1

2 3

A REGULARIZED ROBUST DESIGN CRITERION FOR UNCERTAIN DATA

A. H. SAYED, V. H. NASCIMENTO, AND F. A. M. CIPPARRONE

Abstract. The paper formulates and solves a robust criterion for least-squares designs in the presence of uncertain data. Compared with earlier studies, the proposed criterion incorporates simultaneously both regularization and weighting and allows for a large class of uncertainties. The solution method is based on reducing a vector optimization problem to an equivalent scalar minimization problem of a provably unimodal cost function; thus achieving significant reduction in computational complexity.

Key words. Least-squares, regularization, robustness, min-max, uncertainty, game problem.

AMS subject classifications. 15A06, 15A63, 65K10, 90C47, 91A40.

1. INTRODUCTION.

As is well-known, many estimation and control problems rely on solving regularized least-squares problems of the form

$$\min_x \left[x^T Q x + (Ax - b)^T W (Ax - b)\right] \tag{1.1}$$

where $x^T Qx$ is a regularization term, $Q > 0$ and $W \geq 0$ are Hermitian weighting matrices, x is an unknown $n-$dimensional column vector, A is a known $N \times n$ data matrix, and b is a known $N \times 1$ measurement vector. The solution of (1.1) is

$$\hat{x} = \left[Q + A^T W A\right]^{-1} A^T W b \tag{1.2}$$

where the invertibility of $(Q + A^T WA)$ is guaranteed by the positive-definiteness of Q.

When the nominal data $\{A, b\}$ are subject to disturbances and/or uncertainties, the performance of the optimal estimator (1.1) can degrade appreciably. For example, if the actual data matrix were $(A + \delta A)$, for some unknown perturbation δA, then the estimator (1.1) that is designed based on A alone, and without accounting for the existence of δA, can perform poorly. This fact has motivated numerous works in the literature that attempt to robustify the solution of least-squares designs in the presence of data uncertainties. Some notable methods are the total-least-squares and the $\mathcal{H}_\infty$ formalisms (see, e.g., [1, 2] and the many references therein). These methods

*This work was supported in part by a grant from the National Science Foundation under Award number ECS-9820765. An early version of some of the results in this article, without the proofs and new properties that are offered here, appeared as reference [9] in the Proceedings of the *Workshop on Robust Identification and Control*, held in Siena, Italy, July 1998.

†**Corresponding author**. A. H. Sayed is with the Electrical Engineering Department, University of California, Los Angeles, CA 90095. E-mail: sayed@ee.ucla.edu, fax: (310)206-8495, phone: (310)267-2142.

‡V. H. Nascimento and F. A. M. Cipparrone are with the Dept. of Electronic Systems Engineering (PSI-EPUSP) at the Universidade de São Paulo, Av. Prof. L. Gualberto, trav. 3, n 158, CEP 05508-900, SP, Brazil.

are known to lead to solutions that perform data de-regularization and that can at times be conservative.

In this work, we propose a robust alternative to the regularized and weighted least-squares problem (1.1), which will be shown to lead to a regularized solution, as opposed to a de-regularized solution. This property is useful, especially for on-line implementations, since the regularized solution does not require existence conditions. The special case of $Q = 0$ and $W = I$ (which corresponds to unweighted least-squares problems without regularization), was studied in [3, 4] by different methods; one relies on LMI techniques while the other relies on SVD techniques. However, nontrivial choices for $\{Q, W\}$ require special care and a new technique is developed herein that leads to the following contributions. First, the technique can handle a large class of data uncertainties, as will be explained below. This is achieved by formulating a problem that allows for a general description of the uncertainty set. Second, we show how to replace a vector optimization problem by a scalar minimization problem of a cost function that is provably unimodal. This step is at the heart of the proposed solution method since it leads to significant simplifications in complexity. A complete proof of its validity is provided in the appendices at the end of the paper (in order not to overburden the body of the text with technicalities).

Applications of the proposed methodology to recursive estimation, control, and data fusion problems appear in [5]–[7]; we refer the reader to these articles for motivation, examples, simulations, and comparisons with other related techniques. As a brief motivation, one application in the context of state-space estimation is succinctly described in Sec. 2.3, with full details provided in [6]. In most of the paper, however, we opt to focus on studying the properties and technical aspects of the robust least-squares problem that is formulated further ahead in (2.1).

As mentioned above, the formulation in this article is useful for two main reasons. First, it leads to a robust solution that involves regularization rather than de-regularization. In this way, existence conditions do not arise, which could be a burden for on-line solutions (see, e.g., [6]). Second, the framework incorporates both regularization and weighting into the cost function. Such extensions are needed in order to handle, for example, quadratic control and estimation problems where regularization and weighting are prevalent (see, e.g., [5, 6, 9]). It turns out that the treatment of these generalizations requires some care and is not an immediate extension of the unregularized and unweighted case.

2. PROBLEM FORMULATION. A generalization of the cost function in (1.1) that accounts for uncertainties in the data $\{A, b\}$ can be obtained as follows. Introduce the two-variable cost function

$$J(x, y) \triangleq x^T Q x + R(x, y)$$

where $R(x, y)$ is a modified residual term that is defined by

$$R(x, y) \triangleq \left(Ax - b + Hy\right)^T W \left(Ax - b + Hy\right)$$

Here, H is an $N \times m$ known matrix, and y denotes an $m \times 1$ unknown perturbation vector. When $H = 0$, $J(x, y)$ reduces to the standard regularized cost function in (1.1). The presence of H and y in the expression for $R(x, y)$ allows us to account for uncertainties in the data, as will become more evident from the discussions in the sequel.

To guarantee optimal performance in a worst-case scenario, we consider a min-max optimization problem of the form

$$\hat{x} = \arg\min_x \max_{\|y\| \leq \phi(x)} \quad J(x,y), \tag{2.1}$$

where the notation $\|\cdot\|$ stands for the Euclidean norm of its vector argument or the maximum singular value of its matrix argument. The non-negative function $\phi(x)$ is assumed to be a known bound on the perturbation y and is a function of x only.

Problem (2.1) can be interpreted as a constrained two-player game problem, with the designer trying to pick an $\hat{x}$ that minimizes the cost while the opponent $\{y\}$ tries to maximize the cost (e.g., [8]). The game problem is constrained since it imposes a limit (through $\phi(x)$) on how large (or how damaging) the opponent $\{y\}$ can be. We shall assume in the sequel that H and $\phi(x)$ are not identically zero, i.e,

$$H \neq 0 \quad \text{and} \quad \phi(x) \not\equiv 0,$$

since if either is zero, the game problem (2.1) trivializes to the standard regularized least-squares problem (1.1). The choice of H allows us to handle situations in which the uncertainties are known to be restricted to a certain subspace.

An initial study of problem (2.1) appears in [9] without the full details and new properties that are offered in this article and, in particular, without the arguments and proofs that appear in the appendices for general functions $\phi(x)$.

Two useful special cases of the formulation (2.1) are described below. They correspond to special choices of the function $\phi(x)$. These examples are meant to show how the freedom in selecting $\phi(x)$ allows us to handle different uncertainty models.

2.1. A Special Case: Bounded Uncertainties. Consider uncertainties $\{\delta A, \delta b\}$ that are only known to lie within certain balls of radii $\{\eta, \eta_b\}$, i.e., they are known to be bounded and satisfy

$$\|\delta A\| \leq \eta, \quad \|\delta b\| \leq \eta_b.$$

Now consider an optimization problem of the form

$$\min_x \max_{\substack{\|\delta A\| \leq \eta \\ \|\delta b\| \leq \eta_b}} \left[x^T Q x + \Big((A+\delta A)x - (b+\delta b) \Big)^T W \Big((A+\delta A)x - (b+\delta b) \Big) \right] \tag{2.2}$$

It can be verified that this problem is a special case of (2.1) since it can be shown to be equivalent to a problem of the form

$$\min_x \max_{\|y\| \leq \eta\|x\| + \eta_b} \left[x^T Q x + \Big(Ax - b + y \Big)^T W \Big(Ax - b + y \Big) \right] \tag{2.3}$$

which corresponds to the special choices $H = I$ and $\phi(x) = \eta\|x\| + \eta_b$.

To verify that problems (2.1) and (2.3) are indeed equivalent, we proceed as in [5] and show that the two variables $\{\delta A, \delta b\}$ in (2.1) can be replaced by a single variable y, which would therefore allow us to replace the maximization in (2.1) over *two* constrained variables, by a maximization over a *single* constrained variable as in (2.3).

Indeed, for any fixed value of x, let $\mathcal{Z}_x$ denote the set of all vectors z that are generated as follows:

$$\mathcal{Z}_x = \{z : \ \ z = \delta A x - \delta b \ , \ \ \|\delta A\| \leq \eta \ , \ \ \|\delta b\| \leq \eta_b\}$$

for all possible $\{\delta A, \delta b\}$ within the prescribed bounds. Let also $\mathcal{Y}_x$ denote the set of all vectors y that are generated as follows:

$$\mathcal{Y}_x = \{y : \ \ \|y\| \leq \eta\|x\| + \eta_b\} \ .$$

Then $\mathcal{Z}_x = \mathcal{Y}_x$. That is, if $z \in \mathcal{Z}_x$ then $z \in \mathcal{Y}_x$ (this direction is immediate and follows from the triangle inequality of norms). Conversely, if $y \in \mathcal{Y}_x$ then $y \in \mathcal{Z}_x$. To establish the result for $x \neq 0$, define for a given y the perturbations:

$$\delta A(y) = \frac{\eta}{\eta\|x\| + \eta_b} \frac{y x^T}{\|x\|}, \qquad \delta b(y) = -\frac{\eta_b y}{\eta\|x\| + \eta_b} \tag{2.4}$$

Then $\{\delta A(y), \delta b(y)\}$ are valid perturbations and $y = \delta A(y)x - \delta b(y)$ so that $y \in \mathcal{Z}_x$, which justifies our claim. [When $x = 0$, we select $\delta b = -y$ and δA arbitrary.]

As mentioned before, the special case $Q = 0$ and $W = I$ was treated in [3, 4] by different methods; one uses SVD techniques while the other uses LMI techniques. For this special case, a geometric framework that is similar in nature to the geometry of least-square problems was also developed in [10, 11].

2.2. A Special Case: Uncertainties in Factored Form. Consider now a problem of the form

$$\min_x \max_{\delta A, \delta b} \left[x^T Q x + \Big((A + \delta A)x - (b + \delta b) \Big)^T W \Big((A + \delta A)x - (b + \delta b) \Big) \right], \tag{2.5}$$

where the perturbations $\{\delta A, \delta b\}$ are assumed to satisfy a model of the form

$$\begin{bmatrix} \delta A & \delta b \end{bmatrix} = HS \begin{bmatrix} E_a & E_b \end{bmatrix} \tag{2.6}$$

where S is an arbitrary contraction, $\|S\| \leq 1$, and $\{H, E_a, E_b\}$ are *known* quantities of appropriate dimensions. Perturbation models of this form are common in robust filtering and control and can arise from tolerance specifications on physical parameters (see [12] for an example). The quantity H allows the designer to restrict the range of allowable uncertainties $\{\delta A, \delta b\}$ to a certain column span. Assume, for example, that one wishes to model only uncertainties in the $(0, 0)$ entry of A. Then one could choose

$$H = \mathrm{col}\{1, 0, \ldots, 0\}, \quad E_a = \begin{bmatrix} 1 & 0 & \ldots & 0 \end{bmatrix}, \quad E_b = 0,$$

and S would denote in this case an arbitrary scalar that is less than unity in magnitude. Other choices for $\{H, E_a, E_b\}$ would correspond to different assumptions on the uncertainties.

In order to see how (2.5) is related to (2.1), we rewrite the cost in (2.5) as

$$\left[x^T Q x + \Big(Ax - b + (\delta A x - \delta b) \Big)^T W \Big(Ax - b + (\delta A x - \delta b) \Big) \right]$$

so that with y defined as $y = S(E_a x - E_b)$ and Hy defined as

$$Hy \triangleq \delta A x - \delta b = HS(E_a x - E_b)$$

problem (2.5) can be verified to be equivalent to the following problem

$$\min_{x} \max_{\|y\| \leq \|E_a x - E_b\|} \left[x^T Q x + \Big((Ax - b + Hy \Big)^T W \Big((Ax - b + Hy \Big) \right]$$

which is again a special case of (2.1) for the particular choice $\phi(x) = \|E_a x - E_b\|$.

2.3. An Application: State Estimation. Before proceeding to a discussion of the solution and properties of the general problem (2.1), we motivate this optimization problem by considering an application in the context of state-space estimation.

Thus consider a state-space model of the form

$$x_{i+1} = F_i x_i + G_i u_i, \quad i \geq 0, \tag{2.7}$$

$$y_i = H_i x_i + v_i, \tag{2.8}$$

where $\{x_0, u_i, v_i\}$ are uncorrelated zero-mean random variables with variances

$$E\left(\begin{bmatrix} x_0 \\ u_i \\ v_i \end{bmatrix} \begin{bmatrix} x_0 \\ u_j \\ v_j \end{bmatrix}^T \right) = \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & Q_i \delta_{ij} & 0 \\ 0 & 0 & R_i \delta_{ij} \end{bmatrix} \tag{2.9}$$

that satisfy $\Pi_0 > 0$, $R_i > 0$, and $Q_i > 0$. Here, δ_{ij} is the Kronecker delta function that is equal to unity when $i = j$ and zero otherwise. The well-known Kalman filter [13] provides the optimal linear least-mean-squares (l.l.m.s., for short) estimate of the state variable given prior observations. It admits the following deterministic interpretation [14].

Fix a time instant i and assume that a so-called filtered estimate $\hat{x}_{i|i}$ of x_i has already been computed with the corresponding error variance matrix $P_{i|i}$. Given a new measurement y_{i+1}, one can seek to improve the estimate of x_i, along with estimating u_i, by solving

$$\min_{x_i, u_i} \left[\|x_i - \hat{x}_{i|i}\|^2_{P^{-1}_{i|i}} + \|u_i\|^2_{Q_i^{-1}} + \|y_{i+1} - H_{i+1} x_{i+1}\|^2_{R^{-1}_{i+1}} \right] \tag{2.10}$$

Substituting x_{i+1} by the state-equation $x_{i+1} = F_i x_i + G_i u_i$, the above cost function becomes one of the regularized and weighted least-squares form (1.1) and its solution leads to the Kalman filter recursions.

Now, assume that the state-space model includes parametric uncertainties, say of the form

$$x_{i+1} = (F_i + \delta F_i) x_i + (G_i + \delta G_i) u_i, \quad i \geq 0, \tag{2.11}$$

$$y_i = H_i x_i + v_i, \tag{2.12}$$

where the uncertainties $\{\delta F_i, \delta G_i\}$ lie in a certain domain, say of the form

$$\begin{bmatrix} \delta F_i & \delta G_i \end{bmatrix} = M_i S \begin{bmatrix} E_{f,i} & E_{g,i} \end{bmatrix}$$

for some known matrices $\{M_i, E_{f,i}, E_{g,i}\}$ and an arbitrary contraction S. We can then consider replacing (2.10) by

$$\min_{\{x_i,u_i\}} \max_{\delta F_i, \delta G_i} \left[\|x_i - \hat{x}_{i|i}\|^2_{P^{-1}_{i|i}} + \|u_i\|^2_{Q_i^{-1}} + \|y_{i+1} - H_{i+1}x_{i+1}\|^2_{R^{-1}_{i+1}} \right] \tag{2.13}$$

If we substitute x_{i+1} by its state-equation $x_{i+1} = (F_i + \delta F_i)x_i + (G_i + \delta G_i)u_i$, the above min-max problem becomes again a special case of the robust cost function (2.1); actually one of the form (2.5)–(2.6) — see [6] for the details, including numerical examples and comparison with several other classes of state-space estimation algorithms such as Kalman filters, $\mathcal{H}_\infty$ filters, guaranteed-cost filters, and set-valued estimation filters.

3. SOLUTION OF THE OPTIMIZATION PROBLEM. We now proceed to the solution of problem (2.1). In particular, we shall show that the solution has a regularized form, albeit one that operates on corrected data, i.e., it replaces $\{Q, W\}$ by corrections $\{\widehat{Q}, \widehat{W}\}$. In addition, and significantly, we shall show that the corrected parameters are determined in terms of the *unique* minimizing scalar argument, λ^o, of a *unimodal* cost function. In this way, we end up with a technique that enforces robustness via regularization, rather than de-regularization as is common in many robust procedures in the literature, and whose optimal solution involves determining the minimizing argument of a scalar unimodal function; a step that simplifies the complexity of the solution to great extent.

3.1. Uniqueness of Solution. We start by noting that the condition $Q > 0$ implies that (2.1) has a unique, finite solution. Indeed, for any given y, the residual cost $R(x, y)$ is convex in x. Therefore, the maximum

$$C(x) \stackrel{\Delta}{=} \max_{\|y\| \le \phi(x)} R(x, y) \tag{3.1}$$

is a convex function in x. In addition, the first term in $J(x, y)$, $x^T Qx$, is strictly convex in x and radially unbounded (i.e., $|x^T Qx|$ goes to infinity as $\|x\| \to \infty$) when $Q > 0$. We conclude that $x^T Qx + C(x)$ is also strictly convex in x and radially unbounded, which implies that problem (2.1) has a unique global minimum $\hat{x}$. To determine $\hat{x}$, we proceed in steps.

3.2. The Maximization Problem. We first solve (3.1) for any fixed x. Note that for fixed x, both the cost $R(x, y)$ and the constraint $\|y\| \le \phi(x)$ are convex in y, so that the maximum

$$\max_{\|y\| \le \phi(x)} R(x, y)$$

is achieved at the boundary, $\|y\| = \phi(x)$. We can therefore replace the inequality constraint in (3.1) by an equality. Introducing a Lagrange multiplier λ, the solution to (3.1) can then be found from the unconstrained problem:

$$\max_{y,\lambda} \left[\big(Ax - b + Hy\big)^T W \big(Ax - b + Hy\big) - \lambda\big(\|y\|^2 - \phi^2(x)\big) \right]. \tag{3.2}$$

Differentiating (3.2) with respect to y and λ, and denoting the optimal solutions by $\{y^o, \lambda^o\}$, we obtain the equations

$$(\lambda^o I - H^T WH)y^o = H^T W(Ax - b), \qquad \|y^o\| = \phi(x). \tag{3.3}$$

It turns out that the solution λ^o should satisfy $\lambda^o \geq \|H^T WH\|$. This is because the Hessian of the cost in (3.2) with respect to y, which is equal to

$$H^T WH - \lambda I$$

must be nonpositive-definite [15] at $\lambda = \lambda^o$.[1] We should further stress that the solutions $\{y^o, \lambda^o\}$ depend on x and we shall therefore sometimes denote this dependence explicitly by writing $\{y^o(x), \lambda^o(x)\}$[2].

At this stage, we do not need to solve the equations (3.3) for $\{y^o, \lambda^o\}$. It is enough to know that the optimal $\{y^o, \lambda^o\}$ satisfy (3.3). Using this fact, we can verify that the maximum cost in (3.2) is equal to

$$\begin{aligned} C(x) = (Ax-b)^T \left[W + WH\big(\lambda^o(x)I - H^T WH\big)^{\dagger} H^T W\right](Ax-b) \\ + \lambda^o(x)\phi^2(x), \end{aligned} \tag{3.4}$$

where the notation $X^{\dagger}$ denotes the pseudo-inverse of X.

3.3. The Minimization Problem. The original problem (2.1) is therefore equivalent to

$$\min_x \left[x^T Q x \;+\; C(x)\right] \;. \tag{3.5}$$

However, rather than minimizing the above cost over n variables, which are the entries of the vector x, we can instead reduce the problem to one of minimizing a certain cost function over a single scalar variable (see (3.9) further ahead). For this purpose, we introduce the following function of two independent variables x and λ,

$$C(x,\lambda) = (Ax-b)^T \left[W + WH\big(\lambda I - H^T WH\big)^{\dagger} H^T W\right](Ax-b) + \lambda\phi^2(x),$$

where λ is an independent variable. Then it can be verified, by direct differentiation with respect to λ and by using the expression for $\lambda^o(x)$ from (3.3), that

$$\lambda^o(x) \;=\; \arg \min_{\lambda \geq \|H^T WH\|} \; C(x,\lambda) \;. \tag{3.6}$$

In other words, the optimal $\lambda^o(x)$ from (3.3) coincides with the argument that minimizes $C(x,\lambda)$ over λ (with λ restricted to the interval $\big[\;\|H^T WH\|, \infty\big)$.

In this way, problem (2.1) becomes equivalent to

$$\min_x \min_{\lambda \geq \|H^T WH\|} \left[x^T Q x \;+\; C(x,\lambda)\right] = \min_{\lambda \geq \|H^T WH\|} \min_x \left[x^T Q x \;+\; C(x,\lambda)\right] \;. \tag{3.7}$$

The cost function in the above expression, viz., $J(x,\lambda) = x^T Q x + C(x,\lambda)$, is now a function of two independent variables $\{x,\lambda\}$. This should be contrasted with the cost function in (3.5). Now, for compactness of notation, introduce the quantities:

$$\begin{aligned} W(\lambda) &\stackrel{\Delta}{=} W + WH\big(\lambda I - H^T WH\big)^{\dagger} H^T W, \\ M(\lambda) &\stackrel{\Delta}{=} Q + A^T W(\lambda) A, \\ D(\lambda) &\stackrel{\Delta}{=} A^T W(\lambda) b. \end{aligned}$$

[1]We refer to the case $\lambda^o = \|H^T WH\|$ as the *singular* case, while $\lambda^o > \|H^T WH\|$ is called the *regular* case. Both cases are handled simultaneously in our framework through the use of the pseudo-inverse notation.

[2]In fact, we show in Appendix A that the solution $\lambda^o(x)$ is always a continuous function of x; while there might exist several y^o when $\lambda^o(x) = \|H^T WH\|$.

To solve problem (3.7), we first search for the minimum over x for every fixed value of λ, which can be done (if $\phi^2(x)$ is differentiable) by setting the derivative of $J(x,\lambda)$ w.r.t. x equal to zero. This shows that any minimum x must satisfy the equation

$$M(\lambda)x + \frac{1}{2}\lambda\nabla\phi^2(x) = D(\lambda), \tag{3.8}$$

where $\nabla\phi^2(x)$ is the gradient of $\phi^2(x)$ w.r.t. x. Any x satisfying this equation will of course be a function of λ, and we shall denote it by $x^o(\lambda)$. Now let $G(\lambda)$ denote the minimum value of the cost over x, i.e.,

$$G(\lambda) \triangleq \min_x [x^T Q x + C(x,\lambda)] \;=\; x^{oT}(\lambda) Q x^o(\lambda) + C(x^o(\lambda),\lambda).$$

Then problem (3.7) becomes equivalent to

$$\min_{\lambda \geq \|H^T W H\|} G(\lambda).$$

Thus we see that the solution of (2.1) simply requires that we determine an optimal scalar parameter λ^o. The scalar minimization problem that defines λ^o is simple, since (as we show in Thm. 3.1 below) $G(\lambda)$ is unimodal (i.e., there is a unique global minimum λ^o, and no other local minima).

3.4. Statement of Solution in the General Case. Whenever $\phi(x)$ is a convex function, the cost $J(x,\lambda)$ will be strictly convex in x, and the minimization over x on the right-hand side of Eq. (3.7) will have a unique solution $x^o(\lambda)$ (as was the case with the above two special cases). We thus have a procedure that allows us to determine the minimizing x^o for every λ. This in turn allows us to re-express the resulting cost $J(x^o(\lambda),\lambda)$ as a function of λ alone, say $G(\lambda) = J(x^o(\lambda),\lambda)$. In this way, we concluded above that the solution x^o of the original optimization problem (2.1) can be solved by determining the λ^o that solves

$$\min_{\lambda \geq \|H^T W H\|} G(\lambda)\ , \tag{3.9}$$

and by taking the corresponding $x^o(\lambda^o)$ as x^o. We summarize the solution in the following statement.

THEOREM 3.1 (Solution). *Consider a regularized and weighted robust least-squares problem of the form*

$$\hat{x} = \arg\min_x \max_{\|y\| \leq \phi(x)} \left[x^T Q x + \Big(Ax - b + Hy\Big)^T W \Big(Ax - b + Hy\Big) \right] \tag{3.10}$$

where $\{A, b, H\}$ are known quantities of appropriate dimensions, $W \geq 0$, and $Q > 0$ are known weighting matrices, and $\phi(x)$ is a given convex function. It is further assumed that H and $\phi(x)$ are not identically zero. Then problem (3.10) has a unique global minimum $\hat{x}$ that can be determined as follows:

1. Introduce the modified matrices

$$\begin{aligned} W(\lambda) &\triangleq W + WH\big(\lambda I - H^T W H\big)^{\dagger} H^T W, \\ M(\lambda) &\triangleq Q + A^T W(\lambda) A, \\ D(\lambda) &\triangleq A^T W(\lambda) b. \end{aligned}$$

2. *Let $x^o(\lambda)$ denote the unique solution of the minimization problem*

$$\min_x \left[x^T Q x \ + \ (Ax-b)^T W(\lambda)(Ax-b) + \lambda\phi^2(x)\right]$$

When $\phi^2(x)$ is differentiable, $x^o(\lambda)$ can also be found as the unique solution of the equation

$$M(\lambda)x + \frac{1}{2}\lambda\nabla\phi^2(x) = D(\lambda)$$

where the notation $\nabla\phi^2(x)$ denotes the gradient of $\phi^2(x)$ w.r.t. x.

3. *Introduce the cost function*

$$G(\lambda) \ = \ x^{oT}(\lambda)Qx^o(\lambda) + C[x^o(\lambda), \lambda]$$

4. *Let λ^o denote the solution of the scalar-valued minimization problem*

$$\lambda^o = \arg \min_{\lambda \geq \|H^T W H\|} G(\lambda)$$

5. *Then the optimum solution of (3.10) is $\hat{x} = x^o(\lambda^o)$. In addition, it holds that the cost function $G(\lambda)$ is unimodal, i.e., it has a unique global minimum and no local minima.*

Proof. The only point not yet proven is the fact that $G(\lambda)$ is unimodal. This follows from Lemma C.2 in Appendix C, and from the continuity of $\lambda^o(x)$ in (3.6), which is established in Appendix A.2.

$\square$

We now illustrate the solution method by reconsidering the two special cases we introduced before. In both examples, $\phi(x)$ is convex, so the minimization problem over x in (3.7) has a unique solution and is easily computable. In one of the examples, $\phi^2(x)$ is not differentiable at $x = 0$.

3.5. Uncertainties in Factored Form. Consider first the special case of Sec. 2.2 with

$$\phi(x) = \|E_a x - E_b\|.$$

For this choice of $\phi(x)$, we obtain

$$\nabla\phi^2(x) = 2E_a^T\left(E_a x - E_b\right),$$

so that the solution of Eq. (3.8), which is dependent on λ, becomes

$$x^o(\lambda) = \left[M(\lambda) + \lambda E_a^T E_a\right]^{-1}\left(D(\lambda) + \lambda E_a^T E_b\right). \tag{3.11}$$

Using this expression for $x^o(\lambda)$ we find that the corresponding function $G(\lambda)$ is given by

$$G(\lambda) = \lambda E_b^T E_b + b^T W(\lambda) b - B^T(\lambda)E^{-1}(\lambda)B(\lambda),$$

where $W(\lambda)$ is as before, and the functions $\{B(\lambda), E(\lambda)\}$ are given by

$$B(\lambda) = A^T W(\lambda) b + \lambda E_a^T E_b,$$
$$E(\lambda) = Q + \lambda E_a^T E_a + A^T W(\lambda) A.$$

We are thus led to the following statement.

THEOREM 3.2 (Uncertainties in Factored Form). *Consider a regularized and weighted robust least-squares problem of the form*

$$\min_x \max_{\delta A,\delta b} \left[x^T Q x + \Big((A+\delta A)x - (b+\delta b) \Big)^T W \Big((A+\delta A)x - (b+\delta b) \Big) \right], \tag{3.12}$$

where $\{A,b\}$ are known quantities of appropriate dimensions, $W \geq 0$, and $Q > 0$ are known weighting matrices, and the perturbations $\{\delta A, \delta b\}$ are assumed to satisfy a model of the form

$$\left[\begin{array}{cc} \delta A & \delta b \end{array} \right] \; = \; HS \left[\begin{array}{cc} E_a & E_b \end{array} \right]$$

for some known quantities $\{H, E_a, E_b\}$ and where S denotes an arbitrary contraction. Then problem (3.12) has a unique global minimum $\hat{x}$ that is given by (compare with (1.2)):

$$\hat{x} = \left[\widehat{Q} + A^T \widehat{W} A \right]^{-1} \left[A^T \widehat{W} b + \lambda^o E_a^T E_b \right] \tag{3.13}$$

where the modified weighting matrices $\{\widehat{Q}, \widehat{W}\}$ are obtained from $\{Q, W\}$ via

$$\begin{aligned} \widehat{Q} &\stackrel{\Delta}{=} Q + \lambda^o E_a^T E_a \\ \widehat{W} &\stackrel{\Delta}{=} W + WH(\lambda^o I - H^T W H)^\dagger H^T W \end{aligned}$$

and the nonnegative scalar parameter λ^o is determined from the scalar-valued optimization

$$\lambda^o \; = \; \arg \min_{\lambda \geq \|H^T W H\|} G(\lambda)$$

where the function $G(\lambda)$ is defined as

$$G(\lambda) = \|x^o(\lambda)\|_Q^2 \; + \; \lambda \|E_a x^o(\lambda) - E_b\|^2 \; + \; \|A x^o(\lambda) - b\|_{W(\lambda)}^2$$

Here

$$\begin{aligned} W(\lambda) &\stackrel{\Delta}{=} W + WH\big(\lambda I - H^T W H\big)^\dagger H^T W \\ Q(\lambda) &\stackrel{\Delta}{=} Q + \lambda E_a^T E_a \end{aligned}$$

and

$$x^o(\lambda) \stackrel{\Delta}{=} \left[Q(\lambda) + A^T W(\lambda) A \right]^{-1} \left[A^T W(\lambda) b + \lambda E_a^T E_b \right]$$

$\square$

We thus see that the solution of (3.12) requires that we first determine an optimal nonnegative scalar parameter, λ^o, which corresponds to the minimizing argument of the function $G(\lambda)$ over the semi-open interval $[\|H^T W H\|, \infty)$. Compared with the solution (1.2) of the standard regularized least-squares problem (1.1), we see that the expression for $\hat{x}$ in (3.13) is distinct in two important ways:

a) First, the weighting matrices $\{Q, W\}$ need to be replaced by corrected versions $\{\widehat{Q}, \widehat{W}\}$. These corrections are defined in terms of the optimal parameter λ^o and they are also dependent on the uncertainty model.

b) Second, the right-hand side of (3.13) contains an additional term that is equal to $\lambda^o E_a^T E_b$. This means that, with λ^o given, the $\hat{x}$ in (3.13) can be interpreted as the solution to a regularized least-squares problem of the form

$$\min_{x} \left(\begin{bmatrix} 1 & x^T \end{bmatrix} \begin{bmatrix} \hat{\lambda}\|E_b\|^2 & -\hat{\lambda}E_b^T E_a \\ -\hat{\lambda}E_a^T E_b & \widehat{Q} \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} + \right.$$

$$\left. + (Ax-b)^T \widehat{W} (Ax-b) \right)$$

with a *cross-coupling* term between x and unity.

The complexity of the solution in the factored uncertainty case is therefore comparable to that of a standard regularized least-squares problem with the additional task of determining the optimal scalar parameter λ^o by minimizing the cost function $G(\lambda)$ over the open interval $[\|H^T W H\|, \infty)$. As is clear from the statement of Theorem 3.1 in the general case, this function is unimodal and has a unique global minimum over the interval of interest. Therefore, the determination of λ^o can be pursued by employing standard search procedures without worries about convergence to undesired local minima.

3.6. Bounded Uncertainties. Consider next the special case of Sec. 2.1 with

$$\phi(x) = \eta\|x\| + \eta_b.$$

In this case, solving for x^o is not so immediate since Eq. (3.8) now becomes, for any nonzero x,

$$x = \left[M(\lambda) + \lambda\eta \left(\eta + \frac{\eta_b}{\|x\|} I \right) \right]^{-1} D(\lambda) . \tag{3.14}$$

Note that x appears on both sides of the equality (except when $\eta_b = 0$, in which case the expression for x is complete in terms of $\{M(\lambda), \lambda, \eta, D(\lambda)\}$). To solve for x in the general case we let $\alpha = \|x\|$ and square the above equation to obtain the scalar equation in α:

$$\alpha^2 - D^T(\lambda) \left[M(\lambda) + \lambda\eta \left(\eta + \frac{\eta_b}{\alpha} \right) I \right]^{-2} D(\lambda) \ = 0 . \tag{3.15}$$

It is shown in Appendix B that a unique solution $\alpha^o(\lambda) > 0$ exists for this equation if, and only if, $\lambda\eta\eta_b \ < \|D(\lambda)\|$. Otherwise, $\alpha^o(\lambda) = 0$. In the former case, the expression for x^o, which is a function of λ, becomes

$$x^o(\lambda) = \left[M(\lambda) + \lambda\eta \left(\eta + \frac{\eta_b}{\alpha^o(\lambda)} \right) I \right]^{-1} D(\lambda) . \tag{3.16}$$

In the latter case we clearly have $x^o(\lambda) = 0$.

Substituting the expression for $x^o(\lambda)$ into (3.16) we get

$$G(\lambda) = x^o(\lambda)^T Q x^o(\lambda) + \big(Ax^o(\lambda) - b\big)^T W(\lambda) \big(Ax^o(\lambda) - b\big) + \lambda\phi^2\big(x^o(\lambda)\big).$$

We are thus led to the following statement.

THEOREM 3.3 (Bounded Uncertainties). *Consider a regularized and weighted robust least-squares problem of the form*

$$\min_x \max_{\substack{\|\delta A\| \leq \eta \\ \|\delta b\| \leq \eta_b}} \left[x^T Q x + \Big((A+\delta A)x - (b+\delta b) \Big)^T W \Big((A+\delta A)x - (b+\delta b) \Big) \right] \tag{3.17}$$

where $\{A, b\}$ are known quantities of appropriate dimensions, $W \geq 0$, and $Q > 0$ are known weighting matrices, and the perturbations $\{\delta A, \delta b\}$ are assumed to be bounded by $\{\eta, \eta_b\}$. Then problem (3.3) has a unique global minimum $\hat{x}$ that can be determined as follows:

1. *Introduce the modified matrices*

$$W(\lambda) \triangleq W + W\big(\lambda I - W\big)^{\dagger} W,$$
$$M(\lambda) \triangleq Q + A^T W(\lambda) A,$$
$$D(\lambda) \triangleq A^T W(\lambda) b.$$

2. *For every λ, define*

$$x^o(\lambda) = \begin{cases} 0 & \text{if } \lambda\eta\eta_b < \|D(\lambda)\| \\ \left[M(\lambda) + \lambda\eta\left(\eta + \frac{\eta_b}{\alpha^o(\lambda)}\right) I \right]^{-1} D(\lambda) & \text{otherwise} \end{cases}$$

where in the second case, $\alpha^o(\lambda)$ is the unique positive solution of the equation

$$\alpha^2 - D^T(\lambda) \left[M(\lambda) + \lambda\eta\left(\eta + \frac{\eta_b}{\alpha}\right) I \right]^{-2} D(\lambda) \ = 0$$

3. *Introduce the cost function*

$$G(\lambda) = x^o(\lambda)^T Q x^o(\lambda) + \big(Ax^o(\lambda) - b\big)^T W(\lambda) \big(Ax^o(\lambda) - b\big) + \lambda\phi^2\big(x^o(\lambda)\big)$$

where $\phi(x) = \eta\|x\| + \eta_b$.

4. *Let λ^o denote the solution of the scalar-valued minimization problem*

$$\lambda^o = \arg \min_{\lambda \geq \|W\|} G(\lambda)$$

5. *Then the optimum solution of (3.3) is*

$$\hat{x} = \begin{cases} 0 & \text{if } \lambda^o\eta\eta_b < \|D(\lambda^o)\| \\ x^o(\lambda^o) & \text{otherwise} \end{cases}$$

where in the second case, the solution $\hat{x}$ admits the form

$$\hat{x} = \left[\widehat{Q} + A^T \widehat{W} A \right]^{-1} A^T \widehat{W} b$$

where the modified weighting matrices $\{\widehat{Q}, \widehat{W}\}$ are obtained from $\{Q, W\}$ via

$$\widehat{Q} \triangleq Q + \lambda^o \eta \left(\eta + \frac{\eta_b}{\alpha^o(\lambda^o)} \right) I, \quad \widehat{W} \triangleq W + W(\lambda^o I - W)^{\dagger} W$$

$\square$

Here again we find that the solution requires that we first determine an optimal nonnegative scalar parameter, λ^o, which corresponds to the minimizing argument of the corresponding function $G(\lambda)$ over the semi-open interval $[\|W\|, \infty)$. In the special case $\eta_b = 0$, we do not need to worry about determining $\alpha^o(\cdot)$ anymore since the expression for the solution $\hat{x}$ simplifies to

$$\hat{x} = \left[\widehat{Q} + A^T \widehat{W} A\right]^{-1} A^T \widehat{W} b$$

with

$$\widehat{Q} = Q + \lambda^o \eta^2 I, \qquad \widehat{W} = W + W(\lambda^o I - W)^\dagger W$$

and $G(\lambda)$ is now defined in terms of

$$x^o(\lambda) = \left[M(\lambda) + \lambda \eta^2 I\right]^{-1} D(\lambda)$$

4. CONCLUDING REMARKS. In this paper we formulated and solved a robust optimization problem that involves a least-squares criterion with both regularization and weighting. The solution turns out to be in regularized form, albeit one that involves corrected weighting matrices. Compared with other robust solutions, the technique does not perform de-regularization and, consequently, does not require existence conditions. This fact is useful for applications that involve real-time operations. In such applications, existence conditions can be a burden since when they fail, the optimality of the solution breaks down. Applications of the proposed methodology to recursive Kalman estimation, quadratic control, and data fusion problems in wireless communications appear in [5]–[7] and they show promising performance.

Appendix A. Properties of $\lambda^o(x)$.

In this appendix, we prove that a solution λ^o of (3.3) exists and is unique for every $x \in \mathbb{R}^n$. We also prove that the function $\lambda^o(x)$ is continuous, a fact that was used in Sec. 3.4.

Before we proceed, however, we remark that the arguments are made simpler if we assume that $H^T W H$ is a diagonal matrix. This can be done without any loss of generality, by a change of variables in y. Indeed, define

$$\bar{y} = Uy, \quad \bar{H} = HU^T,$$

where U is an orthogonal matrix ($U^T U = I$) such that $U(H^T W H)U^T = \Omega = \text{diag}(\omega_i)$; and note that the two sets below are equal:

$$\{\bar{y} : \quad \|\bar{y}\| \leq \phi(x)\} = \{y : \quad \|y\| \leq \phi(x)\},$$

since $\|\bar{y}\| = \|Uy\| = \|y\|$, by the orthogonality of U. In addition, $\bar{H}\bar{y} = Hy$ and $\bar{H}^T W \bar{H} = \Omega$. In the following appendices we shall therefore assume that $H^T W H = \text{diag}(\omega_i)$.

A.1. Solution of (3.3)**.** The entries of the diagonal matrix (see above) $H^T W H = \text{diag}(\omega_i)$ can be ordered such that

$$\|H^T W H\| = \omega_1 = \omega_2 = \cdots = \omega_p > \omega_{p+1} \geq \cdots \geq \omega_m \geq 0, \tag{A.1}$$

where p is the multiplicity of the largest eigenvalue of H^TWH, $\omega_1 = \|H^TWH\|$.

Partition H^TWH as follows,

$$H^TWH = \Omega = \begin{bmatrix} \omega_1 I_p & 0 \\ 0 & \Omega_2 \end{bmatrix},$$

where $\Omega_2 = \mathrm{diag}(\omega_{p+1}, \ldots, \omega_m)$. Define also the vector

$$z(x) = \begin{bmatrix} z_1(x) \\ z_2(x) \end{bmatrix} = H^TW(Ax - b),$$

where $z_1(x) \in \mathbb{R}^p$ and $z_2(x) \in \mathbb{R}^{m-p}$. For every $\lambda > \omega_1$, the matrix $(\lambda I - H^TWH)$ is invertible, and we can define

$$y(\lambda, x) = \begin{bmatrix} y_1(\lambda, x) \\ y_2(\lambda, x) \end{bmatrix} = \begin{bmatrix} \frac{1}{\lambda - \omega_1} z_1(x) \\ (\lambda I_{m-p} - \Omega_2)^{-1} z_2(x) \end{bmatrix} = (\lambda I - H^TWH)^{-1} z(x),$$

with $y_1 \in \mathbb{R}^p$ and $y_2 \in \mathbb{R}^{m-p}$. We found previously that the worst-case disturbance y^o and the Lagrange multiplier λ^o must satisfy (3.3), repeated below,

$$(\lambda^o I - H^TWH)y^o = H^TW(Ax - b), \qquad \|y^o\|^2 = \phi^2(x).$$

If $\{y^o, \lambda^o\}$ are such that $\lambda^o > \omega_1$, these conditions reduce to

$$\|y(\lambda^o, x)\| = \phi(x). \tag{A.2}$$

We now study the behavior of $\|y(\lambda, x)\|^2$, to find when there is a $\lambda^o > \omega_1$ satisfying the above condition. Note that, for fixed x, $\|y(\lambda, x)\|^2$ is a differentiable function of λ, with

$$\begin{aligned} \frac{d\|y(\lambda, x)\|^2}{d\lambda} &= z(x)^T \left(\frac{d}{d\lambda} \begin{bmatrix} \frac{1}{(\lambda-\omega_1)^2} & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & \frac{1}{(\lambda-\omega_m)^2} \end{bmatrix} \right) z(x) \\ &= z(x)^T \begin{bmatrix} -\frac{2}{(\lambda-\omega_1)^3} I_p & 0 \\ 0 & -2(\lambda I_{m-p} - \Omega_2)^{-3} \end{bmatrix} z(x). \end{aligned}$$

The derivative is therefore negative for $z(x) \neq 0$, since the above matrix is negative definite when $\lambda > \omega_1$. Note that when $z(x) = 0$, $y(\lambda, x) = 0$ for all $\lambda > \omega_1$. We show further ahead that in this case the solution will be $\lambda^o = \omega_1$.

FACT 1. *We conclude that for $\lambda > \omega_1$, $\|y(\lambda, x)\|^2$ is a strictly decreasing, continuous function of λ (except when $z(x) = 0$). Therefore, the solution to* (A.2), *if it exists, is unique (see Fig. A.1).*

Consider now the following cases:

1. $z_1(x) \neq 0$ (in this case, $\lim_{\lambda \downarrow \omega_1} \|y(\lambda, x)\| = \infty$);
2. $z_1(x) = 0$, but $\|y_2(\omega_1, x)\| > \phi(x)$ (in this case, $\lim_{\lambda \downarrow \omega_1} \|y(\lambda, x)\| > \phi(x)$);
3. $z_1(x) = 0$ and $\|y_2(\omega_1, x)\| \leq \phi(x)$.

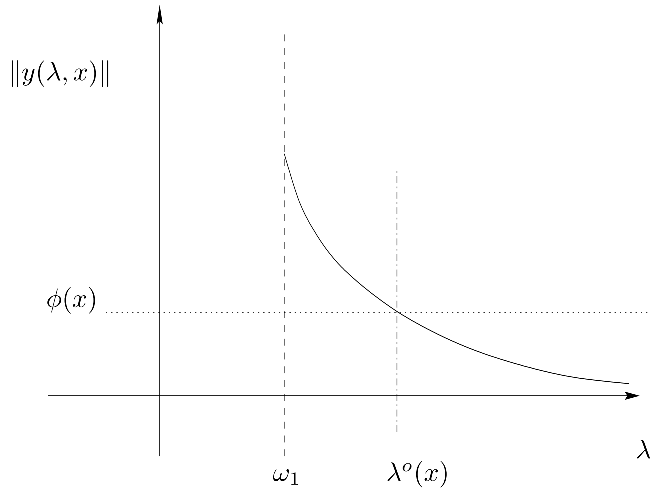

FIG. A.1. *Solution of* (3.3).

FACT 2. *In all cases, the limit of* $\|y(\lambda, x)\|$ *as* λ *goes to infinity is zero. This observation and Fact 1 imply that* (A.2) *will have a solution* $\lambda^o > \omega_1$ *if and only if*

$$\lim_{\lambda \downarrow \omega_1} \|y(\lambda, x)\| > \phi(x),$$

which is the situation in cases 1 and 2. We refer to a point $x \in \mathbb{R}^n$ *for which* $\lambda^o(x) > \omega_1$ *as a* regular *point. A point* x *satisfying the conditions in case 3 will be called a* singular *point.*

We argue now that if x is a singular point (i.e., if case 3 happens), the corresponding Lagrange multiplier must be $\lambda^o(x) = \omega_1$. In case 3, $\|y_2(\omega_1, x)\| \leq \phi(x)$, and condition (A.2) will not be satisfied even in the limit as $\lambda \to \omega_1$. The *original* condition (3.3) can still be satisfied, however, as we show next.

Assume that the conditions in case 3 hold, and choose $\lambda = \omega_1$. Then condition (3.3) reads

$$\begin{bmatrix} 0 & 0 \\ 0 & \omega_1 I_{m-p} - \Omega_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ z_2(x) \end{bmatrix}, \qquad \|y\|^2 = \|y_1\|^2 + \|y_2\|^2 = \phi^2(x).$$

The first condition is satisfied for

$$y_2 = (\omega_1 I_{m-p} - \Omega_2)^{-1} z_2(x),$$

where the inverse exists since by definition $\omega_1 > \omega_{p+1} = \|\Omega_2\|$. The second condition in case 3 is $\|y_2\| \leq \phi(x)$. Therefore, to satisfy the norm condition in (3.3), we just choose any $y_1 \in \mathbb{R}^p$ whose norm satisfies

$$\|y_1\|^2 = \phi^2(x) - \left\| (\omega_1 I_{m-p} - \Omega_2)^{-1} z_2(x) \right\|^2.$$

LEMMA A.1 (Solution to the maximization problem). *If a point* $x \in \mathbb{R}^n$ *is regular, viz.,*

$$z_1(x) \neq 0 \qquad \text{and} \qquad \lim_{\lambda \downarrow \omega_1} \|y(\lambda, x)\| > \phi(x), \tag{A.3}$$

then the Lagrange multiplier at the maximum $\lambda^o(x) > \omega_1$ is the unique solution to (A.2). *(although the first condition implies the second, we want to state it explicitly here for further reference). In this case, the worst-case disturbance*

$$y^o = y(\lambda^o, x) = (\lambda^o I - H^T W H)^{-1} H^T W (Ax - b), \qquad \|y^o\| = \phi(x),$$

is also unique.

On the other hand, if x is singular, viz.,

$$z_1(x) = 0 \qquad\qquad \textit{and} \qquad\qquad \lim_{\lambda \downarrow \omega_1} \|y(\lambda, x)\| \leq \phi(x), \tag{A.4}$$

then the Lagrange multiplier will be $\lambda^o(x) = \omega_1$. Now the worst-case disturbance is no longer unique — any disturbance of the form below will achieve the maximum:

$$y^o = \begin{bmatrix} y_1^o \\ y_2^o \end{bmatrix},$$

with $y_1^o \in I\!R^p$, and

$$y_2^o = (\omega_1 I_{m-p} - \Omega_2)^{-1} z_2(x), \qquad\qquad \|y_1^o\|^2 = \phi^2(x) - \|y_2^o\|^2.$$

In addition, using the pseudo-inverse notation we can write

$$y^o = (\omega_1 I - H^T W H)^\dagger H^T W (Ax - b) + \begin{bmatrix} y_1^o \\ 0 \end{bmatrix},$$

where

$$\|y_1^o\|^2 = \phi^2(x) - \left\|(\omega_1 I - H^T W H)^\dagger H^T W (Ax - b)\right\|^2.$$

□

A.2. Continuity of $\lambda^o(x)$. This property of $\lambda^o(x)$ was invoked in Sec. 3.4 to argue that $G(\lambda)$ is unimodal (see Thm. 3.1). We again treat regular and singular points separately.

Regular points: By definition, at a regular point $\tilde{x}$, $\lambda^o(\tilde{x}) > \omega_1$ and

$$f(\lambda^o, \tilde{x}) \triangleq \phi^2(\tilde{x}) - (A\tilde{x} - b)^T W H (\lambda^o(\tilde{x}) I - H^T W H)^{-2} H^T W (A\tilde{x} - b) = 0.$$

Now from the implicit function theorem [16], the function $\lambda^o(x)$ defined by the above condition is continuous at a given point x if the gradient $\nabla_\lambda f(\lambda, x)$ is nonzero at $\lambda = \lambda^o$. To check this condition, compute the partial derivative

$$\frac{\partial f(\lambda, x)}{\partial \lambda} = 2(Ax - b)^T W H (\lambda I - H^T W H)^{-3} H^T W (Ax - b).$$

At a regular point $\tilde{x}$, recall that we must have either $z_1(\tilde{x}) \neq 0$ or $\|y_2(\omega_1, x)\| > \phi(\tilde{x})$ (se eq. (A.3)). Both these conditions would be violated if $A\tilde{x} - b = 0$, so our assumption that $\tilde{x}$ is regular implies that $A\tilde{x} - b \neq 0$. With this fact, and noting that $(\lambda^o(\tilde{x}) I - H^T W H)^{-3} > 0$ (from the regularity of $\tilde{x}$), we conclude that

$$\frac{\partial f(\lambda^o(x), x)}{\partial \lambda} > 0,$$

satisfying the condition of the implicit function theorem. We have thus proved that $\lambda^o(x)$ is continuous at any regular point $\tilde{x}$.

Singular points: Let now $\bar{x} \in \mathrm{IR}^n$ be a singular point. We prove the continuity of $\lambda^o(\cdot)$ at $x = \bar{x}$ from the definition. Given an $\epsilon > 0$, we shall find $\delta(\epsilon) > 0$ such that ($\lambda^o(\bar{x}) = \omega_1$)

$$\|x - \bar{x}\| < \delta \Rightarrow |\lambda^o(x) - \omega_1| < \epsilon.$$

To find such a δ, we shall need some properties of singular points and of $\phi^2(x)$ and $z(x)$ — if $\bar{x}$ is a singular point, then from the previous sections we have:

1. $\phi(\bar{x}) \geq \|(\omega_1 I - H^T W H)^\dagger H^T W(A\bar{x} - b)\|$,
2. $z_1(\bar{x}) = 0$,
3. $\|y_2(\lambda, x)\|^2$ is continuous in (λ, x) on $(\omega_1, \infty) \times \mathrm{IR}^n$, and continuous and strictly decreasing in $\lambda > \omega_1$ for fixed x.

Recalling that $z(x) = H^T W(Ax - b)$, we also have:

4. $\|z_1(x)\|^2$ and $\|z_2(x)\|^2$ are continuous functions for all $x \in \mathrm{IR}^n$.

Finally, we must make an assumption on the uncertainty bound, namely we assume that

5. $\phi^2(x)$ is continuous for all $x \in \mathrm{IR}^n$ (in fact, this follows from our assumption in Thm. 3.1 that $\phi^2(x)$ is convex).

Two situations may occur:

A. There exists a neighborhood $N(\bar{x})$ whose points are all singular, i.e., $x \in N(\bar{x}) \Rightarrow \lambda^o(x) = \omega_1$. In this situation, the continuity of $\lambda^o(\cdot)$ at $\bar{x}$ is trivial;

B. Every neighborhood of $\bar{x}$ contains a regular point x^*.

Let us consider the second case. We now find a ball $B_\delta(\bar{x}) = \{x : \|x - \bar{x}\| < \delta\}$ for which

$$\sup_{x \in B_\delta(\bar{x})} \lambda^o(x) < \lambda^o(\bar{x}) + \epsilon = \omega_1 + \epsilon.$$

The above properties and assumptions imply that for any K_1, K_2 and K_3, it is possible to find δ_1, δ_2 and $\delta_3 > 0$ such that

$$\begin{aligned} \|x - \bar{x}\| < \delta_1 &\Rightarrow \left| \|z_1(x)\|^2 - \underbrace{\|z_1(\bar{x})\|^2}_{=0} \right| < \frac{\epsilon}{K_1}, \\ \|x - \bar{x}\| < \delta_2 &\Rightarrow \|z_2(x) - z_2(\bar{x})\|^2 < \frac{\epsilon}{K_2}, \\ \|x - \bar{x}\| < \delta_3 &\Rightarrow \left|\phi^2(x) - \phi^2(\bar{x})\right| < \frac{\epsilon}{K_3}. \end{aligned} \tag{A.5}$$

Choose the K_i such that

$$K_1 = \frac{\bar{K}_1}{\epsilon^2}, \qquad \frac{1}{\bar{K}_1} + \frac{1}{(\omega_1 - \omega_{p+1})^2 K_2} + \frac{1}{K_3} < \frac{\|y_2(\omega_1, \bar{x})\|^2}{2(\omega_1 - \omega_{p+1})}, \tag{A.6}$$

and let $\delta = \min\{\delta_1, \delta_2, \delta_3\}$. Remark that the K_i cannot be choosen to satisfy (A.6) only if $y_2(\omega_1, \bar{x}) = 0$. We shall assume for now that $y_2(\omega_1, \bar{x}) \neq 0$, and treat the other case later.

Since we are studying case B, let x^* be a regular point in $B_\delta(\bar{x})$. As a regular point, x^* satisfies $\lambda^o(x^*) > \omega_1$ and

$$\|y(\lambda^o(x^*), x^*)\| = \phi(x^*) \qquad \text{and} \qquad \lim_{\lambda \downarrow \omega_1} \|y(\lambda, x^*)\| > \phi(x^*)$$

(the limit may be infinity).

We now show that for $\lambda^* = \omega_1 + \epsilon$, it necessarily holds that

$$\|y(\lambda^*, x^*)\| < \phi(x^*),$$

which means that $\omega_1 < \lambda^o(x^*) < \omega_1 + \epsilon$, which is our desired result. Let us then evaluate $\|y(\lambda, x^*)\|^2$:

$$\begin{aligned}\|y(\lambda, x^*)\|^2 &= \left\|(\lambda I - H^T W H)^{-1} \begin{bmatrix} z_1(x^*) \\ z_2(x^*) \end{bmatrix}\right\|^2 = \\ &= (\lambda - \omega_1)^{-2}\|z_1(x^*)\|^2 + \left\|\begin{bmatrix} \lambda - \omega_{p+1} & \dots & 0 \\ & \ddots & \\ 0 & \dots & \lambda - \omega_m \end{bmatrix}^{-1} z_2(x^*)\right\|^2.\end{aligned}$$

We use (A.5) to bound these norms,

$$\|y(\lambda, x^*)\|^2 < \frac{\epsilon}{(\lambda - \omega_1)^2 K_1} + \|y_2(\lambda, \bar{x})\|^2 + \frac{\epsilon}{(\lambda - \omega_{p+1})^2 K_2}, \tag{A.7}$$

where we used $\|\operatorname{diag}\big((\lambda - \omega_j)^{-1}\big)\| = (\lambda - \omega_{p+1})^{-1}$, and $z_2(x^*) = z_2(\bar{x}) + \big(z_2(x^*) - z_2(\bar{x})\big)$.

To bound the second term, write

$$\begin{aligned} y_2(\lambda, \bar{x}) &= \begin{bmatrix} \frac{\omega_1 - \omega_{p+1}}{\lambda - \omega_{p+1}} & & \\ & \ddots & \\ & & \frac{\omega_1 - \omega_m}{\lambda - \omega_{p+1}} \end{bmatrix} \begin{bmatrix} \omega_1 - \omega_{p+1} & & \\ & \ddots & \\ & & \omega_1 - \omega_m \end{bmatrix}^{-1} z_2(\bar{x}) \\ &= \begin{bmatrix} \frac{\omega_1 - \omega_{p+1}}{\lambda - \omega_{p+1}} & & \\ & \ddots & \\ & & \frac{\omega_1 - \omega_m}{\lambda - \omega_{p+1}} \end{bmatrix} y_2(\omega_1, \bar{x}) \triangleq P(\lambda) y_2(\omega_1, \bar{x}). \end{aligned}$$

Let $\lambda = \omega_1 + \epsilon$, then the largest element of $P(\lambda)$ will be (if ϵ is small enough)

$$\begin{aligned} \frac{\omega_1 - \omega_{p+1}}{\omega_1 + \epsilon - \omega_{p+1}} &= 1 - \frac{\epsilon}{\omega_1 - \omega_{p+1}} + \frac{\epsilon^2}{(\omega_1 - \omega_{p+1})^2} - \underbrace{\left(\frac{\epsilon^3}{(\omega_1 - \omega_{p+1})^3} + \dots\right)}_{\geq 0 \text{ if } \epsilon/(\omega_1 - \omega_{p+1}) < 1/2} \\ &< 1 - \frac{\epsilon}{\omega_1 - \omega_{p+1}} + \frac{\epsilon^2}{(\omega_1 - \omega_{p+1})^2} < 1 - \frac{\epsilon}{2(\omega_1 - \omega_{p+1})}, \end{aligned}$$

and thus,

$$\begin{aligned}\|y_2(\lambda,\bar{x})\|^2 &< \left(1-\frac{\epsilon}{2(\omega_1-\omega_{p+1})}\right)\|y_2(\omega_1,\bar{x})\|^2\\ &< \phi^2(\bar{x}) - \frac{\epsilon}{2(\omega_1-\omega_{p+1})}\|y_2(\omega_1,\bar{x})\|^2\\ &< \phi^2(x^*) + \frac{\epsilon}{K_3} - \frac{\epsilon}{2(\omega_1-\omega_{p+1})}\|y_2(\omega_1,\bar{x})\|^2.\end{aligned}$$

Using this bound in (A.7), we obtain,

$$\begin{aligned}\|y(\omega_1+\epsilon,x^*)\|^2 &< \frac{\epsilon}{\epsilon^2 K_1} + \frac{\epsilon}{(\omega_1+\epsilon-\omega_{p+1})^2 K_2} + \phi^2(x^*) + \\ &\quad + \frac{\epsilon}{K_3} - \frac{\epsilon}{2(\omega_1-\omega_{p+1})}\|y_2(\omega_1,\bar{x})\|^2 < \phi^2(x^*),\end{aligned}$$

where the last inequality follows from our choice of the K_i in (A.6). The inequality shows that $\lambda^o(x^*) < \omega_1+\epsilon$. Since the above argument holds for any regular point in $B_\delta(\bar{x})$, we have

$$x \in B_\delta(\bar{x}) \Rightarrow \omega_1 \le \lambda^o(x) < \omega_1+\epsilon,$$

which proves the continuity of $\lambda^o(\cdot)$ at singular points $\bar{x}$ for which $y_2(\omega_1,\bar{x}) \neq 0$.

Finally we consider singular points $\bar{x}$ for which $y_2(\omega_1,\bar{x})$ is zero. In this case, $A\bar{x}-b$ is necessarily zero (since $z_1(\bar{x})=0$ for singular points). Again, two situations may happen:

i. $\phi(\bar{x}) = 0$ — in this situation, the solution of the maximization problem is trivial, as the uncertainty will be identically zero;

ii. $\phi(\bar{x}) > 0$ — now, from the continuity of $\|y_2(\lambda,\bar{x})\|^2$ and of $\phi^2(x)$, there exists a ball $B_{\delta_4}(\bar{x})$ such that

$$x \in B_{\delta_4}(\bar{x}) \Rightarrow \|y_2(\omega_1,x)\|^2 < \frac{\phi^2(\bar{x})}{2}.$$

With this inequality, if we choose K_1 and K_3 such that

$$K_1 = \frac{\bar{K}_1}{\epsilon^2}, \qquad \frac{1}{\bar{K}_1} + \frac{1}{2K_3} < \frac{\phi^2(\bar{x})}{2} - \frac{1}{2K_3},$$

then if $\delta = \min\{\delta_1,\delta_3,\delta_4\}$, for all regular points $x^* \in B_\delta(\bar{x})$, we can replace (A.7) by the simpler expression

$$\|y(\lambda,x^*)\|^2 < \frac{\epsilon}{(\lambda-\omega_1)^2 K_1} + \|y_2(\lambda,x^*)\|^2 < \frac{\epsilon}{(\lambda-\omega_1)^2 K_1} + \frac{\phi^2(\bar{x})}{2},$$

where we used the fact that $\|y_2(\lambda,x^*)\|^2$ is decreasing with λ. With our choice of K_1, for $\lambda = \omega_1+\epsilon$ we obtain

$$\begin{aligned}\|y(\omega_1+\epsilon,x^*)\|^2 &< \frac{\epsilon}{\bar{K}_1} + \frac{\phi^2(\bar{x})}{2} < \frac{\epsilon}{\bar{K}_1} + \frac{\epsilon}{2K_3} + \frac{\phi^2(x^*)}{2}\\ &< \frac{\phi^2(\bar{x})}{2} - \frac{\epsilon}{2K_3} + \frac{\phi^2(x^*)}{2} < \phi^2(x^*),\end{aligned}$$

which implies that $\lambda^o(x^*) < \omega_1+\epsilon$.

Appendix B. Computation of $x^o(\lambda)$ in the bounded uncertainty case. With $\phi(x) = \eta\|x\| + \eta_b$, the vector x that achieves the minimum on the right-hand side of (3.7) is the solution to the equation

$$x \stackrel{\Delta}{=} x^o(\lambda) = \left(Q + A^T W(\lambda) A + \lambda\eta^2 I + \frac{\lambda\eta\eta_b}{\|x\|} I\right)^{-1} A^T W(\lambda) b. \tag{B.1}$$

The value of x is clearly a function of λ. Observe however that this equation defines x implicitly since x appears on both sides of the equality. To proceed, we consider two cases.

(1) $\eta_b = 0$. In this case, the expression for $x^o(\lambda)$ collapses to

$$x^o(\lambda) = \left[Q + \lambda\eta^2 I + A^T W(\lambda) A\right]^{-1} A^T W(\lambda) b.$$

That is, the term $\|x\|$ disappears from the right-hand side of (B.1). Consequently, this expression defines $x^o(\lambda)$ explicitly.

(2) $\eta_b \neq 0$. In this case, the term $\|x\|$ does not disappear from the right-hand side of (B.1). In order to solve for x we proceed as follows. First, we introduce the scalar $\alpha = \|x\|$ and square both sides of (B.1). This leads to the following nonlinear equation in α:

$$\alpha^2 - D^T(\lambda)\left[M(\lambda) + \lambda\eta\left(\eta + \frac{\eta_b}{\alpha}\right) I\right]^{-2} D(\lambda) \ = 0, \tag{B.2}$$

where

$$M(\lambda) = Q + A^T W(\lambda) A, \quad D(\lambda) = A^T W(\lambda) b.$$

The value of α is again dependent on λ. The following result indicates that the solution to the above nonlinear equation for α is either at $\alpha = 0$ or at a *unique* positive value.

LEMMA B.1. *Let $x^o(\lambda)$ minimize the inner cost on the right-hand side (RHS) of (3.7). If, and only if, $\lambda\eta\eta_b < \|D(\lambda)\|$, the norm $\|x^o(\lambda)\|$ is equal to the the unique positive solution of equation (B.2), $\alpha^o(\lambda)$. Otherwise, the solution to the minimization problem is $x^o(\lambda) = 0$, i.e., $\alpha^o(\lambda) = 0$.*

Proof. We shall first find the solutions of (B.2) when $\lambda\eta\eta_b < \|D(\lambda)\|$, afterwards we relate these conditions to the solutions of the inner minimization problem on the RHS of (3.7).

Introduce the SVD of the symmetric positive-definite matrix $M(\lambda)$, say $M(\lambda) = U\Sigma U^T$, where $\{U, \Sigma\}$ are also dependent on λ. We denote the entries of Σ by $\{\sigma_i\}$. Substituting this decomposition into the left side of (B.2), it reduces to the function

$$f(\alpha) \stackrel{\Delta}{=} \alpha^2 - \sum_{i=1}^{n} \frac{\bar{d}_i^2}{[\sigma_i + \lambda\eta(\eta + \frac{\eta_b}{\alpha})]^2}$$

where the $\{\bar{d}_i\}$ denote the entries of the transformed vector $U^T D(\lambda)$. We are seeking the roots of $f(\alpha)$.

Note that $\alpha = 0$ is always a solution if $\eta_b > 0$. Let us search for a solution $\alpha > 0$. Assuming $\alpha > 0$, we can write

$$f(\alpha) = \alpha^2 \left[1 - \sum_{i=1}^{n} \frac{\bar{d}_i^2}{(\sigma_i \alpha + \lambda \eta^2 \alpha + \lambda \eta \eta_b)^2} \right] \triangleq \alpha^2 g(\alpha),$$

where we introduced the function $g(\alpha)$. Taking the limits as $\alpha \to 0$ and $\alpha \to \infty$ we find that

$$\lim_{\alpha \to 0} g(\alpha) = 1 - \sum_{i=1}^{n} \frac{\bar{d}_i^2}{(\lambda \eta \eta_b)^2}$$

$$\lim_{\alpha \to \infty} g(\alpha) = 1 > 0$$

Therefore, $g(\alpha)$ will have a zero for $\alpha > 0$ if, and only if, the first limit above is negative, i.e., if $\{\lambda, \eta, \eta_b\}$ satisfy

$$\lambda \eta \eta_b < \|D(\lambda)\|.$$

In addition, since the derivative of $g(\alpha)$ with respect to α is

$$\frac{dg(\alpha)}{d\alpha} = 2 \sum_{i=1}^{n} \frac{\bar{d}_i^2 (\lambda \eta^2 + \sigma_i)}{[\alpha \sigma_i + \lambda \eta (\alpha \eta + \eta_b)]^3} > 0,$$

we conclude that the root is necessarily unique.

Let us verify now that this root really corresponds to the solution of our minimization problem. The point that may cause trouble is $x = 0$, where the cost function is not differentiable. The cost at this point is

$$C(0, \lambda) = b^T W(\lambda) b + \lambda \eta_b^2.$$

If we move a little away from $x = 0$, say to $x = \delta x$, then the cost becomes

$$\begin{aligned} C(\delta x, \lambda) = {} & (A\delta x - b)^T W(\lambda)(A\delta x - b) + \lambda(\eta \|\delta x\| + \eta_b)^2 = \\ & = b^T W(\lambda) b - 2\delta x^T A^T W(\lambda) b + \delta x^T A^T W(\lambda) A \delta x + \lambda \eta^2 \|\delta x\|^2 + \\ & + 2\lambda \eta \eta_b \|\delta x\| + \lambda \eta_b^2, \end{aligned}$$

and thus, for small δx,

$$\begin{aligned} C(\delta x, \lambda) = {} & C(0, \lambda) - 2\delta x^T D(\lambda) + 2\lambda \eta \eta_b \|\delta x\| + O(\|\delta x\|^2) \geq \\ & \geq C(0, \lambda) - 2\|\delta x\| \left(\|D(\lambda)\| - 2\lambda \eta \eta_b \right) + O(\|\delta x\|^2). \end{aligned}$$

We conclude that, for small δx, $C(\delta x, \lambda)$ is smaller than $C(0, \lambda)$ if and only if $\lambda \eta \eta_b < \|D(\lambda)\|$. In this situation, $x = 0$ cannot be a minimum of $C(x, \lambda)$, and the optimum $x^o(\lambda)$ must be such that its norm solves (B.2) with $\alpha^o(\lambda) > 0$.

On the other hand, if $\lambda \eta \eta_b \geq \|D(\lambda)\|$, the cost for small δx satisfies $C(\delta x, \lambda) > C(0, \lambda)$ (we can include the case when equality holds, since the terms in $O(\|\delta x\|^2)$ above are all positive). The point $x = 0$ must thus be a local minimum to $C(x, \lambda)$.

Since we know that this cost is strictly convex in x for fixed λ, $x = 0$ must be the global minimum.

$\square$

Appendix C. A result on convex optimization problems.

In this appendix we establish a result that was used to show that $G(\lambda)$ is unimodal. Let $f(x, y)$ be a real function of variables $x \in X,\ y \in Y$. We shall study the problem

$$\min_{x \in X, y \in Y} f(x, y).$$

Define the functions

$$\begin{aligned} g:\ & X \to \mathbb{R} \\ g(x) = & \min_{y \in Y} f(x, y), \end{aligned}$$

and

$$\begin{aligned} h:\ & Y \to \mathbb{R} \\ h(y) = & \min_{x \in X} f(x, y). \end{aligned}$$

We denote by (x_{op}, y_{op}) one of the (possibly many) global minimum points of $f(x, y)$ in $X \times Y$, by x_g one of the global minima of $g(x)$ in X, and by y_h one of the global minima of $h(y)$ in Y. With these definitions, we prove the following results.

LEMMA C.1. *If any of the minima below is attainable, then it holds that*

$$\min_{(x,y) \in X \times Y} f(x, y) = \min_{x \in X} g(x) = \min_{y \in Y} h(y), \quad \textit{and} \quad (x_{op}, y_{op}) = (x_g, y_{x_g}) = (x_{y_h}, y_h).$$

Proof. This is a classic result. To prove it, simply notice that all points (x, y) are compared in the minimization of all three functions above. If the minima are not attainable, the result is still true if we substitute the min by inf.

$\square$

LEMMA C.2. *Let X, Y be subsets of a metric space, and assume that the functions below,*

$$\begin{aligned} f(\bar{x}, y):\ & Y \to \mathbb{R} && \textit{for all } \bar{x} \in X \textit{ fixed}, \\ f(x, \bar{y}):\ & X \to \mathbb{R} && \textit{for all } \bar{y} \in Y \textit{ fixed}, \\ g(x):\ & X \to \mathbb{R}, && \end{aligned}$$

have unique global minima, and are unimodal in their respective domains, i.e., assume that each function does not admit local minima different than their global minima.

We now define the functions

$$\begin{aligned} y_m:\ & X \to Y \\ y_m(x) = & \arg\min_{y \in Y} f(x, y), \end{aligned}$$

and

$$x_m : Y \to X$$
$$x_m(y) = \arg\min_{x \in X} f(x,y).$$

Under these conditions, and if $y_m(x)$ is continuous in X, then $h(y)$ is also unimodal.

Proof. $y_m(x)$ is a function, since, by hypothesis, $f(\bar{x}, y)$, $\bar{x}$ fixed, is unimodal in Y. A similar argument implies that $x_m(y)$ is a function. Now assume (by contradiction) that $h(y)$ is not unimodal, i.e., it has a local minimum at $y_l \neq y_{op}$. This means that there is an open ball $B_\delta(y_l) \in Y$ such that y_l is the global minimum of $h(y)$ *inside the ball.*

From the previous lemma, we find that

$$\min_{(x,y) \in X \times B_\delta(y_l)} f(x,y) = \min_{y \in B_\delta(y_l)} h(y),$$

and $(x_l, y_l) = (x_m(y_l), y_l)$. This implies that (x_l, y_l) is a local minimum of $f(x,y)$ in $X \times Y$ different from the global minimum (x_{op}, y_{op}). In particular, this means that, fixing x_l, $f(x_l, y)$ has a (local) minimum at $y = y_l$.

Since we assumed that $f(x_l, y)$ is unimodal, it must hold that $y_m(x_l) = y_l$. Function $y_m(\cdot)$ is continuous on X by hypothesis, thus there exists a ball $B_\gamma(x_l)$ whose points satisfy

$$x \in B_\gamma(x_l) \Rightarrow y_m(x) \in B_\delta(y_l).$$

Since (x_l, y_l) is the global minimum of f in $X \times B_\delta(y_l)$, x_l must be the global minimum of $g(x) = f(x, y_m(x))$ in $B_\gamma(x_l)$, that is, x_l is a local minimum of $g(x)$.

Finally, note that we assumed that $y_l \neq y_{op}$. Since $y_m(x_l) = y_l$ and $y_m(x_{op}) = y_{op}$, we must have $x_l \neq x_{op}$ — that is, we found a local minimum of $g(x)$ different than x_{op}, contradicting our initial assumption that $g(x)$ is unimodal.

$\square$

REFERENCES

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

[2] 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, 1999.

[3] 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.

[4] S. Chandrasekaran, G. Golub, M. Gu, and A. H. Sayed. Parameter estimation in the presence of bounded data uncertainties. *SIAM J. Matrix Analysis and Applications*, 19(1):235–252, Jan. 1998.

[5] V. H. Nascimento and A. H. Sayed. Optimal state regulation for uncertain state-space models. In *Proc. ACC*, San Diego, CA, June 1999.

[6] A. H. Sayed, "A framework for state-space estimation with uncertain models," to appear in *IEEE Transactions on Automatic Control*, vol. 46, no. 9, Sep. 2001.

[7] A. H. Sayed, T. Y. Al-Naffouri, and T. Kailath, "Robust estimation for uncertain models in a data fusion scenario," *Proc. IFAC System Identification Symposium*, Santa Barbara, CA, June 2000.

[8] T. Basar and G. J. Olsder. *Dynamic Noncooperative Game Theory.* Academic Press, 1982.

[9] A. H. Sayed and V. H. Nascimento. Design criteria for uncertain models with structured and unstructured uncertainties, in *Robustness in Identification and Control*, A. Garulli, A. Tesi, and A. Vicino, editors, Proceedings of the *Workshop on Robust Identification and Control* (Siena, Italy, July 1998), vol. 245, pp. 159–173, Springer Verlag, London, 1999.

[10] A. H. Sayed, V. H. Nascimento, and S. Chandrasekaran. Estimation and control with bounded data uncertainties. *Linear Algebra and Its Applications*, vol. 284, pp. 259–306, Nov. 1998.
[11] A. H. Sayed and S. Chandrasekaran. Parameter estimation with multiple sources and levels of uncertainties. *IEEE Transactions on Signal Processing*, vol. 48, no. 3, pp. 680–692, Mar. 2000.
[12] Y. Cheng and B. L. de Moor. Robustness analysis and control system design for hydraulic servo system. *IEEE Trans. Contr. Sys. Technol.*, vol. 2, pp. 183–198, 1994.
[13] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, NJ, 2000.
[14] A. E. Bryson and Y. C. Ho, *Applied Optimal Control*, Blaisdell, Waltham, MA, 1969.
[15] R. Fletcher. *Practical Methods of Optimization*. Wiley, 1987.
[16] W. Rudin. *Principles of Mathematical Analysis*. 3rd. ed., McGraw-Hill, 1976.
[17] D. G. Luenberger. *Optimization by Vector Space Methods*. Wiley, 1969.