^T b = -A^T b$ and

$$||A^T b|| = ||\overline{\delta A}^T b|| \;\leq\; ||\overline{\delta A}^T|| \cdot ||b||$$

This implies that $||\overline{\delta A}|| \geq ||A^T b||/||b||$ and, hence, condition (5.2) must hold.

Conversely, assume (5.2) holds and choose

$$\overline{\delta A} = -\frac{1}{||b||^2}\, bb^T A \;.$$

Then

$$||\overline{\delta A}|| \;\leq\; \frac{1}{||b||^2}\,||b||\;||b^T A|| \;=\; \frac{||A^T b||}{||b||} \;\leq\; \eta\;.$$

This shows that $\overline{\delta A}$ is a valid perturbation. Now note that

$$A+\overline{\delta A} = A - \frac{1}{||b||^2}\, bb^T A = \left[I - \frac{bb^T}{||b||^2}\right] A\;,$$

where the matrix $(I - bb^T/||b||^2)$ is the projector onto the orthogonal complement space of b. This implies that $(A+\overline{\delta A})^T b = 0$, as desired.

$\diamondsuit$

Fig. 5.1 is a pictorial representation of the case $\eta < ||A^T b||/||b||$. The uncertainty set is indicated by the dashed area and it is seen not to include a perturbed matrix that is orthogonal to b.

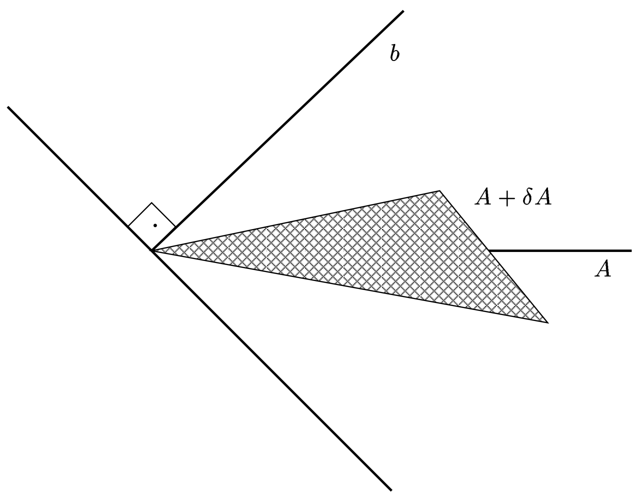

FIG. 5.1. A depiction of the case $\eta < \frac{||A^T b||}{||b||}$.

In the sequel we shall establish that when the uncertainty set is large enough to include a perturbed matrix $(A+\delta A)$ that is orthogonal to b, then the unique solution of (4.3) is $\hat{x} = 0$. Otherwise, the solution is nonzero and has an interesting regularized form.

5.2. Unique Zero Solution. To verify the above claim, we start by considering the case $\eta \geq \|A^T b\|/\|b\|$. We shall show that the solution of (4.3) will be unique and equal to zero, $\hat{x} = 0$. The result makes sense since intuitively it means that in the presence of relatively large uncertainty in the data, the true matrix $(A + \delta A)$ could be orthogonal to the measured vector b, in which case there is no information to extract about x and the best estimate for x would be $\hat{x} = 0$.

To establish the result, we first note that if we set x equal to zero in the BDU cost function (4.3), we obtain that the residual norm is always equal to $\|b\|$ regardless of δA. We now show that when (5.2) holds, for any nonzero x, we can always find a perturbation δA satisfying $\|\delta A\| \leq \eta$ such that the residual norm $\|(A + \delta A)x - b\|$ is strictly larger than $\|b\|$. This would mean that for any nonzero x, the maximum residual $\|(A + \delta A)x - b\|$ over $\{\|\delta A\| \leq \eta\}$ has to be larger than b, so that $\hat{x} = 0$ has to be the unique minimum solution since it leads to the smallest residual norm. In fact, we have a stronger statement.

LEMMA 5.2 (Zero solution vector). *The* BDU *estimation problem (4.3) has a unique solution at $\hat{x} = 0$ if, and only if, (5.2) holds.*

Proof: Assume first that (5.2) holds and let us show that $\hat{x} = 0$ is the unique solution. Choose

$$\overline{\delta A} = -\frac{1}{\|b\|^2}\, bb^T A \ .$$

We already know from the proof of Lemma 5.1 that $\overline{\delta A}$ is a valid perturbation since $\|\overline{\delta A}\| \leq \eta$, and that $(A + \overline{\delta A})^T b = 0$. We now further show that $(A + \overline{\delta A})$ has full column rank. Assume otherwise, then there should exist a nonzero vector p such that

$$\left[I - \frac{bb^T}{\|b\|^2} \right] Ap = 0 \ .$$

If we denote Ap by q (q is also nonzero since A is full column rank), this means that we must have

$$\left[I - \frac{bb^T}{\|b\|^2} \right] q = 0 \ ,$$

which is only possible if q is parallel to the vector b, say $q = \alpha b$ for some $\alpha \neq 0$, since the matrix $(I - bb^T/\|b\|^2)$ is the projector onto the orthogonal complement space of b. Hence, we must have $Ap = \alpha b$. This contradicts our assumption that b does not lie in the column span of A. Therefore, the matrix $(A + \overline{\delta A})$ has full column rank.

Now since b is orthogonal to $(A + \overline{\delta A})$, it follows that

$$\|(A + \overline{\delta A})x - b\| \ > \ \|b\| \ ,$$

for any nonzero vector x. Hence, the smallest residual over x is $\|b\|$ and it is attained at $x = 0$. All other nonzero choices for x would lead to a larger residual. We can now write, for any nonzero x,

$$\max_{\|\delta A\| \leq \eta} \ \|(A + \delta A)x - b\| \ \geq \ \|(A + \overline{\delta A})x - b\| \ > \ \|b\| \ .$$

This shows that $\hat{x} = 0$ is the unique solution of (4.3).

Conversely, assume $\hat{x} = 0$ is the unique solution of (4.3) and let us establish that (5.2) must hold. Indeed, if $\hat{x} = 0$ is the unique solution then for every x,

$$\max_{\|\delta A\| \leq \eta} \|(A + \delta A)x - b\|^2 \ \geq \ \|b\|^2 \ .$$

That is, for every x,

$$\max_{\|\delta A\| \leq \eta} \left[x^T(A + \delta A)^T(A + \delta A)x - 2b^T(A + \delta A)x\right] \ \geq \ 0 \ .$$

Choose x as a scaled multiple of A^Tb, say $x = \gamma A^T b$. Then the above inequality implies that for any γ,

$$\max_{\|\delta A\| \leq \eta} \left[\gamma^2 b^T A(A + \delta A)^T(A + \delta A)A^T b - 2\gamma b^T(A + \delta A)A^T b\right] \ \geq \ 0 \ .$$

We now claim that for the above inequality to hold, it must be true that

$$\max_{\|\delta A\| \leq \eta} \left[-2b^T(A + \delta A)A^T b\right] \ \geq \ 0 \ . \tag{5.3}$$

Indeed, assume not, say

$$\max_{\|\delta A\| \leq \eta} \left[-2b^T(A + \delta A)A^T b\right] \ = \ -\rho \ < 0 \ ,$$

for some $\rho > 0$. Then

$$2b^T(A + \delta A)A^T b \geq \rho > 0 \ ,$$

for all δA. Choose γ such that

$$\gamma \ < \ \frac{\rho}{\|b\|^2 \|A\|^2 (\|A\| + \eta)^2} \ .$$

Then it is easy to verify that

$$\max_{\|\delta A\| \leq \eta} \left[\gamma^2 b^T A(A + \delta A)^T(A + \delta A)A^T b - 2\gamma b^T(A + \delta A)A^T b\right] \ < \ 0 \ ,$$

which is a contradiction. Therefore, (5.3) must hold, *i.e.,*

$$\max_{\|\delta A\| \leq \eta} \left[-2b^T A A^T b - 2b^T \delta A A^T b\right] \ \geq \ 0 \ . \tag{5.4}$$

The maximum the expression between parenthesis can be is

$$-2b^T A A^T b + 2\eta \|b\| \cdot \|A^T b\| \ .$$

This maximum is in fact achievable. Indeed, choose

$$\delta A = -\eta \ \frac{bb^T A}{\|b\| \ \|b^T A\|} \ .$$

Then δA is a valid perturbation since $\|\delta A\| = \eta$ and it achieves the above value. Therefore, we must have

$$-2\|A^T b\|^2 + 2\eta \|b\| \ \|A^T b\| \ \geq 0 \ ,$$

which leads to the desired conclusion (5.2).

$\diamondsuit$

5.3. Worst-Case Perturbations. We have therefore identified completely the condition under which $\hat{x} = 0$ is the unique solution, *viz.*, when (5.2) holds. We now assume to the contrary that

$$\eta < \frac{\|A^T b\|}{\|b\|} \; . \tag{5.5}$$

Hence, if the problem has a solution $\hat{x}$ then it has to be nonzero. In fact, we shall show that a unique *nonzero* solution exists. For this reason, we shall focus in the sequel on nonzero vectors x,

$$x \neq 0 \; ,$$

and determine conditions for a nonzero x to be a solution.

Returning to (4.3), we shall first identify the perturbations δA that maximize the residual norm. To begin with, note that in view of the triangle inequality of norms, it holds that for any δA,

$$\|\,(A + \delta A)\,x - b)\| \;\leq\; \|Ax - b\| + \|\delta A x\| \; ,$$

with equality if, and only if, the perturbation δA is such that the vector $\delta A x$ is collinear with the vector $(Ax - b)$, *i.e.*,

$$\delta A x = \beta(Ax - b) \; , \tag{5.6}$$

for some scalar $\beta \geq 0$. Moreover, $\|\delta A x\| \leq \eta \|x\|$ with equality if, and only if, the perturbation δA is also such that

$$\|\delta A x\| \;=\; \eta \|x\| \; . \tag{5.7}$$

Combining (5.7) with (5.6), we see that (5.7) will hold only if

$$\beta = \eta \, \frac{\|x\|}{\|Ax - b\|} \; .$$

This expression for β is well-defined since $\|Ax - b\| \neq 0$ in view of our earlier assumption that b does not lie in the column span of A.

The above discussion shows that if for a nonzero x there exists a perturbation δA in the valid domain $\|\delta A\| \leq \eta$ that satisfies

$$\delta A x = \eta \, \frac{\|x\|}{\|Ax - b\|} \, (Ax - b) \; , \tag{5.8}$$

then

$$\max_{\|\delta A\| \leq \eta} \; \|(A + \delta A)x - b\| \;=\; \|Ax - b\| + \eta \|x\| \; . \tag{5.9}$$

It also follows from (5.6) that any δA that attains the maximum residual in (5.9) leads to a residual vector $(A + \delta A)x - b$ that is necessarily collinear with $(Ax - b)$, since

$$(A + \delta A)x - b = (1 + \beta)(Ax - b) \; . \tag{5.10}$$

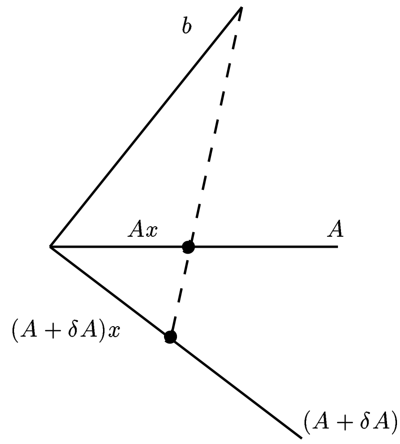

FIG. 5.2. The residual vectors $[(A+\delta A)x-b]$ and $Ax-b$ are collinear for any δA that attains the maximum residual norm.

Therefore, for any nonzero x, the vectors connecting b to Ax, in the column span of A, and to $(A+\delta A)x$, in the column span of $(A+\delta A)x$, are collinear and point in the same direction. This is depicted in Fig. 5.2 and summarized below.

LEMMA 5.3 (Direction of residual vectors). *For any nonzero x, if δA is a perturbation that achieves the maximum residual norm in (5.9), viz., $\|Ax-b\|+\eta\|x\|$, then the residual vectors $(A+\delta A)x-b$ and $Ax-b$ are collinear. They also point in the same direction (i.e., one is a positive multiple of the other).*

It is easy to verify that the following choice for δA satisfies (5.8) and has norm equal to η,

$$\delta A^o(x) \triangleq \eta \frac{(Ax-b)}{\|Ax-b\|} \frac{x^T}{\|x\|} .$$

Note that δA^o is a function of x. [Often, we shall drop the argument x and write simply δA^o for compactness of notation.] Therefore, the maximum residual in (5.9) is attainable. It is enough for our purposes to identify one of the perturbations δA that achieve the maximum residual, *e.g.,* the δA^o above.

We remark, however, that in general there can exist many other $\delta A's$ that satisfy (5.8) for any given x. The following statement parametrizes all such perturbations. The result holds for any η, regardless of (5.2) or (5.5). Although we shall not use the next two results in the sequel, we include them here for completeness.

LEMMA 5.4 (Worst-case perturbation). *For any given nonzero x, and for any η, the perturbations that satisfy (5.8), and therefore attain the maximum residual in (5.9), occur on the boundary $\|\delta A\|=\eta$ and they are fully parametrized by the following expression*

$$\delta A = \delta A^o + Y\left(I - \frac{xx^T}{\|x\|^2}\right) , \tag{5.11}$$

for any matrix $Y \in \mathbb{R}^{N\times n}$ *that satisfies*

$$\left(I - \frac{xx^T}{\|x\|^2}\right) Y^T(Ax-b) = 0 \quad \textit{and} \quad \left\| Y\left(I - \frac{xx^T}{\|x\|^2}\right)\right\| \leq \eta \ , \tag{5.12}$$

and where

$$\delta A^o \triangleq \eta \frac{(Ax-b)}{\|Ax-b\|} \ \frac{x^T}{\|x\|} \ .$$

Proof: Before we begin the proof let us first note that

$$\left(I - \frac{xx^T}{\|x\|^2}\right)^2 = \left(I - \frac{xx^T}{\|x\|^2}\right) \ ,$$

since it is the projection matrix onto the orthogonal complement space of x.

Now any perturbation δA that satisfies (5.8) must be such that (5.7) holds. This implies that $\|\delta A\| \geq \eta$, which in view of the restriction $\|\delta A\| \leq \eta$ means that any such δA has to lie on the boundary $\|\delta A\| = \eta$. This establishes part of the Lemma.

To establish the parametrization of all valid $\delta A's$, we note that for any given nonzero x, δA has to be the solution of the under-determined linear system of equations (5.8). It is well-known that all solutions are given by

$$\delta A \ = \ \frac{\eta\|x\|}{\|Ax-b\|} \, (Ax-b) \, x^\dagger \ + \ Y\left(I - xx^\dagger\right) \ , \tag{5.13}$$

for an arbitrary matrix Y and where $x^\dagger$ denotes the pseudo-inverse of x,

$$x^\dagger = \frac{x^T}{\|x\|^2} \ ,$$

and where the two terms

$$\frac{\eta\|x\|}{\|Ax-b\|} \, (Ax-b) \, x^\dagger \quad \text{and} \quad Y\left(I - xx^\dagger\right)$$

are orthogonal, *i.e.,*

$$\left(\frac{\eta\|x\|}{\|Ax-b\|} \, (Ax-b) \, x^\dagger\right)^T \left(Y\left[I - xx^\dagger\right]\right) = 0 \ .$$

The expression for $x^\dagger$ leads to

$$\begin{aligned} \delta A &= \eta \frac{(Ax-b)}{\|Ax-b\|} \ \frac{x^T}{\|x\|} \ + \ Y\left(I - \frac{xx^T}{\|x\|^2}\right) \ , \\ &= \delta A^o \ + \ Y\left(I - \frac{xx^T}{\|x\|^2}\right) \ , \end{aligned}$$

and in view of the above orthogonality,

$$\begin{aligned} \delta A \delta A^T &= \delta A^o(\delta A^o)^T \ + \ Y\left(I - \frac{xx^T}{\|x\|^2}\right) Y^T \\ &= \frac{\eta^2}{\|Ax-b\|^2}(Ax-b)(Ax-b)^T + Y\left(I - \frac{xx^T}{\|x\|^2}\right) Y^T \ . \end{aligned} \tag{5.14}$$

But recall that δA has to satisfy $\|\delta A\| = \eta$. This imposes restrictions on Y. We shall now verify that the requirement $\|\delta A\| = \eta$ is satisfied if, and only if, Y is chosen according to (5.12).

Indeed, assume first that (5.12) holds. Using $(I - xx^T/\|x\|^2)Y^T(Ax-b) = 0$, we easily conclude from the above that

$$\delta A \delta A^T (Ax - b) = \eta^2 (Ax - b) \; .$$

This shows that the vector $(Ax - b)$ is an eigenvector for $\delta A \delta A^T$ with eigenvalue η^2. Let us verify that all other eigenvalues of $\delta A \delta A^T$ have to be smaller or equal to η^2 so that we conclude that $\|\delta A\| = \eta$. Note first that $\delta A \delta A^T$ is a symmetric nonnegative definite matrix and, hence, all its eigenvalues are non-negative. It also has orthogonal eigenvectors. Let w denote any of the unit-norm eigenvectors of $\delta A \delta A^T$ that is orthogonal to the eigenvector $(Ax-b)$. Let λ^2 be the corresponding eigenvalue. Then from (5.14),

$$w^T \delta A \delta A^T w = \lambda^2 w^T w = \lambda^2 = w^T Y \left(I - \frac{xx^T}{\|x\|^2} \right) Y^T w \; .$$

Using the second condition on Y in (5.12) we see that we must have

$$\lambda^2 \leq \left\| Y \left(I - \frac{xx^T}{\|x\|^2} \right) \right\|^2 \; \leq \; \eta^2 \; ,$$

and therefore $\lambda \leq \eta$, as desired. We thus showed that if conditions (5.12) hold, then any δA in (5.13) satisfies $\|\delta A\| = \eta$.

Conversely, let us show that if Y is not chosen according to (5.12) then $\|\delta A\| > \eta$. So assume that one of the conditions fails, say $(I - xx^T/\|x\|^2)Y^T(Ax - b) \neq 0$. It then follows from (5.14) that

$$\begin{aligned} (Ax-b)^T \delta A \delta A^T (Ax-b) &= \eta^2 \|Ax - b\|^2 \; + \; (Ax-b)^T Y \left(I - \frac{xx^T}{\|x\|^2} \right) Y^T (Ax - b) \\ &> \eta^2 \|Ax - b\|^2 \; . \end{aligned}$$

Therefore, $\|\delta A\| > \eta$. Assume, on the other hand, that the second condition in (5.12) fails. Hence, there exists a nonzero vector q such that

$$q^T Y \left(I - \frac{xx^T}{\|x\|^2} \right) Y^T q \; > \eta^2 \|q\|^2 \; .$$

Then using (5.14) again we obtain

$$q^T \delta A \delta A^T q = \frac{\eta^2}{\|Ax - b\|^2} q^T (Ax - b)(Ax - b)^T q + q^T Y \left(I - \frac{xx^T}{\|x\|^2} \right) Y^T q > \eta^2 \|q\|^2 \; ,$$

which again implies that $\|\delta A\| \; > \; \eta$.

$\diamondsuit$

Remark 1. Had we originally stated the BDU problem (4.3) with the constraint $\|\delta A\|_{\mathrm{F}} \leq \eta$, in terms of the Frobenius norm of δA rather than its 2−induced norm, then all the arguments in this section will still apply except that the δA that achieves the maximum residual in (5.9) would be unique and equal to δA^o! This is because it is known, and can be easily verified, that the solution δA of (5.13) that corresponds to the choice $Y = 0$ has the smallest Frobenius norm. Hence, $\|\delta A^o\|_F = \eta$ and all other $\delta A's$ solving (5.13) would have $\|\delta A\|_F > \eta$.

$\diamondsuit$

Remark 2. Note also that the parametrization in Lemma 5.4 is not unique. Two distinct matrices Y can result in the same δA. For example, given a Y, we can add to it any matrix that is orthogonal to $(I - xx^T/\|x\|^2)$ and obtain the same δA. This nonuniqueness is not relevant to our future discussions.

$\diamondsuit$

Lemma 5.4 parametrizes all possible δA's that attain the maximum residual in terms of δA^o. The following result now follows immediately.

COROLLARY 5.5. *For any given nonzero x, and for any η, if δA_1 and δA_2 are two perturbations that satisfy (5.8), and therefore attain the maximum residual in (5.9), then there exists a $Y \in \mathbb{R}^{N\times n}$ satisfying*

$$\left(I - \frac{xx^T}{\|x\|^2}\right) Y^T(Ax - b) = 0 \ , \tag{5.15}$$

and such that

$$\delta A_1 = \delta A_2 + Y\left(I - \frac{xx^T}{\|x\|^2}\right) \ .$$

Proof: For δA_1 we can write

$$\delta A_1 = \delta A^o + Y_1\left(I - \frac{xx^T}{\|x\|^2}\right) \ ,$$

for some Y_1 that satisfies (5.12). Likewise, for δA_2 we can write

$$\delta A_2 = \delta A^o + Y_2\left(I - \frac{xx^T}{\|x\|^2}\right) \ ,$$

for some Y_2 that satisfies (5.12). Subtracting we obtain

$$\delta A_2 = \delta A_1 + (Y_2 - Y_1)\left(I - \frac{xx^T}{\|x\|^2}\right) \ ,$$

where it is easy to see that the difference $(Y_2 - Y_1)$ satisfies (5.15).

$\diamondsuit$

It also turns out that for any nonzero x, the worst-case perturbation δA^o has a very useful property. Recall that A is full column rank by assumption. Now we have the following.

LEMMA 5.6 (Full rank property). *For any nonzero x, the worst-case perturbation δA^o is such that*

$$A + \delta A^o \quad \textit{is still full column rank} \, .$$

Proof: Assume $A + \delta A^o$ is rank deficient. This means that there exists a nonzero vector p such that

$$\left[A + \eta \frac{(Ax-b)}{\|Ax-b\|} \ \frac{x^T}{\|x\|}\right] p = 0 \ .$$

The vector p is necessarily not orthogonal to x since otherwise we would obtain $Ap = 0$, which contradicts our earlier assumption that A itself has full column rank. Define the scalar nonzero quantity

$$\beta = \frac{\eta(x^T p)}{\|Ax-b\| \ \|x\|}$$

It then follows from the above equality that

$$A\left[x + \frac{1}{\beta}p\right] = b \ .$$

This means that b should lie in the column span of A, which again contradicts our earlier assumption about b.

$\diamondsuit$

5.4. The Orthogonality Condition. Once the maximum residual norm over δA, or equivalently the worst-case perturbations, have been identified, we are reduced to studying the equivalent problem

$$\min_x \ (\|Ax-b\| + \eta\|x\|) \ . \tag{5.16}$$

[Note that the expression for the maximum residual over the set $\{\|\delta A\| \leq \eta\}$ also holds for $x = 0$.]

The formulation (5.16) is deceptively similar, but significantly distinct, from the regularized least-squares formulation (3.1), where the *squared* Euclidean norms $\{\|x\|^2, \|Ax-b\|^2\}$ are used rather than the norms themselves!

Lemma 5.7 (Existence of nonzero solutions). *Assume condition (5.5) holds. Then a nonzero solution $\hat{x}$ of (4.3) should exist.*

Proof: The equivalence between both problems (4.3) and (5.16) holds for all x. Now the cost function in (5.16) is convex in x, which means that at least one global minimum is guaranteed to exist. When $\eta < \|A^T b\|/\|b\|$, we already know that $\hat{x} = 0$ can not be a global minimum. Therefore, at least one nonzero global minimum must exist.

$\diamondsuit$

To make the connection with least-squares theory more explicit, we shall rewrite the BDU estimation problem alternatively as

$$\min_{x \neq 0} \ \| \, [A + \delta A^o]\, x - b \, \| \ , \tag{5.17}$$

where, using the result of Lemma 5.4, we know that the perturbed matrix $(A + \delta A^o)$ attains the maximum residual norm. [We shall explain later – see Remark 3 further ahead – that we could have used any of the perturbations that result in the maximum norm residual for x, and not just δA^o. The conclusions would be the same.]

For compactness of notation we shall denote the worst-case perturbed matrix used in (5.17) by $A(x)$,

$$A(x) = A + \delta A^o \ = \ A + \eta \frac{(Ax - b)}{\|Ax - b\|} \ \frac{x^T}{\|x\|} \ ,$$

then we can write (5.17) more compactly as

$$\min_{x \neq 0} \ \|A(x)x - b\| \ . \tag{5.18}$$

This statement looks similar to the least-squares problem (2.2) with two important distinctions. First, the coefficient matrix A is replaced by a perturbed version of it, $A + \delta A(x)$ and, secondly, the new coefficient matrix is dependent on the unknown x as well. Hence, what we have is a nonlinear least-squares problem with a special form for the coefficient matrix $A(x)$. If $A(x)$ were a constant matrix, and therefore not dependent on x, say $\bar{A}$, then we know from the geometry of least-squares estimation that the solution $\hat{x}$ is obtained by imposing the orthogonality condition (or normal equations – recall (2.4))

$$\bar{A}^T(\bar{A}\hat{x} - b) = 0 \ .$$

In the BDU case, the coefficient matrix is a nonlinear function of x. Interestingly enough, however, it turns out that the solution $\hat{x}$ can still be completely characterized by the same orthogonality condition, with $\bar{A}$ replaced by $A(\hat{x})$ (see Fig. 5.3),

$$A^T(\hat{x}) \left[A(\hat{x})\hat{x} - b\right] = 0 \ . \tag{5.19}$$

Since, from (5.6), the residual vector $A(\hat{x})\hat{x} - b$ is collinear with $A\hat{x} - b$, we obtain the equivalent condition

$$A^T(\hat{x}) \left[A\hat{x} - b\right] = 0 \ .$$

or, equivalently,

$$\left[A + \eta \frac{(A\hat{x} - b)}{\|A\hat{x} - b\|} \ \frac{\hat{x}^T}{\|\hat{x}\|}\right]^T \left[A\hat{x} - b\right] = 0 \ , \tag{5.20}$$

or even more compactly,

$$\left[A + \delta A^o(\hat{x})\right]^T (A\hat{x} - b) = 0 \ .$$

Compared with least-squares theory, we may interpret the result (5.20) as an oblique projection onto A rather than an orthogonal projection. The orthogonality conditions (5.19)–(5.20) do not hold for all nonlinear least-squares problems, *i.e.,* for more general nonlinear functions $A(x)$. They hold for the particular $A(x)$ that arises in the BDU context. We now establish the above claims.

THEOREM 5.8 (Orthogonality of residual vector). *Assume (5.5) holds. Then a nonzero vector $\hat{x}$ is a solution of (5.16) or equivalently (5.18) if, and only if, the*

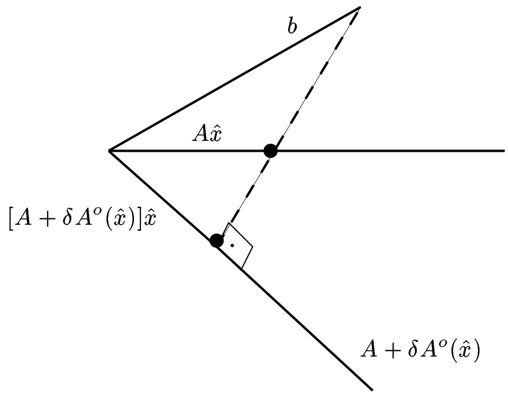

FIG. 5.3. Orthogonality condition for BDU estimation.

residual vector $A\hat{x}-b$ is orthogonal to the following rank-one modification of the data matrix A,

$$A(\hat{x}) = A + \eta \frac{(A\hat{x}-b)}{\|A\hat{x}-b\|} \; \frac{\hat{x}^T}{\|\hat{x}\|} \; . \tag{5.21}$$

That is, if and only if either (5.19) or (5.20) hold.

Proof: Let $\hat{x}$ be a nonzero vector that satisfies the orthogonality condition

$$A^T(\hat{x})[A\hat{x}-b] = 0 \; .$$

Let us show that it has to be a minimizer of the cost function in (5.16). Indeed, pick any other vector x. Then we necessarily have

$$\|A(x)x-b\| \;\; \geq \;\; \|A(\hat{x})x-b\| \; .$$

This is because we know from Lemma 5.4 that for a given x, $A(x)$ is a matrix that maximizes $\|(A+\delta A)x-b\|$. We now verify that

$$\|A(\hat{x})\hat{x}-b\| \; \leq \; \|A(\hat{x})x-b\| \; ,$$

in order to conclude that

$$\|A(\hat{x})\hat{x}-b\| \; \leq \; \|A(x)x-b\| \; ,$$

so that $\hat{x}$ is a minimizer. To establish this fact, we perform the following calculations:

$$\begin{aligned} \|A(\hat{x})x-b\|^2 &= \|A(\hat{x})(x+\hat{x}-\hat{x})-b\|^2 \\ &= \|[A(\hat{x})\hat{x}-b] \; + \; A(\hat{x})(x-\hat{x})\|^2 \\ &= \|A(\hat{x})\hat{x}-b\|^2 \; + \; \|A(\hat{x})(x-\hat{x})\|^2 \\ &\geq \|A(\hat{x})\hat{x}-b\|^2 \; , \end{aligned}$$

where in the third step we used the fact that $A^T(\hat{x})[A(\hat{x})\hat{x}-b]=0$. We thus established that if $\hat{x}$ satisfies the orthogonality condition (5.21) then $\|A(\hat{x})\hat{x}-b\| \leq \|A(x)x-b\|$ for any nonzero x and, therefore, $\hat{x}$ is a minimizer.

Conversely, suppose that $\hat{x}$ is a nonzero minimizer of the cost function in (5.16). The gradient of the cost function with respect to x has to evaluate to zero at $x = \hat{x}$. This leads directly to the condition $A^T(\hat{x})[A\hat{x} - b] = 0$.

$\diamondsuit$

Remark 3. In view of the parameterization (5.11), and of the first condition on Y in (5.12), it is easy to verify that the orthogonality condition holds for any δA that achieves the maximum residual at $\hat{x}$, *i.e.*,

$$\left[A + \eta \frac{A\hat{x} - b}{\|A\hat{x} - b\|} \frac{\hat{x}^T}{\|\hat{x}\|} + Y\left(I - \frac{xx^T}{\|x\|^2}\right)\right]^T [A\hat{x} - b] = 0 \; .$$

We can now establish uniqueness of the solution.

LEMMA 5.9 (Uniqueness of solution). *Assume (5.5) holds, then there exists a unique nonzero solution $\hat{x}$ of (4.3) or, equivalently, a unique nonzero vector $\hat{x}$ that solves the nonlinear equation*

$$\left(A + \eta \frac{(A\hat{x} - b)}{\|A\hat{x} - b\|} \frac{\hat{x}^T}{\|\hat{x}\|}\right)^T [A\hat{x} - b] \;=\; 0 \; . \tag{5.22}$$

Proof: We already know from Lemma 5.7 that a nonzero solution $\hat{x}$ exists. Now assume $\hat{x}_1$ and $\hat{x}_2$ are two distinct nonzero solutions. Then

$$\|A(\hat{x}_1)\hat{x}_1 - b\| \;\geq\; \|A(\hat{x}_2)\hat{x}_1 - b\| \; ,$$

This is because we know from Lemma 5.4 that $A(\hat{x}_1)$ is a matrix that maximizes $\|(A + \delta A)\hat{x}_1 - b\|$. We claim that since both $\hat{x}_1$ and $\hat{x}_2$ are assumed to be minimizers, the above inequality has to be an equality. To see this, assume to the contrary that we can have strict inequality,

$$\|A(\hat{x}_1)\hat{x}_1 - b\| \;>\; \|A(\hat{x}_2)\hat{x}_1 - b\| \; . \tag{5.23}$$

Then note that

$$\begin{aligned}
\|A(\hat{x}_2)\hat{x}_1 - b\|^2 &= \|A(\hat{x}_2)(\hat{x}_1 + \hat{x}_2 - \hat{x}_2) - b\|^2 \\
&= \|[A(\hat{x}_2)\hat{x}_2 - b] \;+\; A(\hat{x}_2)(\hat{x}_2 - \hat{x}_2)\|^2 \\
&= \|A(\hat{x}_2)\hat{x}_2 - b\|^2 \;+\; \|A(\hat{x}_2)(\hat{x}_1 - \hat{x}_2)\|^2 \\
&\geq \|A(\hat{x}_2)\hat{x}_2 - b\|^2 \; ,
\end{aligned}$$

where in the third step we used the fact that $\hat{x}_2$ is a solution and therefore satisfies the orthogonality condition $A^T(\hat{x}_2)[A(\hat{x}_2)\hat{x}_2 - b] = 0$.

Combining the above inequality with (5.23) we find that we must have

$$\|A(\hat{x}_1)\hat{x}_1 - b\| > \|A(\hat{x}_2)\hat{x}_2 - b\| \; ,$$

which contradicts the fact that both $\hat{x}_1$ and $\hat{x}_2$ are solutions and must therefore have equal maximum residual norms. Hence, equality must hold,

$$\|A(\hat{x}_1)\hat{x}_1 - b\| \;=\; \|A(\hat{x}_2)\hat{x}_1 - b\| \; . \tag{5.24}$$

This means that for $\hat{x}_1$, $A(\hat{x}_1)$ and $A(\hat{x}_2)$ are two perturbed matrices that achieve the maximum residual. It follows that $A(\hat{x}_2)\hat{x}_1 - b$ should be orthogonal to $A(\hat{x}_2)$ just like $A(\hat{x}_2)\hat{x}_2 - b$ is (by the assumed optimality of $\hat{x}_2$). To verify this, assume not. Then the norm of the residual $A(\hat{x}_2)\hat{x}_1 - b$ has to be strictly larger than the norm of the residual $A(\hat{x}_2)\hat{x}_2 - b$,

$$\|A(\hat{x}_2)\hat{x}_2 - b\| \ < \ \|A(\hat{x}_2)\hat{x}_1 - b\| \ .$$

By (5.24), we obtain $\|A(\hat{x}_2)\hat{x}_2 - b\| < \|A(\hat{x}_1)\hat{x}_1 - b\|$. This is a contradiction since both $\hat{x}_1$ and $\hat{x}_2$ are minima and must therefore have equal maximum residual norms.

We thus conclude that

$$A^T(\hat{x}_2)[A(\hat{x}_2)\hat{x}_1 - b] = 0 \ . \tag{5.25}$$

Using this condition and the orthogonality condition of $\hat{x}_2$, *viz.,*

$$A^T(\hat{x}_2)[A(\hat{x}_2)\hat{x}_2 - b] = 0 \ ,$$

in addition to the full rank property of $A(\hat{x}_2)$ we obtain that

$$\hat{x}_1 = [A^T(\hat{x}_2)A(\hat{x}_2)]^{-1}A^T(\hat{x}_2)b = \hat{x}_2 \ .$$

$\diamondsuit$

Remark 4. The last part of the above proof could have also been established by resorting to the parametrization of perturbations that lead to maximum residual norms. Indeed, equality (5.24) means that for $\hat{x}_1$, $A(\hat{x}_1)$ and $A(\hat{x}_2)$ are two perturbed matrices that achieve the maximum residual. Using Corollary 5.5, they must be related via

$$A(\hat{x}_2) = A(\hat{x}_1) \ + \ Y\left(I - \frac{\hat{x}_1\hat{x}_1^T}{\|\hat{x}_1\|^2}\right) \ , \tag{5.26}$$

for some Y that satisfies

$$\left(I - \frac{\hat{x}_1\hat{x}_1^T}{\|\hat{x}_1\|^2}\right)Y^T(A\hat{x}_1 - b) = 0 \ .$$

It further follows from (5.10) that the residual vector $A(\hat{x}_2)\hat{x}_1 - b$ is collinear with $A\hat{x}_1 - b$. Using this fact and the orthogonality condition of $\hat{x}_1$, *viz.,*

$$A^T(\hat{x}_1)[A\hat{x}_1 - b] = 0 \ ,$$

if we multiply (5.26) by $[A(\hat{x}_2)\hat{x}_1 - b]^T$ from the left we obtain (5.25), and the argument can now be continued as above.

$\diamondsuit$

We therefore established that the solution of (4.3) is unique and nonzero when (5.5) holds. Using the orthogonality condition of the solution, it is now immediate to confirm that when (5.5) holds, the nonzero solution indeed has a smaller cost than the one associated with the zero vector. To see this, recall from Lemma 5.1 that when (5.5) holds, none of the matrices in the set $\{A + \delta A\}$ will be orthogonal to b. Hence, the distance from b to any of these matrices will always be strictly smaller than b. In particular, the optimal nonzero solution $\hat{x}$ will satisfy

$$\|A(\hat{x})\hat{x} - b\| \; < \; \|b\| \, ,$$

since the left-hand side measures the distance from b to the vector $A(\hat{x})\hat{x}$ in the column space of $A(\hat{x})$. Recalling that the optimal cost associated with $x = 0$ is $\|b\|$, and since $\|A(\hat{x})\hat{x} - b\|$ is the maximum residual associated with $\hat{x}$, we see that the nonzero solution $\hat{x}$ does have a smaller cost.

Remark 5. The cost function in (5.16) can be shown to be strictly convex when the assumptions (5.1) hold. Therefore, a unique global minimum $\hat{x}$ should exist. This minimum can occur either at the points where the cost function $\|Ax - b\| + \eta\|x\|$ is not differentiable (*viz.*, $\hat{x} = 0$) or at the points where the gradient with respect to x is zero. Since $\hat{x} = 0$ is not a solution when $\eta < \|A^T b\|/\|b\|$, we conclude that a unique nonzero solution $\hat{x}$ exists and it is equal to the vector where the gradient of the cost function is zero. Differentiating the cost function with respect to x, and setting the gradient equal to zero at $x = \hat{x}$, we obtain the orthogonality condition (5.20). While this optimization-based argument provides a short route to the solution, it nevertheless obscures the geometry of the problem. For this reason, in our presentation in this paper we have opted for emphasizing the geometric and linear algebraic aspects of the BDU formulation and its solution.

$\diamondsuit$

5.5. Statement of Solution. Returning to the orthogonality condition (5.22), we introduce the auxiliary positive number

$$\hat{\alpha} \triangleq \frac{\eta\|A\hat{x} - b\|}{\|\hat{x}\|} \, . \tag{5.27}$$

Then we can rewrite (5.22) in the form

$$\left(A^T A + \hat{\alpha} I\right) \hat{x} = A^T b \, , \tag{5.28}$$

where $\hat{\alpha}$ is clearly a function of $\hat{x}$ as well.

Expressions (5.27)-(5.28) define a system of equations with two unknowns $\{\hat{x}, \hat{\alpha}\}$. We already know that this system of equations has a unique solution $\{\hat{x}, \hat{\alpha}\}$.

We summarize here the conclusions of the earlier sections.

THEOREM 5.10 (Solution of BDU estimation). *Consider a full rank matrix $A \in \mathbb{R}^{N\times n}$ with $N > n$ and a nonzero vector b that does not belong to the column span of A. The solution of the* BDU *estimation problem*

$$\min_{x} \max_{\|\delta A\| \le \eta} \|(A + \delta A)x - b\| \, ,$$

is always unique. Two scenarios arise depending on the size of η.

1. *The solution is zero ($\hat{x} = 0$) if, and only if, $\eta \ge \|A^T b\|/\|b\|$.*
2. *The solution is nonzero if, and only if, $\eta < \|A^T b\|/\|b\|$. In this case, it is given by the solution of the nonlinear system of equations*

$$\left(A + \eta \frac{(A\hat{x} - b)}{\|A\hat{x} - b\|} \; \frac{\hat{x}^T}{\|\hat{x}\|}\right)^T [A\hat{x} - b] \; = \; 0 \, .$$

Alternatively, the unique $\hat{x}$ can be found by solving the nonlinear system of equations (5.27)-(5.28) in $\hat{x}$ and $\hat{\alpha}$, viz.,

$$\left(A^T A + \hat{\alpha} I\right) \hat{x} = A^T b \, ,$$
$$\hat{\alpha} - \eta \, \frac{\|A\hat{x} - b\|}{\|\hat{x}\|} = 0 \, .$$

$\diamondsuit$

If we replace (5.28) into (5.27) we obtain a nonlinear equation in $\hat{\alpha}$,

$$\hat{\alpha} = \eta \, \frac{\|[A(A^T A + \hat{\alpha} I)^{-1} A^T b - I]b\|}{\|(A^T A + \hat{\alpha} I)^{-1} A^T b\|} \, . \tag{5.29}$$

The mapping between the variables $\hat{x}$ and $\hat{\alpha}$ is bijective. Given $\hat{x}$ we can evaluate $\hat{\alpha}$ uniquely via (5.27) and given $\hat{\alpha}$ we can evaluate $\hat{x}$ uniquely via (5.28). Hence, since the solution $\hat{x}$ is nonzero and unique when $\eta < \|A^T b\|/\|b\|$, the above nonlinear equation in $\hat{\alpha}$ has a unique positive solution $\hat{\alpha}$. In [47], a method is presented for finding this root by introducing the SVD of the matrix A in order to further simplify the nonlinear equation (5.29). The scalar $\hat{\alpha}$ can be determined, for example, by employing a bisection-type algorithm to solve the nonlinear equation, thus requiring $O\left(n \log \frac{\hat{\alpha}}{\epsilon}\right)$ operations, where ϵ is the desired precision.

5.6. Connection to Regularized Least-Squares. We remarked earlier that the cost function (5.16) looks deceptively similar, but significantly distinct, from the regularized least-squares formulation (3.1), where the *squared* Euclidean norms $\{\|x|^2, \|Ax - b\|^2\}$ are used rather than the norms themselves. Indeed, the arguments in the earlier sections highlighted several of the subtleties involved in solving the BDU estimation problem, compared to the solution of regularized least-squares.

Interesting enough however, the solution of the BDU problem turns out to have a regularized form since

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

This can be regarded as the exact solution of a regularized least-squares problem of the form:

$$\min_{\hat{x}} \; \left(\hat{\alpha}\|x\|^2 \; + \; \|Ax - b\|^2\right) \tag{5.30}$$

with squared Euclidean distances, and where the regularization parameter $\hat{\alpha}$ is determined by the algorithm itself rather than specified by the designer. In this sense, the solution of the BDU problem (5.16) (with norms only rather than squared norms) can be seen to perform *automatic regularization*; it first determines a regularization parameter $\hat{\alpha}$ and then uses it to solve a regularized least-squares problem of the above form.

This observation allows us to also establish the following robustness property for the classical regularized least-squares solution.

THEOREM 5.11 (Robustness of regularized least-squares). *Consider a regularized least-squares problem of the form*

$$\min_{x} \; \left[\gamma\|x\|^2 \; + \; \|Ax - b\|^2\right] \, ,$$

where γ is a given positive number. Let $\hat{x}_{\text{rls}}$ denote its unique solution. Assume $A \in \mathbb{R}^{N\times n}$ is full rank with $N > n$ and that b does not belong to the column span of A. Assume also that $A^T b \neq 0$ so that $\hat{x}_{\text{rls}}$ is nonzero. The solution of every such problem is also the solution of a BDU *problem of the form*

$$\min_x \max_{\|\delta A\|\leq\eta} \|(A+\delta A)x - b\| \ , \tag{5.31}$$

for the following η:

$$\eta = \frac{\gamma\|\hat{x}_{\text{rls}}\|}{\|A\hat{x}_{\text{rls}} - b\|} \ . \tag{5.32}$$

Proof: To prove the result we need to verify that

$$\eta^2 \ = \ \frac{\gamma^2\|\hat{x}_{\text{rls}}\|^2}{\|A\hat{x}_{\text{rls}} - b\|^2} \ < \ \frac{\|A^T b\|^2}{\|b\|^2} \ , \tag{5.33}$$

so that the unique solution of the BDU problem (5.31)–(5.32) is $\hat{x}_{\text{rls}}$. For this purpose, we introduce the SVD of A, say

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

where U is $N\times N$ unitary, Σ is $n \times n$ diagonal, and V is $n\times n$ unitary. We denote the entries of Σ by $\{\sigma_1, \ldots, \sigma_n\}$. Let $\bar{b} = U^T b$ with entries $\{\bar{b}_i, 1 \leq i \leq N\}$. Then

$$\begin{aligned}
\gamma\|\hat{x}_{\text{rls}}\|^2 &= \sum_{i=1}^{n} \bar{b}_i^2 \sigma_i^2 \left(\frac{\gamma}{\gamma+\sigma_i^2}\right)^2 \ , \\
\|A\hat{x}_{\text{rls}} - b\|^2 &= \sum_{i=1}^{n} \bar{b}_i^2 \left(\frac{\gamma}{\gamma+\sigma_i^2}\right)^2 \ + \ \sum_{i=n+1}^{N} \bar{b}_i^2 \ , \\
\|A^T b\|^2 &= \sum_{i=1}^{n} \bar{b}_i^2\sigma_i^2 \ , \\
\|b\|^2 &= \sum_{i=1}^{N} \bar{b}_i^2 \ .
\end{aligned}$$

The fact that b does not belong to $\mathcal{R}(A)$ guarantees that $\sum_{i=n+1}^{N} \bar{b}_i^2 \neq 0$. The result (5.33) now follows by verifying that

$$\frac{\sum_{i=1}^{n} \bar{b}_i^2\sigma_i^2 \left(\frac{\gamma}{\gamma+\sigma_i^2}\right)^2}{\sum_{i=1}^{n} \bar{b}_i^2 \left(\frac{\gamma}{\gamma+\sigma_i^2}\right)^2 \ + \ \sum_{i=n+1}^{N} \bar{b}_i^2} \ < \ \frac{\sum_{i=1}^{n} \bar{b}_i^2\sigma_i^2}{\sum_{i=1}^{N} \bar{b}_i^2} \ ,$$

using $\gamma/(\gamma+\sigma_i^2) < 1$.

$\diamondsuit$

5.7. Back to the Image Processing Example. In order to demonstrate the performance of the BDU method, we reconsider the image processing example in Fig. 5.4. Fig. 5.4(a) shows the original image. Fig. 5.4(b) shows the blurred image with approximately 8.5% perturbation in A (*i.e.,* $\|\delta A\|/\|A\|$ is approximately 8.5%). Fig. 5.4(c) shows the failed least-squares restoration, while Fig. 5.4(d) shows a reasonably good restoration by the BDU solution. Figs. 5.4(e) and 5.4(f) show that both the LS and the BDU solutions perform well on the original blurred image when there are no perturbations in A.

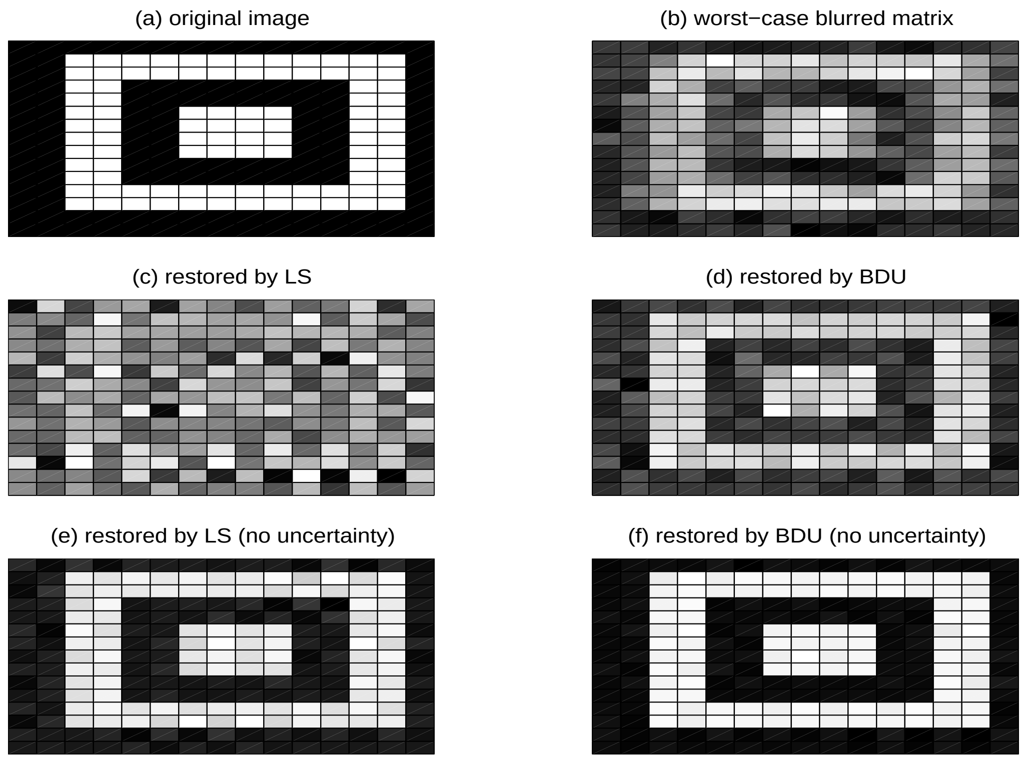

FIG. 5.4. Image processing example revisited.

6. BDU CONTROL. For the one-dimensional state-space regulation problem of Sec. 2.4, we consider the cost function (4.7), *viz.,*

$$\min_{x} \left(\max_{\|\delta A\|\leq\eta,\ \|\delta b\|\leq\beta} \|\,(A+\delta A)\,x - (b+\delta b)\|^2 \ + \ \rho\|x\|^2 \right) \ ,$$

where we allow for uncertainties in both A and b, in addition to a further weighting on x. While we shall treat this cost function and (4.8) in more detail elsewhere [50], here we only summarize its solution.

Let $A \in \mathbb{R}^{N\times n}$ be full rank with $N > n$ and assume b does not belong to the column span of A. If $\eta \geq \|A^T b\|/\|b\|$ then the unique solution is again $\hat{x} = 0$. Otherwise, the unique solution is given by

$$\hat{x} = \left(A^T A + \hat{\lambda} I\right)^{-1} A^T b \tag{6.1}$$

where $\hat{\lambda}$ is the unique positive root of the nonlinear equation:

$$\hat{\lambda} = \frac{\eta||A\hat{x}-b||}{||\hat{x}||} + \frac{\rho||A\hat{x}-b||}{||A\hat{x}-b||+\eta||\hat{x}||+\beta} \;. \tag{6.2}$$

[In our problem below, however, b lies in the range space of A. The solution will generally have the same form (6.1)–(6.2) except in two cases where we choose either $\hat{\lambda}=0$ or $\hat{\lambda}=\infty$ – details are given in [50].]

For the quadratic regulator problem of Sec. 2.4 with parametric uncertainty, we can reformulate each step of the LQR design as follows:

$$\min_{u_N} \max_{\begin{pmatrix} |\delta f_N| \leq \eta_f \\ |\delta g_N| \leq \eta_g \end{pmatrix}} \left[p x_{N+1}^2 + q u_N^2 + r x_N^2 \right] \;.$$

Here, δf_N and δg_N denote the uncertainties in f and g at step N. They are both bounded by η_f and η_g, respectively. If we now replace x_{N+1} by

$$x_{N+1} = (f+\delta f_N)x_N + (g+\delta g_N)u_N \;,$$

the above cost reduces, after grouping terms, to one of the form

$$\min_{u_N} \left(\max_{|\delta a|\leq\eta,\, |\delta b|\leq\beta} |\,(a+\delta a)\,u_N - (b+\delta b)|^2 \;+\; q|u_N|^2 \right) \;,$$

where

$$a = p^{1/2}g, \;\; b = -p^{1/2}f x_N, \;\; \eta = p^{1/2}\eta_g, \;\; \beta = p^{1/2}\eta_f |x_N| \;.$$

Using the solution of the BDU control cost we obtain the following state-feedback law (when the expression for λ_N below evaluates to a positive number):

$$\left\{ \begin{array}{lcl} \hat{u}_N & = & -k_N x_N \\ k_N & = & \frac{f g p_{N+1}}{\lambda_N + g^2 p_{N+1}} \\ p_N & = & f^2 p_{N+1} \left[\frac{\lambda_N + \frac{\eta_g}{|g|} g^2 p_{N+1}}{\lambda_N + g^2 p_{N+1}} + \frac{\eta_f}{|f|} \right]^2 + \frac{f^2 g^2 q p_{N+1}^2}{(\lambda_N + g^2 p_{N+1})^2} + r \\ \lambda_N & = & \frac{1}{1+\frac{\eta_f}{|f|}} \left\{ \frac{q}{1-\frac{\eta_g}{|g|}} - p_{N+1} g^2 \left(\frac{\eta_g}{|g|} + \frac{\eta_f}{|f|} \right) \right\} \end{array} \right.$$

The difference between the above solution and the LQR solution is that the gain constant k_N has a term λ_N in the denominator rather than q. The λ_N is propagated by the algorithm and enters into the recursion for p_N. [When the expression for λ_N evaluates to a negative value, it can be shown that λ_N should be set to zero, $\lambda_N = 0$ [50]. Also, when $\eta_g/|g| > 1$, we must set $\lambda_N = \infty$.]

The BDU control law has some interesting and meaningful features. When $\eta_f = \eta_g = 0$, it collapses to the Riccati recursion of the LQR case. In other words, the BDU solution collapses to the expected one in the absence of uncertainties. Moreover, when $\lambda_N = 0$ (which occurs for large uncertainties), the gain constant becomes $k_N = f/g$, which is the optimal $\mathcal{H}_\infty$ solution in this case for the largest possible diagonal uncertainty $\Delta(z)$ (as we saw earlier at the end of Sec. 3.3). Finally, when $\eta_g/|g| > 1$ the uncertainty in g is so large that the sign of g itself is unknown (it can be positive as well as negative). In this case, the BDU solution cancels the control and sets it equal to zero.

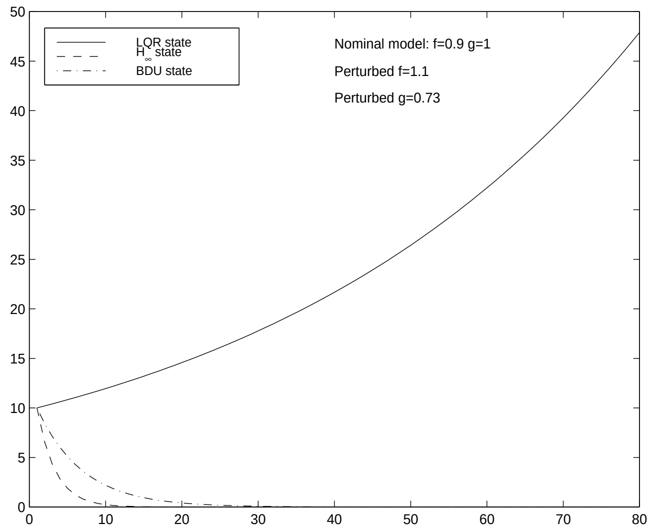

FIG. 6.1. Comparison of the LQR, $\mathcal{H}_\infty$, and BDU designs.

Once the problem has been solved at step N, we can proceed to the next step and solve

$$\min_{u_{N-1}} \max_{\left(\begin{array}{c} |\delta f_{N-1}| \leq \eta_f \\ |\delta g_{N-1}| \leq \eta_g \end{array}\right)} \left[p_N x_N^2 + q u_{N-1}^2 + r x_{N-1}^2\right] .$$

Fig. 6.1 shows the results obtained with this design procedure. The solid line shows the divergence of the LQR design. The dashed line shows the convergence of the $\mathcal{H}_\infty$ state to zero, while the dash-dotted line shows the convergence of the BDU state to zero at a total cost of 56.65. Also, the closed-loop pole is now located at 0.8449. This is in contrast to the $\mathcal{H}_\infty$ cost of 71.53 and to the location of the $\mathcal{H}_\infty$ closed-loop pole at 0.6595.

Fig. 6.2 compares the performance (cost) of the LQR, $\mathcal{H}_\infty$, and BDU designs in terms of the resulting control and state energies over 300 random runs. The figure demonstrates a consistent performance of the BDU method (dark line). The almost horizontal line refers to the $\mathcal{H}_\infty$ design. The curve with occasional spikes refers to the LQR design. Still, despite these results, there are several important issues to be addressed, such as stability results, comparison with parametric approaches in the literature, and extensions to MIMO systems. We shall pursue these studies elsewhere.

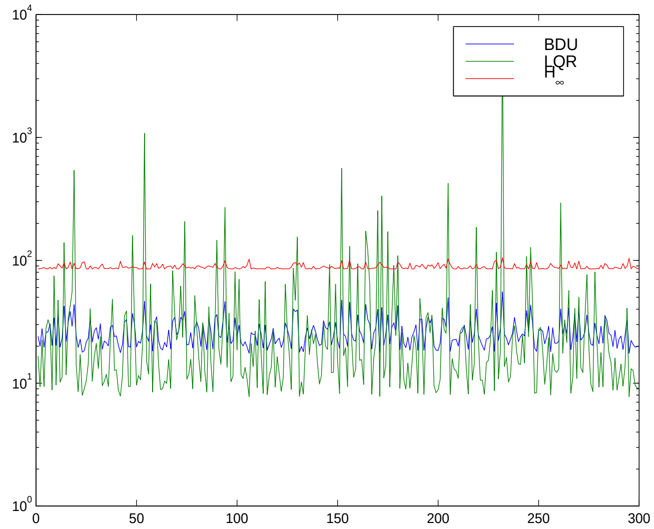

FIG. 6.2. 300 random runs with $\eta_f = 0.2$ and $\eta_g = 0.27$.

7. BDU ESTIMATION WITH MULTIPLE UNCERTAINTIES. We now demonstrate briefly an application of the BDU cost function (4.5) that deals with the case of multiple sources of uncertainties in the data [49], *viz.*,

$$\min_x \left(\max_{\|\delta A_j\| \le \eta_j} \left\| \begin{bmatrix} A_1 + \delta A_1 & \dots & A_K + \delta A_K \end{bmatrix} x - b \right\| \right) , \tag{7.1}$$

where the $\{A_j\}$ denote column-wise partitions of A.

Again, it can be verified that the nontrivial solution $\hat{x}$ is of the form

$$\hat{x} = \left(A^T A + \text{diag}\{\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_K\} I\right)^{-1} A^T b \,,$$

with K regularization parameters that are now used in diagonal form. If we partition $\hat{x}$ accordingly with the A_j, say $\hat{x} = \text{col}\{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_K\}$, then the $\hat{\alpha}_j' s$ are found by solving the K coupled nonlinear equations,

$$\hat{\alpha}_j = \eta_j \frac{\|A\hat{x} - b\|}{\|\hat{x}_j\|} \,, \quad 1 \le j \le K \,.$$

An application arises in the context of co-channel interference cancellation, as depicted in a simplified form in Fig. 7.1 for the case of two sources.

Assume there are 2 emitters sending at time i the signals $\{x_i, \theta_i\}$ from different angles to an antenna array. The antenna array has 4 elements that are equally spaced. The signal received by the elements of the antenna array can be presented in vector form as

$$b_i = A_x x_i + A_\theta \theta_i + v_i \,, \tag{7.2}$$

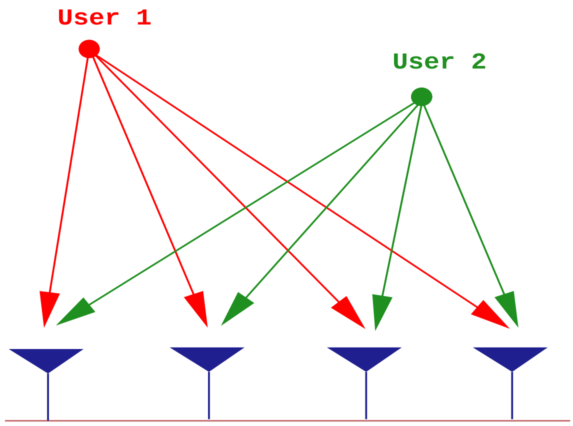

FIG. 7.1. Spatial-processing with multiple users.

where v_i denotes a measurement noise vector. Moreover, A_x and A_θ are column vectors. The $j-$th entry of A_x is the gain from source x to the $j-$th antenna. Likewise, the $j-$th entry in A_θ is the gain from source θ to the $j-$th antenna. In practice, these gains are estimated by a variety of methods (*e.g.*, MUSIC, ESPRIT, and many others – see [51, 52] and the many references therein) and are therefore subject to errors. They can also be subject to different levels of errors. The BDU formulation allows us to handle such situations with multiple sources of uncertainties, say

$$\|\delta A_x\| \leq \eta_x , \quad \|\delta A_\theta\| \leq \eta_\theta .$$

We can recover the $\{x_i, \theta_i\}$ by solving

$$\min_{x_i,\theta_i} \max_{\|\delta A_x\|\leq\eta_x, \|\delta A_\theta\|\leq\eta_\theta} \left\| \begin{bmatrix} A_x + \delta A_x & A_\theta + \delta A_\theta \end{bmatrix} \begin{bmatrix} x_i \\ \theta_i \end{bmatrix} - b_i \right\| ,$$

which is a special case of (4.5). Fig. 7.2 compares the performance (in terms of mean-square error) of the BDU solution with alternative methods such as least-squares, total least-squares, and cross-validation [36] for 4PAM modulation with 7% and 22% relative uncertainties in the path gains. The top curve corresponds to total-least-squares while the bottom curve corresponds to BDU. The second curve from top is least-squares and the third curve is generalized cross-validation.

Figure 7.3 repeats the same experiment in a different context, where the signals $\{x_i, \theta_i\}$ now represent the pixels of two 128×128 images that are being transmitted over different paths. Hence, the purpose is to identify and separate the superimposed images. In this particular simulation, we took $\eta_x = \eta_\theta = 7\%$. We see that the result from the BDU solution is the clearest. In Fig. 7.4 we further perform median filtering on the outputs of Fig. 7.3. Again, the BDU solution comes out most enhanced.

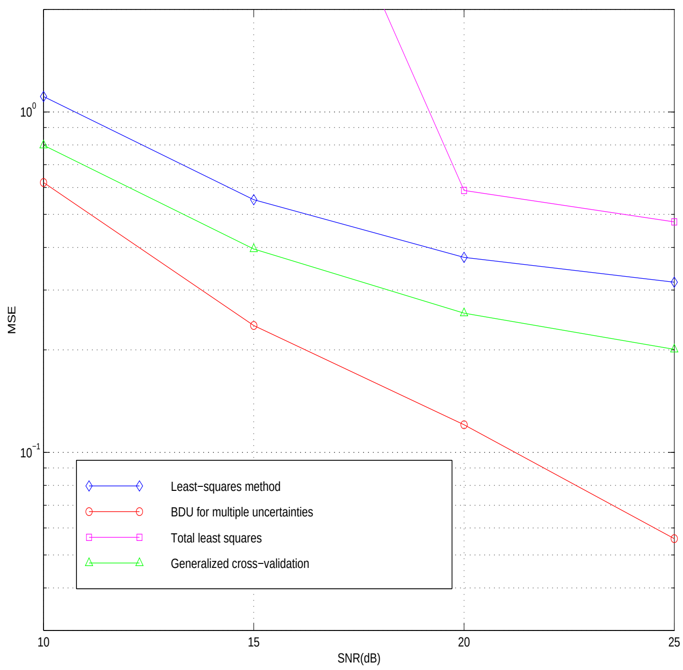

FIG. 7.2. *4PAM modulation, $N = 4000$ runs, $\eta_x \approx 7\%$, $\eta_\theta \approx 22\%$.*

8. CONCLUDING REMARKS. This paper developed a geometric framework for BDU problems and exhibited several examples that demonstrate the performance of the BDU methods in estimation and control. The results show that there is merit to these methods, but there are many issues and extensions that remain to be addressed.

In particular, it would be useful to study the statistical properties of the BDU estimators in terms of bias and consistency. It would also be useful to study the stability properties of BDU designs for closed-loop estimation and control, as well as extend the control results to higher-dimensional state-space models.

For on-line operation, it is also useful to develop recursive (adaptive) variants for BDU estimation. Preliminary and encouraging results in this direction appear in [53, 54], where an RLS-type result was developed for BDU estimation. The algorithm exploits a fundamental contraction property of the nonlinear mapping (5.29) for $\hat{\alpha}$ and uses it to determine the fixed point $\hat{\alpha}$ recursively.

Solution methods that exploit structure, as well as sparsity, in the data are also of interest in order to further reduce the computational cost.

Extensions to continuous-time results can also be pursued, where now operators should replace matrices. Also, more general BDU cost functions can be studied, as well as stochastic formulations where the variables $\{x, \delta A\}$ are described statistically.

Finally, we may add that some of the cost functions that we introduced here are convex in the unknown x. They can therefore be solved via convex optimization techniques. These methods, however, are costlier than the direct solution methods of this paper. They are also iterative techniques that obscure the geometry of the underly-

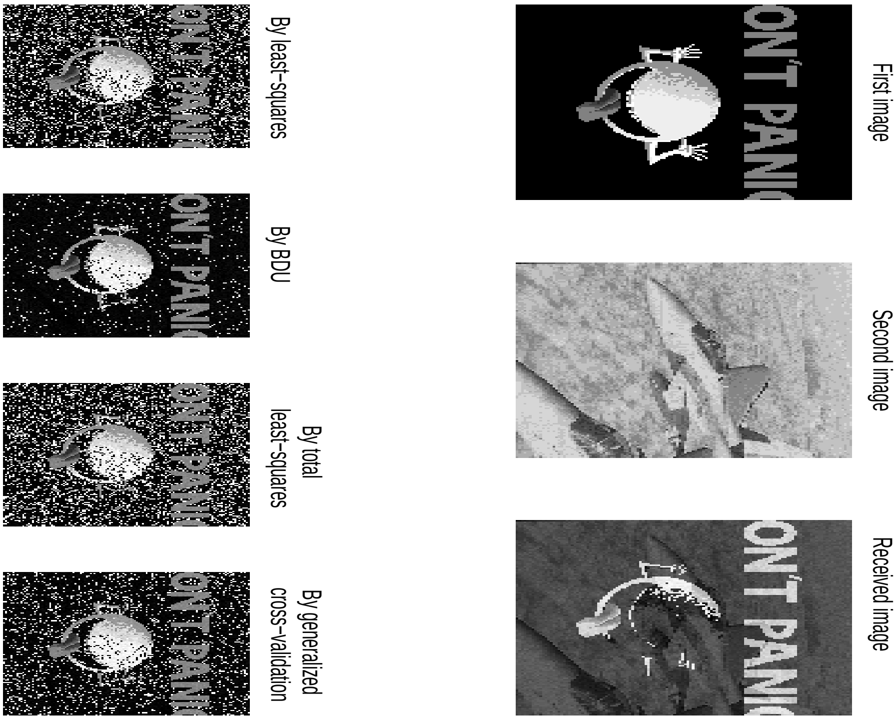

FIG. 7.3. Image separation.

ing problems. In our approach, we have relied almost entirely on the geometry of the BDU problem and have characterized its solution in these terms. Nevertheless, the convex optimization approach can handle situations that are possibly more difficult to handle directly or for which no direct solutions are yet known (see, *e.g.,* [48, 55]).

We end by mentioning a non-convex BDU cost function introduced in [56], *viz.,*

$$\min_{(x, \|\delta A\| \leq \eta)} \ \| (A + \delta A) x - b \| .$$

This cost is useful for design purposes where the objective is to select system parameters that would result in the smallest cost possible. This is a good example of a non-convex optimization problem with some of the headaches that come with it (*e.g.,* multiple minima) – but is still solvable [56].

Acknowledgments

The authors are grateful to Mr. F. Xu of ULCA for his help with the simulations of Sec. 7. They are also thankful to Profs. Ming Gu of UCLA and Gene Golub of Stanford University for helpful discussions and feedback on the topic of this paper.

REFERENCES

[1] B. D. O. Anderson and J. B. Moore, *Optimal Filtering*, Prentice-Hall Inc., NJ, 1979.

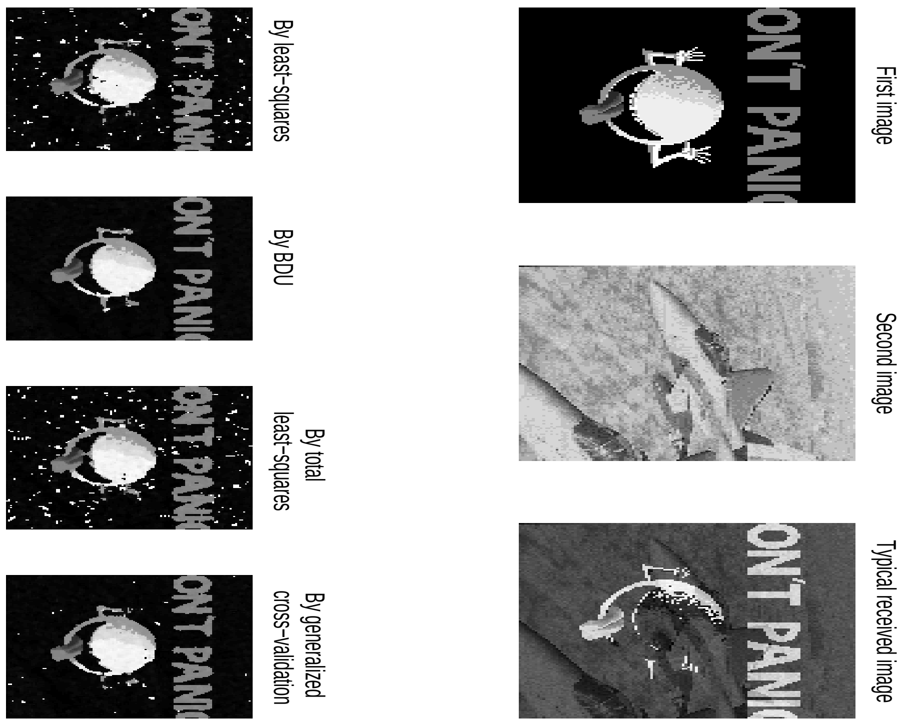

FIG. 7.4. Image separation after median filtering.

[2] T. Kailath, A. H. Sayed, and B. Hassibi, *State-Space Estimation Theory*, Prentice Hall, NJ, to appear.

[3] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, NJ, 1987.

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

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

[6] J. G. Proakis, C. M. Rader, F. Ling, and C. L. Nikias, *Advanced Digital Signal Processing*. Macmillan Publishing Co., NY 1992.

[7] J. G. Proakis, Digital Communications, 3rd edition, McGraw-Hill, NY, 1995.

[8] S. Haykin, *Adaptive Filter Theory*, 3rd edition, Prentice Hall, NJ, 1996.

[9] M. Green and D. J. N. Limebeer, *Linear Robust Control*, Prentice Hall, NJ, 1995.

[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] K. Zhou, J. C. Doyle, and K. Glover, *Robust and Optimal Control*, Prentice Hall, NJ, 1996.

[12] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd edition, The Johns Hopkins University Press, Baltimore, 1996.

[13] A. Björck, *Numerical Methods for Least-Squares Problems*, SIAM, Philadelphia, 1996.

[14] G. W. Stewart, *Theory of the Combination of Observations Least Subject to Errors*, Classics in Applied Mathematics, SIAM, PA, 1995. [Translation of original works by C. F. Gauss under the title *Theoria Combinationis Observationum Erroribus Minimis Obnoxiae*.]

[15] A. N. Tikhonov, Regularization of incorrectly posed problems, *Soviet Math.*, vol. 4, pp. 1624–1627, 1963.

[16] C. L. Lawson and R. J. Hanson, *Solving Least-Squares Problems*, SIAM, Philadelphia, PA, 1995.

[17] G. H. Golub and C. F. Van Loan, An analysis of the total least squares problem, *SIAM J.*

Numer. Anal., vol. 17, pp. 883–893, 1980.

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

[19] M. A. Dahleh and I. J. Diaz-Bobillo, *Control of Uncertain Systems: A Linear Programming Approach*, Prentice Hall, NJ, 1995.

[20] J.R. Deller, Set membership identification in digital signal processing, *IEEE ASSP Magazine*, vol. 6, no. 4, Oct. 1989.

[21] M. Milanese, J. Norton, H. Piet-Lahanier, and E. Walter, *Bounding Approaches to System Identification*, Plenum Press, NY, 1996.

[22] A. H. Sayed and T. Kailath, A state-space approach to adaptive RLS filtering, *IEEE Signal Processing Magazine*, 11(3):18–60, July 1994.

[23] E. Anderson, Z. Bai, C. Bischof, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, S. Oustrouchov, and D. Sorensen, *LAPACK Users' Guide*, SIAM, PA, 1992.

[24] G. Demoment, Image reconstruction and restoration: overview of commom estimation structures and problems, *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 2024–2036, 1989.

[25] D.M. Titterington, General structure of regularization procedures in image reconstruction, *Astronomy and Astrophysics*, vol. 144, pp. 381–387, 1985.

[26] D.L. Phillips, A technique for the numerical solution of certain integral equations of the first kind, *J. Ass. Comput. Mach.*, vol. 9, pp. 84–97, 1962.

[27] N.P. Galatsanos and A.K. Katsaggelos, Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation, *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 322–336, July 1992.

[28] V.Z. Mesarović and N.P. Galatsanos and A.K. Katsaggelos, Regularized constrained total least squares image restoration, *IEEE Transactions on Image Processing*, vol. 4, no. 8, pp. 1096–1108, August 1995.

[29] 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(8):831–847, August 1989.

[30] P. Dorato, C. Abdallah, and V. Cerone, *Linear-Quadratic Control — An Introduction*, Prentice-Hall, 1995.

[31] H. Heffes, The effect of erroneous models on the Kalman filter response, *IEEE Trans. Aut. Control*, vol. AC-11, pp. 541–543, July 1966.

[32] F. H. Schlee, C. J. Standish, and N. F. Toda, Divergence in the Kalman filter, *AIAA Journal*, vol. 5, no. 6, pp. 1114-1120, June 1967.

[33] A. P. Sage and J. L. Melsa, *Estimation theory with applications to Communications and Control*, McGraw Hill, 1971.

[34] H. W. Sorenson, ed., *Kalman filtering: Theory and Application,* IEEE Press, 1985.

[35] R. K. Mehra, On the identification of variances and adaptive Kalman filtering, *IEEE Trans. Aut. Control,* ol. AC-15, pp. 175–184, April 1970.

[36] G. H. Golub and M. Heath, Generalized cross-validation as a method for choosing a good ridge parameter, *Technometrics*, vol. 21, no. 2, pp. 195–7, May 1979.

[37] R. J. Adcock, A problem in least-squares, *The Analyst*, vol. 5, pp. 53–54, 1878.

[38] T. Basar and P. Bernhard, *$\mathcal{H}^\infty$-Optimal Control and Related Minimax Design Problems - A Dynamic Game Approach*, Birkhauser, Boston, 1991.

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

[40] B. Hassibi, A. H. Sayed, and T. Kailath, Linear estimation in Krein spaces - Part I: Theory, *IEEE Trans. Automatic Control*, 41(1):18–33, January 1996.

[41] A. H. Sayed, B. Hassibi and T. Kailath, Inertia conditions for the minimization of quadratic forms in indefinite metric spaces, *Operator Theory: Advances and Applications*, I. Gohberg, P. Lancaster, and P. N. Shivakumar, Eds., vol. 87, pp. 309–347, Birkhauser, 1996.

[42] B. Hassibi, A. H. Sayed, and T. Kailath, *Indefinite Quadratic Estimation and Control: A Unified Approach to $\mathcal{H}_2$ and $\mathcal{H}_\infty$ Theories*, SIAM, PA, 1998.

[43] J. C. Doyle, Analysis of feedback systems with structured uncertainties, *IEE Proceedings*, Part D, vol. 129, no. 6, pp. 242–250, 1982.

[44] A. Packard and J. C. Doyle, The complex structured singular value, *Automatica*, vol. 29, no. 1, pp. 71–109, 1993.

[45] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*, Wiley, England, 1996.

[46] S. Chandrasekaran, G. Golub, M. Gu, and A. H. Sayed, "Parameter estimation in the presence

of bounded modeling errors," *IEEE Signal Processing Letters,* vol. 4, no. 7, pp. 195–197, July 1997.

[47] S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, Parameter estimation in the presence of bounded data uncertainties, *SIAM J. Matrix Analysis and Applications*, vol. 19, no. 1, pp. 235–252, Jan. 1998.

[48] L. E. Ghaoui and H. Hebret, Robust solutions to least-squares problems with uncertain data, *SIAM J. Matrix Anal. Appl.*, vol. 18, pp. 1035–1064, 1997.

[49] A. H. Sayed and S. Chandrasekaran, Estimation in the presence of multiple sources of uncertainties with applications, to appear in *Proc. Asilomar Conference*, Oct. 1998.

[50] V. Nascimento and A. H. Sayed, Optimal control design with bounded parametric uncertainties, *in preparation.*

[51] H. Krim and M. Viberg, Two decades of array signal processing research, *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 67–94, July 1996.

[52] A. J. Paulraj and C. B. Papadias, Space-time processing for wireless communications, *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49–83, Nov. 1997.

[53] A. H. Sayed, A. Garulli, and S. Chandrasekaran, A fast iterative solution for worst-case parameter estimation with bounded model uncertainties, in *Proc. American Control Conference*, vol. 3, pp. 1499–1503, NM, June 1997.
June 1997.

[54] A. H. Sayed, A. Garulli, V. Nascimento, and S. Chandrasekaran, Exponentially-weighted iterative solutions for worst-case parameter estimation, in *Proc. 5th IEEE Mediterranean Conference On Control and Systems*, Paphos, Cyprus, July 1997.

[55] M.S. Lobo, L.V. Vandenberghe, S. Boyd, and H. Lebret, "Second-order cone programming", *Technical report*, Information Systems Laboratory, Electrical Engineering Department, Stanford University, April 1997.

[56] S. Chandrasekaran, G. Golub, M. Gu, and A. H. Sayed, An efficient algorithm for a bounded errors-in-variables model, to appear in *SIAM J. Matrix Anal. Appl.*