A UNIQUENESS RESULT CONCERNING A ROBUST REGULARIZED LEAST-SQUARES SOLUTION

ALI H. SAYED AND HONG CHEN

ELECTRICAL ENGINEERING DEPARTMENT
UNIVERSITY OF CALIFORNIA
LOS ANGELES, CA 90095
PHONE: (310)267-2142
FAX: (310)206-8495
E-MAIL: SAYED@EE.UCLA.EDU
URL: WWW.EE.UCLA.EDU/ASL

Abstract. In solving a robust version of regularized least-squares with weighting, a certain scalar-valued optimization problem is required in order to determine the regularized robust solution and the corresponding robustified weighting parameters. This letter establishes that the required optimization problem does not have local, non-global minima over the interval of interest. This property is proved by resorting to a useful Schur complementation argument. The result is reassuring in that it demonstrates that the robust design procedure is well defined and that its optimal global solution can be determined without concerns about local minima.

Key words. Regularization, least-squares, Schur complement, game problem, uncertainty, optimization.

1. INTRODUCTION. Many estimation and control techniques rely on quadratic objective functions and employ, in one way or another, regularized least-squares costs of the form

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

Here x denotes the unknown parameter vector, $x^T Q x$ is a regularization term with $Q = Q^T > 0$, and $W = W^T \geq 0$ is a weighting matrix. The unknown x is $n-$dimensional, while A is $N \times n$ and b is $N \times 1$. Both A and b are assumed to be known with A called the data matrix and b the measurement vector. Although not necessary, all quantities are assumed to be real-valued. The solution of (1.1) is

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

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

In practice, the nominal data $\{A, b\}$ are often subject to uncertainties, which can degrade the performance of the otherwise optimal estimator (1.2). For example, if the actual data matrix were $(A + \delta A)$, for some unknown perturbation δA, then the estimator (1.2) that is designed based on A alone, and without accounting for the existence of δA, can perform poorly.

In the works [1, 2], and motivated by earlier studies in [3]–[6] for standard least-squares problems ($Q = 0$ and $W = I$), we have formulated a robustified version of the regularized and weighted objective function (1.1) that can account for uncertainties

in the data $\{A, b\}$. It turns out that nontrivial choices for $\{Q, W\}$ require special care and a technique was developed for this case in [1, 2].

Two situations that can be modeled in this way arise in the context of state regulation [1] and state estimation for state-space models with parametric uncertainties [7]. These applications are motivated in Secs. 3.1 and 3.2. In both cases, the design requires solving a robust version of the regularized least-squares problem (1.2). This robust formulation is described in Sec. 2, where it is shown that a key step to its solution involves minimizing a certain scalar-valued function $G(\lambda)$ over a certain interval $[\lambda_l, \infty)$ – see Eq. (2.10). This letter establishes that the function $G(\lambda)$ does not have local, non-global minima over the interval of interest. The result is reassuring in that it demonstrates that the optimal global solution of the robust design procedure can be determined without concerns about local minima.

The function $G(\lambda)$, which is described in Theorem 2.1, behaves like a secular function; secular functions are rational functions with poles at eigenvalues of a certain model-dependent matrix – see, e.g., [8]–[11] and the references therein. Connections with secular functions in the context of robust designs were exploited earlier in [5, 12]. However, in this letter we choose to study the behavior of $G(\lambda)$ by resorting instead to a useful and simplifying Schur complementation argument. Rather than examining the signatures of the first- and second-order derivatives of $G(\lambda)$, which would be a difficult task in this problem, we show how the second-order derivative is related to the first-order derivative via a Schur complementation step. This fact is then used to conclude that $G(\lambda)$ has a unique global minimum over the interval of interest. Details are provided in Sec. 4.

2. ROBUST REGULARIZED LEAST-SQUARES. Returning to the least-squares problem (1.1), let $J(x, y)$ denote a two-variable cost function of the form

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

where the residual $R(x, y)$ 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 is an $m \times 1$ unknown perturbation vector. Comparing the expression for $R(x, y)$ with the term $(Ax - b)^T W (Ax - b)$ that appears in (1.1), we see that we are modeling sources of uncertainties in A and b by the additional term Hy. The choice of the matrix H provides the designer with the freedom of restricting the uncertainty y to certain range spaces. While y itself is not known, we shall assume that what is known is a bound on its Euclidean norm, say

$$\|y\| \leq \phi(x)$$

for some known (linear or nonlinear) nonnegative function $\phi(x)$. Observe that the bound on y is allowed to depend on the parameter x.

Consider now the problem of solving

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

where the notation $\|\cdot\|$ stands for the Euclidean norm of its vector argument (or the maximum singular value of a matrix argument). Problem (2.1) can be interpreted as

a *constrained* two-player game problem, with the designer trying to pick an estimate $\hat{x}$ that minimizes the cost while the opponent $\{y\}$ tries to maximize the cost. The game problem is constrained since it imposes a bound (through $\phi(x)$) on how large (or how damaging) the opponent can be. Observe further that the strength of the opponent can vary with the choice of x. We shall assume in the sequel that H and $\phi(x)$ are not identically zero,

$$H \neq 0 \quad \text{and} \quad \phi(\cdot) \neq 0, \tag{2.2}$$

since if either is zero, then the game problem (2.1) trivializes to the standard regularized least-squares problem (1.1).

However, in this letter we focus on the following specialization of (2.1) because this case is simpler to analyze while also arising in applications (see Secs. 3.1 and 3.2):

$$\min_{x} \max_{\delta A, \delta b} \left[x^T Q x + \left((A + \delta A)x - (b + \delta b) \right)^T W \left((A + \delta A)x - (b + \delta b) \right) \right] \tag{2.3}$$

Here $\{\delta A\}$ denotes an $N \times n$ perturbation matrix to the nominal matrix A, δb denotes an $N \times 1$ perturbation vector to the nominal vector b, and $\{\delta A, \delta b\}$ are assumed to satisfy a model of the form

$$\begin{bmatrix} \delta A & \delta b \end{bmatrix} = H \Delta \begin{bmatrix} E_a & E_b \end{bmatrix} \tag{2.4}$$

where Δ is an arbitrary contraction, $\|\Delta\| \leq 1$, and $\{H, E_a, E_b\}$ are known quantities of appropriate dimensions (say, H is $N \times m$, Δ is $m \times m$, E_a is $m \times n$ and E_b is $m \times 1$). Perturbation models of the form (2.4) are common in robust filtering and control and can arise from tolerance specifications on physical parameters (see, e.g., [16]).

In order to see how (2.3)–(2.4) is a special case of (2.1), we rewrite the cost function in (2.3) as

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

so that with Hy defined as

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

and y as $y = \Delta(E_a x - E_b)$, problem (2.3)–(2.4) reduces to

$$\min_{x} \max_{\|y\| \leq \|E_a x - E_b\|} \left[x^T Q x + \left(Ax - b + Hy \right)^T W \left(Ax - b + Hy \right) \right] \tag{2.5}$$

which is a special case of (2.1) for the particular choice

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

Conversely, we can verify that every problem of the form (2.5) reduces to one of the form (2.3)–(2.4) so that both formulations (2.3)–(2.4) and (2.5) are equivalent. Indeed, for any y satisfying $\|y\| \leq \|E_a x - E_b\|$, there should exist a contraction Δ relating the vectors y and $E_a x - E_b$, say $y = \Delta(E_a x - E_b)$. Now choose $\delta A = H \Delta E_a$ and $\delta b = H \Delta E_b$, and problem (2.5) reduces to (2.3)–(2.4).

The formulation (2.1) is more general than (2.3)–(2.4) in that it allows for other classes of perturbations through the choice of the function $\phi(x)$ — see [2]. In this

letter we focus on the form (2.3)–(2.4). The following result is proven in [1, 2]. In the rest of the paper, the notation $X^\dagger$ denotes the pseudo-inverse of X.

THEOREM 2.1 ([1]). *The problem (2.3)–(2.4) has a unique solution $\hat{x}$ that is given by (compare with (1.2))*

$$\hat{x} = \left[\widehat{Q} + A^T\widehat{W}A\right]^{-1}\left[A^T\widehat{W}b + \hat{\lambda}E_a^TE_b\right] \tag{2.6}$$

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

$$\widehat{Q} \triangleq Q + \hat{\lambda}E_a^TE_a \tag{2.7}$$
$$\widehat{W} \triangleq W + WH(\hat{\lambda}I - H^TWH)^\dagger H^TW \tag{2.8}$$

and the nonnegative scalar parameter $\hat{\lambda}$ is determined from the optimization

$$\hat{\lambda} = \arg\min_{\lambda \geq \|H^TWH\|} G(\lambda) \tag{2.9}$$

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

$$G(\lambda) \triangleq x^T(\lambda)Qx(\lambda) + \lambda\|E_ax(\lambda) - E_b\|^2 + [Ax(\lambda) - b]^TW(\lambda)[Ax(\lambda) - b] \tag{2.10}$$

Here

$$W(\lambda) \triangleq W + WH(\lambda I - H^TWH)^\dagger H^TW \tag{2.11}$$
$$Q(\lambda) \triangleq Q + \lambda E_a^TE_a \tag{2.12}$$

and

$$x(\lambda) \triangleq \left[Q(\lambda) + A^TW(\lambda)A\right]^{-1}\left[A^TW(\lambda)b + \lambda E_a^TE_b\right] \tag{2.13}$$

We thus see that the solution of (2.3)–(2.4) requires that we first determine an optimal nonnegative scalar parameter, $\hat{\lambda}$, which corresponds to the minimizing argument of the function $G(\lambda)$ over the semi-open interval $[\|H^TWH\|, \infty)$. For convenience of notation, we shall denote the lower bound on λ by λ_l, i.e.,

$$\lambda_l \triangleq \|H^TWH\|. \tag{2.14}$$

The parameter $\hat{\lambda}$ is then used to modify the regularization matrix Q and the weighting matrix W according to (2.7) and (2.8).

3. TWO APPLICATIONS. Before proceeding to studying the minimization of $G(\lambda)$, we illustrate the application of the robust least-squares problem (2.3)–(2.4) to two examples. The first example is in the context of state-regulation.

4

3.1. State Regulation. Consider a linear state-space model $x_{i+1} = F_i x_i + G_i u_i$, with initial state x_0 and control sequence $\{u_i, i \geq 0\}$. The classical linear quadratic regulator (LQR) problem is concerned with the determination of a control sequence, $\{\hat{u}_i\}$, that regulates the state vector to zero while keeping the control cost low. This is achieved by minimizing the quadratic cost function

$$\min_{\{u_0, u_1, \ldots, u_N\}} \left(x_{N+1}^T P_{N+1} x_{N+1} + \sum_{j=0}^{N} \left[u_j^T Q_j u_j + x_j^T R_j x_j \right] \right)$$

say over a horizon of duration $N + 1$, with $Q_j > 0$, $R_j \geq 0$, and $P_{N+1} \geq 0$.

Using a dynamic programming argument [13], it is well-known that $\hat{u}_N$ can be determined by solving

$$\min_{u_N} \left(x_{N+1}^T P_{N+1} x_{N+1} + u_N^T Q_N u_N \right) \tag{3.1}$$

Substituting x_{N+1} by its state-equation $x_{N+1} = F_N x_N + G_N u_N$ leads to a quadratic cost function of the form (1.1) and, subsequently, to the state-feedback law:

$$\begin{cases} \hat{u}_N = -K_N x_N \\ K_N = (Q_N + G_N^T P_{N+1} G_N)^{-1} G_N^T P_{N+1} F_N \\ P_N = R_N + K_N^T Q_N K_N + (F_N - G_N K_N)^T P_{N+1} (F_N - G_N K_N) \end{cases} \tag{3.2}$$

The process can be repeated by determining $\hat{u}_{N-1}$ via

$$\min_{u_{N-1}} \left(x_N^T P_N x_N + u_{N-1}^T Q_{N-1} u_{N-1} \right)$$

and so forth. Now, assume that the state-space model includes parametric uncertainties, say

$$x_{i+1} = (F_i + \delta F_i) x_i + (G_i + \delta G_i) u_i \tag{3.3}$$

with known x_0, and where $\{\delta F_i, \delta G_i\}$ denote the uncertain parameters (assumed to satisfy a constraint similar to (2.4)). We can then consider replacing the optimization problem (3.1) by

$$\min_{u_N} \max_{\delta F_N, \delta G_N} \left[x_{N+1}^T P_{N+1} x_{N+1} + u_N^T Q_N u_N \right]$$

If we substitute x_{N+1} by its state-equation $x_{N+1} = (F_N + \delta F_N) x_N + (G_N + \delta G_N) u_N$, the above min-max problem becomes a special case of the robust cost function (2.1) — see [1] for more details.

3.2. State Estimation. The second example is in the context of state estimation. Thus consider a model of the form

$$x_{i+1} = F_i x_i + G_i u_i, \quad i \geq 0 \tag{3.4}$$

$$y_i = H_i x_i + v_i \tag{3.5}$$

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

$$E \left(\begin{bmatrix} x_0 \\ u_i \\ v_i \end{bmatrix} \begin{bmatrix} x_0 \\ u_j \\ v_j \end{bmatrix}^T \right) = \begin{bmatrix} \Pi_0 & 0 & 0 \\ 0 & Q_i \delta_{ij} & 0 \\ 0 & 0 & R_i \delta_{ij} \end{bmatrix} \tag{3.6}$$

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 [14] 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 [15].

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

$$\min_{x_i, u_i} \quad \|x_i - \hat{x}_{i|i}\|^2_{P_{i|i}^{-1}} + \|u_i\|^2_{Q_i^{-1}} + \|y_{i+1} - H_{i+1}x_{i+1}\|^2_{R_{i+1}^{-1}} \tag{3.7}$$

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 form (1.1) and its solution leads to the Kalman filter recursions.

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

$$x_{i+1} = (F_i + \delta F_i)x_i + (G_i + \delta G_i)u_i, \quad i \geq 0 \tag{3.8}$$
$$y_i = H_i x_i + v_i \tag{3.9}$$

where the uncertainties $\{\delta F_i, \delta G_i\}$ are assumed to satisfy a constraint similar to (2.4). We can then consider replacing (3.7) by

$$\min_{\{x_i, u_i\}} \max_{\delta F_i, \delta G_i} \left[\|x_i - \hat{x}_{i|i}\|^2_{P_{i|i}^{-1}} + \|u_i\|^2_{Q_i^{-1}} + \|y_{i+1} - H_{i+1}x_{i+1}\|^2_{R_{i+1}^{-1}} \right] \tag{3.10}$$

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) — see [7] for more details.

4. THE MINIMIZATION OF $G(\lambda)$. Returning to the statement of Theorem 2.1, the contribution of this letter is to prove that the function $G(\lambda)$ does not have local, non-global minima over the interval $[\lambda_l, \infty)$; and that therefore the desired $\hat{\lambda}$ corresponds to a global minimum. Actually, the argument further shows that $G(\lambda)$ has a unique global minimum (except in the unlikely trivial case of $G(\lambda) \equiv$ constant for all $\lambda \geq \lambda_l$). In this way, the robustified least-squares problem (2.3)–(2.4) is well defined and the determination of the optimal parameter $\hat{\lambda}$ can be sought without concerns about local minima.

To begin with, we note that for any value of λ in the interval $[\lambda_l, \infty)$, the matrix $W(\lambda)$ in (2.11) is positive semi-definite,

$$W(\lambda) \geq 0 \quad \text{for } \lambda \in [\lambda_l, \infty)$$

so that the function $G(\lambda)$ is nonnegative for all such λ, i.e., $G(\lambda) \geq 0$ for $\lambda \geq \lambda_l$. Notice however that $G(\lambda)$ may become negative for $\lambda < \lambda_l$. Figures 4.1 and 4.2 illustrate typical behaviors of $G(\lambda)$. The top plot in Figure 4.1 shows one particular $G(\lambda)$ that was generated with the data

$$A = \begin{bmatrix} 1 & -1 \\ 0 & 1 \\ -2 & 0 \\ 0 & 2 \end{bmatrix}, \, b = \begin{bmatrix} -1 \\ 0 \\ 1 \\ -2 \end{bmatrix}, \, Q = \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \, W = I$$

$$E_a = \begin{bmatrix} 1 & 0 \\ 0 & -2 \\ 1 & 1 \\ 0 & 0 \end{bmatrix}, \quad E_b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad H = \begin{bmatrix} -1 & 0 \\ 1 & 2 \\ 0 & 0 \\ 1 & -1 \end{bmatrix}$$

for which the lower bound on λ can be seen to be

$$\lambda_l = \|H^T W H\| = 5.4142.$$

The bottom plot in the same figure zooms on the behavior of $G(\lambda)$ over the interval $\lambda \geq \lambda_l$. Figure 4.2 shows similar plots generated with random data $\{A, b, Q, W, E_a, E_b, H\}$ over the respective intervals $\lambda \geq \lambda_l$. In all cases, $G(\lambda)$ is seen to be nonnegative (but not necessarily convex). Moreover, in each case $G(\lambda)$ is seen to have a unique global minimum in the interval $[\|H^T W H\|, \infty)$.
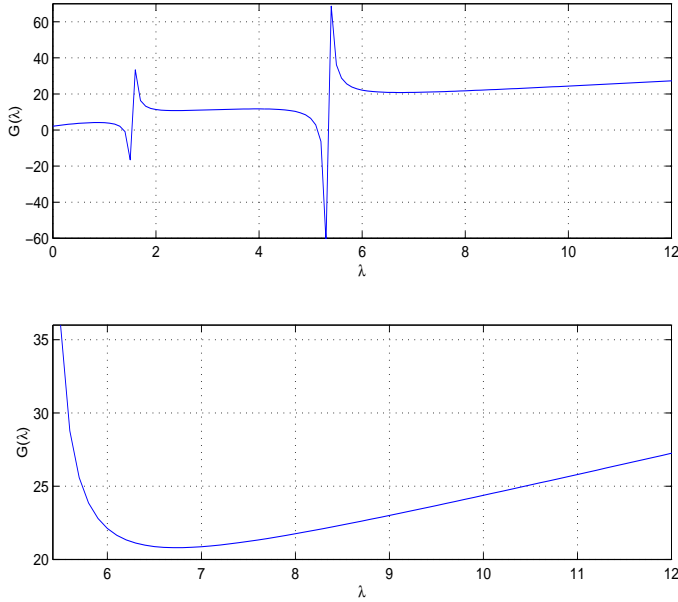

FIG. 4.1. Plots of $G(\lambda)$ over the intervals $\lambda \geq 0$ (top) and $\lambda \geq \lambda_l = 5.4142$ (bottom). Note that $G(\lambda)$ is nonnegative and has a unique minimum over the latter interval.

From now on we assume that $W > 0$, which is often satisfied in practice. We shall then study the behavior of $G(\lambda)$ in two cases. First, we focus on the open interval $(\|H^T W H\|, \infty)$, and later we discuss the boundary point $\lambda_l = \|H^T W H\|$.

For $\lambda > \|H^T W H\|$, it is easy to see that $(\lambda I - H^T W H)$ will be positive-definite (and, hence, also invertible), so that the pseudo-inverse operation in (2.11) can be replaced by normal matrix inversion,

$$W(\lambda) = W + WH(\lambda I - H^T W H)^{-1} H^T W. \tag{4.1}$$

Using the matrix inversion lemma [18], also known as the Sherman-Morrison-Woodbury formula [19], we arrive at the more compact representation

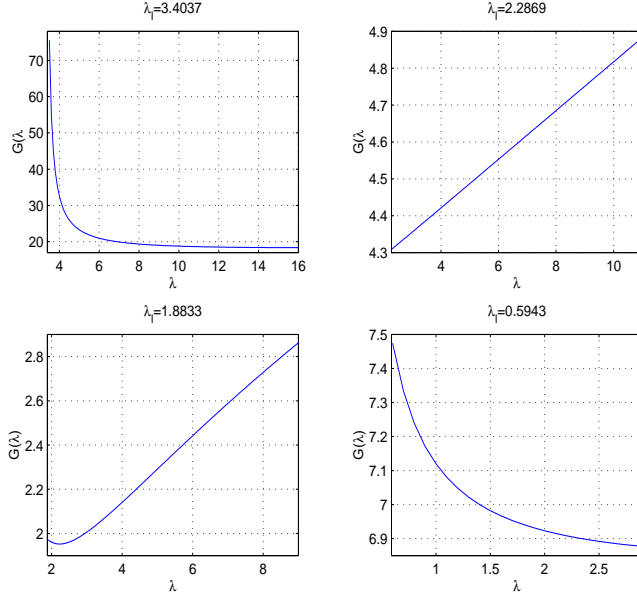$$W^{-1}(\lambda) = W^{-1} - \lambda^{-1} H H^T. \tag{4.2}$$

7

FIG. 4.2. Plots of $G(\lambda)$ over the intervals $\lambda \geq \lambda_l$ in four random simulations. The values of λ_l are indicated at the top of each plot. The plots show typical behaviors of $G(\lambda)$. In particular, it is seen that $G(\lambda)$ is nonnegative and has a unique minimum in all cases.

We shall henceforth use this simpler expression.

Moreover, it can also be verified that an equivalent representation of (2.10) is

$$G(\lambda) = \lambda E_b^T E_b + b^T W(\lambda)b - B^T(\lambda)E^{-1}(\lambda)B(\lambda) \tag{4.3}$$

where the functions $\{W(\lambda), B(\lambda), E(\lambda)\}$ are defined by

$$W(\lambda) = \left(W^{-1} - \lambda^{-1}HH^T\right)^{-1} \tag{4.4}$$

$$B(\lambda) = A^T W(\lambda)b + \lambda E_a^T E_b \tag{4.5}$$

$$E(\lambda) = Q + \lambda E_a^T E_a + A^T W(\lambda)A \tag{4.6}$$

In order to study the nature of the stationary points of $G(\lambda)$ over the interval (λ_l, ∞), one would in principle evaluate the first- and second-order derivatives of $G(\lambda)$ with respect to λ and check their signatures. However, performing this task directly is not easy in the problem at hand. For this reason, we choose to proceed by instead relating the first-and second-order derivatives and by using Schur complementation arguments to establish a certain positivity result. The details are provided below.

First, the following facts about differentiation are useful. Consider a continuous and differentiable matrix-valued function, $K(\lambda)$, of a scalar parameter λ. If $K(\lambda)$ is invertible, say,

$$K(\lambda)K^{-1}(\lambda) = I$$

then by differentiating both sides of this equality with respect to λ we find that

$$\frac{dK^{-1}(\lambda)}{d\lambda} = -K^{-1}(\lambda) \cdot \frac{dK(\lambda)}{d\lambda} \cdot K^{-1}(\lambda).$$

8

Likewise, differentiating one more time we arrive at the expression

$$\frac{d^2 K^{-1}(\lambda)}{(d\lambda)^2} = 2K^{-1}(\lambda) \cdot \frac{dK(\lambda)}{d\lambda} \cdot K^{-1}(\lambda) \cdot \frac{dK(\lambda)}{d\lambda} \cdot K^{-1}(\lambda) - K^{-1}(\lambda) \cdot \frac{d^2 K(\lambda)}{(d\lambda)^2} \cdot K^{-1}(\lambda).$$

Applying these results to the function $G(\lambda)$ we arrive at the following result.

LEMMA 1. *Consider the function $G(\lambda)$ defined by (4.3)–(4.6) with $W > 0$. Let*

$$S(\lambda) = E_b - E_a E^{-1}(\lambda) B(\lambda), \qquad L(\lambda) = H^T W(\lambda)(b - AE^{-1}(\lambda)B(\lambda))$$

Then it holds that

$$\frac{dG(\lambda)}{d\lambda} = S^T(\lambda)S(\lambda) - \frac{1}{\lambda^2} L^T(\lambda)L(\lambda)$$

and $\frac{1}{2}\frac{d^2 G(\lambda)}{d\lambda^2}$ is the Schur complement with respect to the $(2,2)$ block entry of a matrix $\mathcal{X}$ of the form

$$\mathcal{X} = F(\lambda) + \frac{1}{\lambda} \begin{bmatrix} -\frac{dG(\lambda)}{d\lambda} & 0 \\ 0 & 0 \end{bmatrix}$$

where $F(\lambda)$ is a positive semi-definite matrix.

Proof: Differentiating $G(\lambda)$ with respect to λ, we get

$$\frac{dG(\lambda)}{d\lambda} = E_b^T E_b + b^T \frac{dW(\lambda)}{d\lambda} b - 2\left(\frac{dB(\lambda)}{d\lambda}\right)^T E^{-1}(\lambda)B(\lambda) - B^T(\lambda)\frac{dE^{-1}(\lambda)}{d\lambda} B(\lambda)$$

which can be rearranged into the form given in the statement of the lemma above. Differentiating again with respect to λ, we get after some manipulations:

$$\frac{d^2 G(\lambda)}{(d\lambda)^2} = 2(X_1 - X_2 X_3^{-1} X_2^T)$$

where we defined

$$Z = b - AE^{-1}(\lambda)B(\lambda)$$

$$X_1 = Z^T \left(W(\lambda)\frac{HH^T}{\lambda^2}W(\lambda)\frac{HH^T}{\lambda^2}W(\lambda) + W(\lambda)\frac{HH^T}{\lambda^3}W(\lambda)\right)Z$$

$$X_2 = \left[-A^T \frac{W(\lambda)HH^T W(\lambda)}{\lambda^2}\left(b - AE^{-1}(\lambda)B(\lambda)\right) + E_a^T\left(E_b - E_a E^{-1}(\lambda)B(\lambda)\right)\right]^T$$

$$X_3 = E(\lambda) = Q + \lambda E_a^T E_a + A^T W(\lambda)A$$

We can therefore regard $\frac{1}{2}\frac{d^2 G(\lambda)}{d\lambda^2}$ as the Schur complement of the block matrix

$$\mathcal{X} \triangleq \begin{bmatrix} X_1 & X_2 \\ X_2^T & X_3 \end{bmatrix}.$$

Now the matrix $\mathcal{X}$ can be expanded as a sum of several terms, namely

$$
\mathcal{X} = \begin{bmatrix} 0 & 0 \\ 0 & Q \end{bmatrix}
$$

$$
+ \begin{bmatrix} \frac{1}{\sqrt{\lambda}}\left(E_b - E_a E^{-1}(\lambda)B(\lambda) \right)^T \\ \sqrt{\lambda}E_a^T \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{\lambda}}\left(E_b - E_a E^{-1}(\lambda)B(\lambda) \right) & \sqrt{\lambda}E_a \end{bmatrix}
$$

$$
+ \begin{bmatrix} -\dfrac{\left(b - AE^{-1}(\lambda)B(\lambda) \right)^T W(\lambda)HH^T}{\lambda^2} \\ A^T \end{bmatrix} W(\lambda) \begin{bmatrix} -\dfrac{HH^T W(\lambda)\left(b - AE^{-1}(\lambda)B(\lambda) \right)}{\lambda^2} & A \end{bmatrix}
$$

$$
+ \frac{1}{\lambda} \begin{bmatrix} \dfrac{\left(b - AE^{-1}(\lambda)B(\lambda) \right)^T W(\lambda)H}{\lambda} \\ 0 \end{bmatrix} \begin{bmatrix} \dfrac{H^T W(\lambda)\left(b - AE^{-1}(\lambda)B(\lambda) \right)}{\lambda} & 0 \end{bmatrix}
$$

$$
- \frac{1}{\lambda} \begin{bmatrix} \left(E_b - E_a E^{-1}(\lambda)B(\lambda) \right)^T \\ 0 \end{bmatrix} \begin{bmatrix} E_b - E_a E^{-1}(\lambda)B(\lambda) & 0 \end{bmatrix}
$$

and it can be seen that the sum of the first three terms results in a positive semi-definite matrix (which we denote by $F(\lambda)$), while the sum of the other two terms gives a matrix that is closely related to $dG(\lambda)/d\lambda$. Actually, using the expression for $dG(\lambda)/d\lambda$ given in Lemma 1 we find that

$$
\mathcal{X} = F(\lambda) + \frac{1}{\lambda} \begin{bmatrix} -\frac{dG(\lambda)}{d\lambda} & 0 \\ 0 & 0 \end{bmatrix}.
$$

$\blacksquare$

It follows from Lemma 1 that whenever

$$
dG(\lambda)/d\lambda \le 0
$$

we get $\mathcal{X} \ge 0$ and, consequently, its Schur complement is also positive semi-definite, i.e.,

$$
d^2 G(\lambda)/(d\lambda)^2 \ge 0.
$$

In other words, it holds that for all $\lambda \in (\lambda_l, \infty)$

$$
dG(\lambda)/d\lambda \le 0 \implies d^2 G(\lambda)/(d\lambda)^2 \ge 0.
$$

This property guarantees that $G(\lambda)$ does not have any (local or global) disjoint maxima in the interval (λ_l, ∞), since a maximum would require the simultaneous conditions $dG(\lambda)/d\lambda = 0$ and $d^2 G(\lambda)/d\lambda^2 < 0$. It then follows that $G(\lambda)$ cannot have two or more disjoint (local or global) minima since this would imply the existence of (local or global) maxima.

The above argument indicates that $G(\lambda)$ can only have a unique global minimum or a continuum of minima or maxima. The latter case occurs in either of the following two situations:

1. $G(\lambda)$ is flat (i.e., equal to a constant) over the entire interval (λ_l, ∞). This is a trivial case that is not of interest. It occurs, for example, when the quantities $E_b^T E_b, H H^T, E_a^T E_a, E_a^T E_b$ are all zero, which is not relevant in the case of uncertain models.

2. $G(\lambda)$ is constant over a subinterval of (λ_l, ∞). This case is excluded since it would imply that some high-order derivative of $G(\lambda)$ is discontinuous. As defined by (2.10), $G(\lambda)$ is an analytic function of λ and, for any n, its nth-order derivative is also analytic in λ.

We thus conclude that $G(\lambda)$ has a unique global minimum or infimum over the open interval (λ_l, ∞). Now by the continuity of $G(\lambda)$, and in view of the above arguments, we conclude that $G(\lambda)$ has a unique global minimum over the closed interval $[\lambda_l, \infty)$.

5. CONCLUDING REMARKS. Theorem 2.1 shows that the robust solution of the constrained game problem (2.3)–(2.4) involves a scalar-valued optimization step that is defined by (2.9). This step requires the determination of a scalar parameter $\hat{\lambda}$ by minimizing a function $G(\lambda)$ over an interval $[\lambda_l, \infty)$. In this letter, we proved that $G(\lambda)$ cannot have local, non-global minima over the interval $[\lambda_l, \infty)$. More specifically, except in a trivial degenerate case of $G(\lambda) \equiv$ constant, the arguments show that $G(\lambda)$ has a unique global minimum over $[\lambda_l, \infty)$. The result is reassuring in that it demonstrates that the robust design procedure of Theorem 2.1 is well defined and that its optimal global solution can be determined without concerns about local minima for (2.9). Applications of the procedure described herein in the context of state-space regulation and state-space estimation can be found in [1, 7, 17].

REFERENCES

[1] 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, vol. 245, pp. 159–173, Springer Verlag, London, 1999.

[2] A. H. Sayed, V. H. Nascimento, and F. A. M. Cipparrone, A regularized robust design criterion for uncertain data, *SIAM J. Matrix Anal. Appl.*, vol. 23, no. 4, 1120–1142, 2002.

[3] 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, Jul. 1997.

[4] L. E. Ghaoui and H. Hebret, Robust solutions to least-squares problems with uncertain data, *SIAM J. Matrix Anal. Appl.*, vol. 18, no.4, pp. 1035–1064, 1997.

[5] 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.

[6] 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.

[7] A. H. Sayed, A framework for state-space estimation with uncertain models. *IEEE Transactions on Automatic Control*, vol. 46, no. 7, pp. 998–1013, July 2001.

[8] R. H. Bartles, G. H. Golub, and M. A. Saunders, Numerical techniques in mathematical programming, in *Nonlinear Programming*, J. B. Rosen, O. L. Mangasarian, and K. Ritter, editors, pp. 123–176, Academic Press, 1970.

[9] J. R. Bunch, C. P. Nielsen, and D. C. Sorensen, Rank-one modification of the symmetric eigenproblem, *Numer. Math.*, vol. 31, pp. 31–48, 1978.

[10] A. Melman, Analysis of third-order methods for secular equations, *Math. Comp.,* 67(221), pp. 271–286, 1998.

[11] C. H. Reinsch, Smoothing by spline functions II, *Numer. Math.*, vol. 16, pp. 451–454, 1971.

[12] S. Chandrasekaran, G. Golub, M. Gu, and A. H. Sayed, "An efficient algorithm for a bounded errors-in-variables model," *SIAM J. Mat. Anal. Appl.*, vol. 20, no. 4, pp. 839–859, Oct. 1999.

[13] K. J. Astrom, *Introduction to Stochastic Control Theory*, Academic Press, NY, 1970.

[14] T. Kailath, A. H. Sayed, and B. Hassibi, *Linear Estimation*, Prentice Hall, NJ, 2000.

[15] A. E. Bryson and Y. C. Ho, *Applied Optimal Control*, Blaisdell, Waltham, MA, 1969.

[16] 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.

[17] A. H. Sayed and A. Subramanian, "State-space estimation with uncertain models,"in *Total Least Squares and Errors-in-Variables Modeling, III: Analysis, Algorithms and Applications*, S. Van Huffel and P. Lemmerling, editors, Kluwer, pp. 191–202, 2002.

[18] T. Kailath, *Linear Systems*, Prentice-Hall, NJ, 1980.

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