ROBUST SOLUTIONS TO LEAST-SQUARES PROBLEMS WITH UNCERTAIN DATA [*]

LAURENT EL GHAOUI[†] AND HERVÉ LEBRET[†]

Abstract. We consider least-squares problems where the coefficient matrices A, b are unknown but bounded. We minimize the worst-case residual error using (convex) second-order cone programming, yielding an algorithm with complexity similar to one singular value decomposition of A. The method can be interpreted as a Tikhonov regularization procedure, with the advantage that it provides an exact bound on the robustness of solution and a rigorous way to compute the regularization parameter. When the perturbation has a known (e.g., Toeplitz) structure, the same problem can be solved in polynomial-time using semidefinite programming (SDP). We also consider the case when A, b are rational functions of an unknown-but-bounded perturbation vector. We show how to minimize (via SDP) upper bounds on the optimal worst-case residual. We provide numerical examples, including one from robust identification and one from robust interpolation.

Key words. least-squares problems, uncertainty, robustness, second-order cone programming, semidefinite programming, ill-conditioned problem, regularization, robust identification, robust interpolation

AMS subject classifications. 15A06, 65F10, 65F35, 65K10, 65Y20

PII. S0895479896298130

Notation. For a matrix X, $\|X\|$ denotes the largest singular value and $\|X\|_F$ the Frobenius norm. If x is a vector, $\max_i |x_i|$ is denoted by $\|x\|_\infty$. For a matrix A, $A^\dagger$ denotes the Moore–Penrose pseudoinverse of A. For a square matrix S, $S \geq 0$ (resp., $S > 0$) means S is symmetric and positive semidefinite (resp., definite). For $S \geq 0$, $S^{1/2}$ denotes the symmetric square root of S. For $S > 0$, and given vector x, we define $\|x\|_S = \|S^{-1/2}x\|$. The notation I_p denotes the $p \times p$ identity matrix; sometimes the subscript is omitted when it can be inferred from context. For given matrices X, Y, the notation $X \oplus Y$ refers to the block-diagonal matrix with X, Y as diagonal blocks.

1. Introduction. Consider the problem of finding a solution x to an overdetermined set of equations $Ax \simeq b$, where the data matrices $A \in \mathbf{R}^{n \times m}$, $b \in \mathbf{R}^n$ are given. The least squares (LS) fit minimizes the residual $\|\Delta b\|$ subject to $Ax = b + \Delta b$, resulting in a consistent linear model of the form $(A, b + \Delta b)$ that is closest to the original one (in the Euclidean norm sense). The total least squares (TLS) solution described by Golub and Van Loan [17] finds the smallest error $\|[\Delta A \ \Delta b]\|_F$ subject to the consistency equation $(A + \Delta A)x = b + \Delta b$. The resulting closest consistent linear model $(A + \Delta A, b + \Delta b)$ is even more accurate than the LS one, since modifications of A are allowed.

Accuracy is the primary aim of LS and TLS, so it is not surprising that both solutions may exhibit very sensitive behavior to perturbations in the data matrices (A, b). Detailed sensitivity analyses for the LS and TLS problems may be found in [12, 18, 2, 44, 22, 14]. Many regularization methods have been proposed to decrease sensitivity and make LS and TLS applicable. Most regularization schemes for LS, including Tikhonov regularization [43], amount to solve a weighted LS problem for

---

an augmented system. As pointed out in [18], the choice of weights (or regularization parameter) is usually not obvious and application dependent. Several criteria for optimizing the regularization parameter(s) have been proposed (see, e.g., [23, 11, 15]). These criteria are chosen according to some additional a priori information, of deterministic or stochastic nature. The extensive surveys [31, 8, 21] discuss these problems and some applications.

In contrast with the extensive work on sensitivity and regularization, relatively little has been done on the subject of *deterministic robustness* of LS problems in which the perturbations are deterministic and unknown but bounded (not necessarily small). Some work has been done on a qualitative analysis of the problem, where entries of (A, b) are unspecified except for their sign [26, 39]. In many papers mentioning least squares and robustness, the latter notion is understood in some stochastic sense; see, e.g., [20, 47, 37]. A notable exception concerns the field of identification, where the subject has been explored using a framework used in control system analysis [40, 9], or using regularization ideas combined with additional a priori information [34, 42].

In this paper, we assume that the data matrices are subject to (not necessarily small) deterministic perturbations. First, we assume that the given model is not a single pair (A, b) but a family of matrices $(A + \Delta A, b + \Delta b)$, where $\Delta = [\Delta A \ \Delta b]$ is an unknown-but-bounded matrix; precisely, $\|\Delta\| \leq \rho$, where $\rho \geq 0$ is given. For x fixed, we define the worst-case residual as

$$(1) \qquad r(A, b, \rho, x) \triangleq \max_{\|\Delta A \ \Delta b\|_F \leq \rho} \|(A + \Delta A)x - (b + \Delta b)\|.$$

We say that x is a robust least squares (RLS) solution if x minimizes the worst-case residual $r(A, b, \rho, x)$. The RLS solution trades accuracy for robustness at the expense of introducing bias. In our paper, we assume that the perturbation bound ρ is known, but in section 3.5 we also show that TLS can be used as a preliminary step to obtain a value of ρ that is consistent with data matrices A, b.

In many applications, the perturbation matrices ΔA, Δb have a known structure. For instance, ΔA might have a Toeplitz structure inherited from A. In this case, the worst-case residual (1) might be a very conservative estimate. We are led to consider the following structured RLS (SRLS) problem. Given $A_0, \ldots, A_p \in \mathbf{R}^{n \times m}$, $b_0, \ldots, b_p \in \mathbf{R}^n$, we define, for every $\delta \in \mathbf{R}^p$,

$$(2) \qquad \mathbf{A}(\delta) \triangleq A_0 + \sum_{i=1}^p \delta_i A_i, \quad \mathbf{b}(\delta) \triangleq b_0 + \sum_{i=1}^p \delta_i b_i.$$

For $\rho \geq 0$ and $x \in \mathbf{R}^m$, we define the structured worst-case residual as

$$(3) \qquad r_S(\mathbf{A}, \mathbf{b}, \rho, x) \triangleq \max_{\|\delta\| \leq \rho} \|\mathbf{A}(\delta)x - \mathbf{b}(\delta)\|.$$

We say that x is an SRLS solution if x minimizes the worst-case residual $r_S(\mathbf{A}, \mathbf{b}, \rho, x)$.

Our main contribution is to show that we can compute the *exact value* of the optimal worst-case residuals using convex, second-order cone programming (SOCP) or semidefinite programming (SDP). The consequence is that the RLS and SRLS problems can be solved in polynomial time and with great practical efficiency using, e.g., recent interior-point methods [33, 46]. Our exact results are to be contrasted with those of Doyle et al. [9], who also use SDP to compute *upper bounds* on the worst-case residual for identification problems. In the preliminary draft [5] sent to us shortly

after submission of this paper, the authors provide a solution to an (unstructured) RLS problem, which is similar to that given in section 3.2.

Another contribution is to show that the RLS solution is continuous in the data matrices A, b. RLS can thus be interpreted as a (Tikhonov) regularization technique for ill-conditioned LS problems: the additional a priori information is ρ (the perturbation level), and the regularization parameter is optimal for robustness. Similar regularity results hold for the SRLS problem.

We also consider a generalization of the SRLS problem, referred to as the linear-fractional SRLS problem in what follows, in which the matrix functions $\mathbf{A}(\delta)$, $\mathbf{b}(\delta)$ in (2) depend rationally on the parameter vector δ. (We describe a robust interpolation problem that falls in this class in section 7.6.) Using the framework of [9], we show that the problem is NP-complete in this case, but we may compute and optimize upper bounds on the worst-case residual using SDP. In parallel with RLS, we interpret our solution as one of a weighted LS problem for an augmented system, the weights being computed via SDP.

The paper's outline is as follows. The next section is devoted to some technical lemmas. Section 3 is devoted to the RLS problem. In section 4, we consider the SRLS problem. Section 5 studies the linear-fractional SRLS problem. Regularity results are given in section 6. Section 7 shows numerical examples.

2. Preliminary results.

2.1. Semidefinite and second-order cone programs.
We briefly recall some important results on semidefinite programs (SDPs) and second-order cone programs (SOCPs). These results can be found, e.g., in [4, 33, 46].

A linear matrix inequality is a constraint on a vector $x \in \mathbf{R}^m$ of the form

$$(4) \qquad \mathcal{F}(x) = \mathcal{F}_0 + \sum_{i=1}^m x_i \mathcal{F}_i \geq 0,$$

where the symmetric matrices $\mathcal{F}_i = \mathcal{F}_i^T \in \mathbf{R}^{N \times N}$, $i = 0, \ldots, m$, are given. The minimization problem

$$(5) \qquad \text{minimize } c^T x \text{ subject to } \mathcal{F}(x) \geq 0,$$

where $c \in \mathbf{R}^m$, is called an SDP. SDPs are convex optimization problems and can be solved in polynomial time with, e.g., primal-dual interior-point methods [33, 45].

The problem dual to problem (5) is

$$(6) \qquad \begin{array}{ll} \text{maximize} & -\mathbf{Tr}\mathcal{F}_0\mathcal{Z} \\ \text{subject to} & \mathcal{Z} \geq 0, \ \ \mathbf{Tr}\mathcal{F}_i\mathcal{Z} = c_i, \ i = 1, \ldots, m, \end{array}$$

where $\mathcal{Z}$ is a symmetric $N \times N$ matrix and c_i is the ith coordinate of vector c. When both problems are strictly feasible (that is, when there exists $x, \mathcal{Z}$ which satisfy the constraints strictly), the existence of optimal points is guaranteed [33, Thm. 4.2.1], and both problems have equal optimal objectives. In this case, the optimal primal-dual pairs $(x, \mathcal{Z})$ are those pairs $(x, \mathcal{Z})$ such that x is feasible for the primal problem, $\mathcal{Z}$ is feasible for the dual one, and $\mathcal{F}(x)\mathcal{Z} = 0$.

An SOCP problem is one of the form

$$(7) \qquad \begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \|C_i x + d_i\| \leq e_i^T x + f_i, \ \ i = 1, \ldots, L, \end{array}$$

where $C_i \in \mathbf{R}^{n_i \times m}$, $d_i \in \mathbf{R}^{n_i}$, $e_i \in \mathbf{R}^m$, $f_i \in \mathbf{R}$, $i = 1, \dots, L$. The dual problem of problem (7) is

$$(8) \qquad \begin{aligned} &\text{maximize} & &-\sum_{i=1}^{L} \left(d_i^T z_i + f_i s_i \right) \\ &\text{subject to} & &\sum_{i=1}^{L} \left(C_i^T z_i + e_i s_i \right) = c, \ \ \|z_i\| \le s_i, \ i = 1, \dots, L, \end{aligned}$$

where $z_i \in \mathbf{R}^{n_i}$, $s_i \in \mathbf{R}$, $i = 1, \dots, L$ are the dual variables. Optimality conditions similar to those for SDPs can be obtained for SOCPs. SOCPs can be expressed as SDPs; therefore, they can be solved in polynomial time using interior-point methods for SDPs. However, the SDP formulation is not the most efficient numerically, as special interior-point methods can be devised for SOCPs [33, 28, 1].

Precise complexity results on interior-point methods for SOCPs and SDPs are given by Nesterov and Nemirovsky [33, pp. 224, 236]. In practice, it is observed that the number of iterations is almost constant, independent of problem size [46]. For the SOCP, each iteration has complexity $O((n_1 + \cdots + n_L)m^2 + m^3)$; for the SDP, we refer the reader to [33].

2.2. $\mathcal{S}$-procedure. The following lemma can be found, e.g., in [4, p. 24]. It is widely used, e.g., in control theory and in connection with trust region methods in optimization [41].

LEMMA 2.1 ($\mathcal{S}$-procedure). *Let* $F_0, \dots, F_p$ *be quadratic functions of the variable* $\zeta \in \mathbf{R}^m$:

$$F_i(\zeta) \triangleq \zeta^T T_i \zeta + 2u_i^T \zeta + v_i, \ i = 0, \dots, p,$$

where $T_i = T_i^T$. *The following condition on* $F_0, \dots, F_p$:

$$F_0(\zeta) \ge 0 \ \text{for all } \zeta \text{ such that } F_i(\zeta) \ge 0, \ i = 1, \dots, p,$$

holds if

$$\text{there exist } \tau_1 \ge 0, \dots, \tau_p \ge 0 \ \text{ such that } \begin{bmatrix} T_0 & u_0 \\ u_0^T & v_0 \end{bmatrix} - \sum_{i=1}^{p} \tau_i \begin{bmatrix} T_i & u_i \\ u_i^T & v_i \end{bmatrix} \ge 0.$$

When $p = 1$, *the converse holds, provided that there is some* ζ_0 *such that* $F_1(\zeta_0) > 0$.

The next lemma is a corollary of the above result in the case $p = 1$.

LEMMA 2.2. *Let* $T_1 = T_1^T$, T_2, T_3, T_4 *be real matrices of appropriate size. We have* $\det(I - T_4 \Delta) \ne 0$ *and*

$$(9) \qquad T(\Delta) = T_1 + T_2 \Delta (I - T_4 \Delta)^{-1} T_3 + T_3^T (I - T_4 \Delta)^{-T} \Delta^T T_2^T \ge 0$$

for every Δ, $\|\Delta\| \le 1$, *if and only if* $\|T_4\| < 1$ *and there exists a scalar* $\tau \ge 0$ *such that*

$$(10) \qquad \begin{bmatrix} T_1 - \tau T_2 T_2^T & T_3^T - \tau T_2 T_4^T \\ T_3 - \tau T_4 T_2^T & \tau (I - T_4 T_4^T) \end{bmatrix} \ge 0.$$

Proof. If T_2 or T_3 equal zero, the result is obvious. Now assume $T_2, T_3 \neq 0$. Then, (10) implies $\tau > 0$, which in turn implies $\|T_4\| < 1$. Thus, for a given τ, (10) holds if and only if $\|T_4\| < 1$, and for every (u, p) we have

$$u^T(T_1 u + 2T_3^T p) - \tau(q^T q - p^T p) \geq 0,$$

where $q = T_2^T u + T_4^T p$. Since $T_2 \neq 0$, the constraint $q^T q \geq p^T p$ is qualified, that is, satisfied strictly for some (u_0, p_0) (choose $p_0 = 0$ and u_0 such that $T_2^T u_0 \neq 0$). Using the $\mathcal{S}$-procedure, we obtain that there exists $\tau \in \mathbf{R}$ such that (10) holds if and only if $\|T_4\| < 1$, and for every (u, p) such that $q^T q \geq p^T p$ we have $u^T(T_1 u + 2T_3^T p) \geq 0$. We end our proof by noting that for every pair (p, q), $p = \Delta^T q$ for some Δ, $\|\Delta\| \leq 1$ if and only if $p^T p \leq q^T q$. ☐

The following lemma is a "structured" version of the above, which can be traced back to [13].

LEMMA 2.3. *Let* $T_1 = T_1^T$, T_2, T_3, T_4 *be real matrices of appropriate size. Let* $\mathcal{D}$ *be a subspace of* $\mathbf{R}^{N \times N}$ *and denote by* $\mathcal{S}$ *(resp.,* $\mathcal{G}$*) the set of symmetric (resp., skew-symmetric) matrices that commute with every element of* $\mathcal{D}$*. We have* $\det(I - T_4\Delta) \neq 0$ *and* (9) *for every* $\Delta \in \mathcal{D}$, $\|\Delta\| \leq 1$, *if there exist* $S \in \mathcal{S}$, $G \in \mathcal{G}$ *such that*

$$\begin{bmatrix} T_1 - T_2 S T_2^T & T_3^T - T_2 S T_4^T + T_2 G \\ T_3 - T_4 S T_2^T - G T_2^T & S - G T_4^T + T_4 G - T_4 S T_4^T \end{bmatrix} > 0, \quad S > 0.$$

If $\mathcal{D} = \mathbf{R}^{\mathcal{N} \times \mathcal{N}}$, *the condition is necessary and sufficient.*

Proof. The proof follows the scheme of that of Lemma 2.2, except that $p^T p \leq q^T q$ is replaced with $p^T S p \leq q^T S q$, $p^T G q = 0$, for given $S \in \mathcal{S}$, $S > 0$, $G \in \mathcal{G}$. Note that for $G = 0$, the above result is a simple application of Lemma 2.2 to the scaled matrices T_1, $T_2 S^{-1/2}$, $S^{1/2} T_3$, $S^{1/2} T_4 S^{-1/2}$. ☐

2.3. Elimination lemma. The last lemma is proven in [4, 24].

LEMMA 2.4 (elimination). *Given real matrices* $W = W^T$, U, V *of appropriate size, there exists a real matrix* X *such that*

$$(11) \qquad\qquad W + UXV^T + VX^T U^T > 0$$

if and only if

$$(12) \qquad\qquad \tilde{U}^T W \tilde{U} > 0 \text{ and } \tilde{V}^T W \tilde{V} > 0,$$

where $\tilde{U}$, $\tilde{V}$ *are orthogonal complements of* U, V*. If* U, V *are full column rank, and* (12) *holds, a solution* X *to the inequality* (11) *is*

$$(13) \qquad\qquad X = \sigma(U^T Q^{-1} U)^{-1} U^T Q^{-1} V,$$

where $Q \triangleq W + \sigma V V^T$*, and* σ *is any scalar such that* $Q > 0$ *(the existence of which is guaranteed by* (12)*).*

3. Unstructured RLS. In this section, we consider the RLS problem, which is to compute

$$(14) \qquad \phi(A, b, \rho) \triangleq \min_x \max_{\|\Delta A\ \Delta b\|_F \leq \rho} \|(A + \Delta A)x - (b + \Delta b)\|.$$

For $\rho = 0$, we recover the standard LS problem. For every $\rho > 0$, $\phi(A, b, \rho) = \rho\phi(A/\rho, b/\rho, 1)$, so we take $\rho = 1$ in what follows, unless otherwise stated. In the remainder of this paper, $\phi(A, b)$ (resp., $r(A, b, x)$) denotes $\phi(A, b, 1)$ (resp., $r(A, b, 1, x)$).

In the preceding definition, the norm used for the perturbation bound is the Frobenius norm. As will be seen, the worst-case residual is the same when the norm used is the largest singular value norm.

3.1. Optimizing the worst-case residual. The following results yield a numerically efficient algorithm for solving the RLS problem in the unstructured case.

THEOREM 3.1. *When $\rho = 1$, the worst-case residual* (1) *is given by*

$$r(A, b, x) = \|Ax - b\| + \sqrt{\|x\|^2 + 1}.$$

The problem of minimizing $r(A, b, x)$ over $x \in \mathbf{R}^m$ has a unique solution x_{RLS}, referred to as the RLS solution. This problem can be formulated as the SOCP

$$(15) \qquad minimize\ \lambda\ subject\ to\ \|Ax - b\| \leq \lambda - \tau,\ \left\| \begin{bmatrix} x \\ 1 \end{bmatrix} \right\| \leq \tau.$$

Proof. Fix $x \in \mathbf{R}^m$. Using the triangle inequality, we have

$$(16) \qquad r(A, b, x) \leq \|Ax - b\| + \sqrt{\|x\|^2 + 1}.$$

Now choose $\Delta = [\Delta A\ \Delta b]$ as

$$[\Delta A\ \Delta b] = \frac{u}{\sqrt{\|x\|^2 + 1}} \begin{bmatrix} x^T & 1 \end{bmatrix},\ \text{where}\ u = \begin{cases} \dfrac{Ax - b}{\|Ax - b\|} & \text{if } Ax \neq b, \\ \text{any unit-norm vector otherwise.} \end{cases}$$

Since Δ is rank one, we have $\|\Delta\|_F = \|\Delta\| = 1$. In addition, we have

$$\|(A + \Delta A)x - (b + \Delta b)\| = \|Ax - b\| + \sqrt{\|x\|^2 + 1},$$

which implies that Δ is a worst-case perturbation (for both the Frobenius and maximum singular value norms) and that equality always holds in (16). Finally, unicity of the minimizer x follows from the strict convexity of the worst-case residual. □

Using an interior-point primal-dual potential reduction method for solving the unstructured RLS problem (15), the number of iterations is almost constant [46]. Furthermore, each iteration takes $O((n + m)m^2)$ operations. A rough summary of this analysis is that the method has the same order of complexity as one singular value decomposition (SVD) of A.

3.2. Analysis of the optimal solution. Using duality results for SOCPs, we have the following theorem.

THEOREM 3.2. *When $\rho = 1$, the (unique) solution x_{RLS} to the RLS problem is given by*

$$(17) \qquad x_{\mathrm{RLS}} = \begin{cases} (\mu I + A^T A)^{-1} A^T b & \text{if } \mu \triangleq (\lambda - \tau)/\tau > 0, \\ A^\dagger b & \text{else,} \end{cases}$$

where (λ, τ) are the (unique) optimal points for problem (15).

Proof. Using the results of section 2.1, we obtain that the problem dual to (15) is

$$maximize\ b^T z - v\ subject\ to\ A^T z + u = 0,\ \|z\| \leq 1,\ \left\| \begin{bmatrix} u \\ v \end{bmatrix} \right\| \leq 1.$$

Since both primal and dual problems are strictly feasible, there exist optimal points for both of them. If $\lambda = \tau$ at the optimum, then $Ax = b$, and

$$\lambda = \tau = \sqrt{\|x\|^2 + 1}.$$

In this case, the optimal x is the (unique) minimum-norm solution to $Ax = b$: $x = A^\dagger b$.

Now assume $\lambda > \tau$. Again, both primal and dual problems are strictly feasible; therefore, the primal- and dual-optimal objectives are equal:

$$(18) \quad \|Ax - b\| + \left\|\begin{bmatrix} x^T & 1 \end{bmatrix}\right\| = \lambda = b^T z - v = -(Ax - b)^T z - \begin{bmatrix} x^T & 1 \end{bmatrix}\begin{bmatrix} -A^T z \\ v \end{bmatrix}.$$

Using $\|z\| \le 1$, $\|[u^T \ v]^T\| \le 1$, $u = -A^T z$, we get

$$z = -\frac{Ax - b}{\|Ax - b\|} \text{ and } \begin{bmatrix} u^T & v \end{bmatrix} = -\frac{\begin{bmatrix} x^T & 1 \end{bmatrix}}{\sqrt{\|x\|^2 + 1}}.$$

Replace these values in $A^T z + u = 0$ to obtain the expression of the optimal x:

$$x = \left(A^T A + \mu I\right)^{-1} A^T b, \text{ with } \mu = \frac{\lambda - \tau}{\tau} = \frac{\|Ax - b\|}{\sqrt{\|x\|^2 + 1}}. \qquad \square$$

REMARK 3.1. *When $\lambda > \tau$, the RLS solution can be interpreted as the solution of a weighted LS problem for an augmented system:*

$$x_{\mathrm{RLS}} = \arg\min \left\|\begin{bmatrix} A \\ I \\ 0 \end{bmatrix} x - \begin{bmatrix} b \\ 0 \\ 1 \end{bmatrix}\right\|_\Theta,$$

where $\Theta = \mathbf{diag}((\lambda-\tau)I, \tau I, \tau)$. The RLS method amounts to computing the weighting matrix Θ that is optimal for robustness via the SOCP (15). We shall encounter a generalization of the above formula for the linear-fractional SRLS problem of section 5.

REMARK 3.2. *It is possible to solve the problem when only A is perturbed ($\Delta b = 0$). In this case, the worst-case residual is $\|Ax - b\| + \|x\|$, and the optimal x is determined by (17), where $\mu\|x\| = \|Ax - b\|$. (See the example in section 7.2.)*

3.3. Reduction to a one-dimensional search. When the SVD of A is available, we can use it to reduce the problem to a one-dimensional convex differentiable problem. The following analysis will also be useful in section 6.

Introduce the SVD of A and a related decomposition for b:

$$A = U \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} V^T, \quad U^T b = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix},$$

where $\Sigma = \mathbf{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbf{R}^{r \times r}$, $\Sigma > 0$, and $b_1 \in \mathbf{R}^r$, $r = \mathbf{Rank}A$.

Assume that $\lambda > \tau$ at the optimum of problem (15). From (18), we have

$$\lambda = b^T z - v = \frac{b^T (b - Ax)}{\|Ax - b\|} + \frac{1}{\sqrt{\|x\|^2 + 1}}$$

$$= \frac{1}{\tau} + \frac{b_2^T b_2}{\lambda - \tau} + b_1^T ((\lambda - \tau)I + \tau\Sigma^2)^{-1} b_1.$$

Since $\lambda = 0$ is never feasible, we may define $\theta = \tau/\lambda$. Multiplying by λ, we obtain that

$$\lambda^2 = \frac{1}{\theta} + \frac{b_2^T b_2}{1 - \theta} + b_1^T((1 - \theta)I + \theta\Sigma^2)^{-1}b_1.$$

From $\lambda \leq \|b\| + 1$ and $\tau \geq 1$, we deduce $\theta \geq \theta_{\min} \triangleq 1/(\|b\| + 1)$. Thus, the optimal worst-case residual is

(19) $$\phi(A, b)^2 = \inf_{\theta_{\min} \leq \theta < 1} f(\theta),$$

where f is the following function:

(20) $$f(\theta) \triangleq \begin{cases} \dfrac{1}{\theta} + b^T((1 - \theta)I + \theta AA^T)^{-1}b & \text{if } \theta_{\min} \leq \theta < 1, \\ \infty & \text{if } \theta = 1, \ b \notin \mathbf{Range}(A), \\ 1 + \|A^\dagger b\|^2 & \text{if } \theta = 1, \ b \in \mathbf{Range}(A). \end{cases}$$

The function f is convex and twice differentiable on $[\theta_{\min} \ 1[$. If $b \notin \mathbf{Range}(A)$, f is infinite at $\theta = 1$; otherwise, f is twice differentiable on the closed interval $[\theta_{\min} \ 1]$. Therefore, the minimization of f can be done using standard Newton methods for differentiable optimization.

THEOREM 3.3. *When $\rho = 1$, the solution of the unstructured RLS can be computed by solving the one-dimensional convex differentiable problem* (19) *or by computing the unique real root inside $[\theta_{\min} \ 1]$ (if any) of the equation*

$$\frac{1}{\theta^2} = \frac{\|b_2\|^2}{(1 - \theta)^2} + \sum_{i=1}^{r} \frac{b_{1i}^2(1 - \sigma_i^2)}{(1 + \theta(\sigma_i^2 - 1))^2}.$$

The above theorem yields an alternative method for computing the RLS solution. This method is similar to the one given in [5]. A related approach was used for quadratically constrained LS problems in [19].

The above solution, which requires one SVD of A, has cost $O(nm^2 + m^3)$. The SOCP method is only a few times more costly (see the end of section 3.1), with the advantage that we can include all kinds of additional constraints on x (nonnegativity and/or quadratic constraints, etc.) in the SOCP (15), with low additional cost. Also, the SVD solution does not extend to the structured case considered in section 4.

3.4. Robustness of LS solution. It is instructive to know when the RLS and LS solutions coincide, in which case we can say the LS solution is robust. This happens if and only if the optimal θ in problem (19) is equal to 1. The latter implies $b_2 = 0$ (that is, $b \in \mathbf{Range}(A)$). In this case, f is differentiable at $\theta = 1$, and its minimum over $[\theta_{\min} \ 1]$ is at $\theta = 1$ if and only if

$$\frac{df}{d\theta}(1) = b_1^T \Sigma^{-4} b_1 - (1 + b_1^T \Sigma^{-2} b_1) \leq 0.$$

We obtain a necessary and sufficient condition for the optimal θ to be equal to 1. This condition is

(21) $$b \in \mathbf{Range}(A), \quad b^T(AA^T)^{2\dagger}b \leq 1 + b^T(AA^T)^\dagger b.$$

If (21) holds, then the RLS and LS solutions coincide. Otherwise, the optimal $\theta < 1$, and x is given by (17). We may write the latter condition in the case when the norm bound of the perturbation ρ is different from 1 as the following: $\rho > \rho_{\min}$, where

$$(22) \qquad \rho_{\min}(A, b) \triangleq \begin{cases} \dfrac{\sqrt{1 + \|A^\dagger b\|^2}}{\|(AA^T)^\dagger b\|} & \text{if } b \in \mathbf{Range}(A), \ A \neq 0, \ b \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $\rho_{\min}$ can be interpreted as the perturbation level that the LS solution allows. We note that when $b \in \mathbf{Range}(A)$, the LS and TLS solutions also coincide.

COROLLARY 3.4. *The LS, TLS, and RLS solutions coincide whenever the norm bound on the perturbation matrix ρ satisfies $\rho \leq \rho_{\min}(A, b)$, where $\rho_{\min}(A, b)$ is defined in (22). Thus, $\rho_{\min}(A, b)$ can be seen as a robustness measure of the LS (or TLS) solution.*

When A is full rank, the robustness measure $\rho_{\min}$ is nonzero and decreases as the condition number of A increases.

REMARK 3.3. *We note that the TLS solution x_{TLS} is the most accurate, in the sense that it minimizes the distance function (see [18])*

$$a(A, b, x) = \frac{\|Ax - b\|}{\sqrt{\|x\|^2 + 1}},$$

and is the least robust, in the sense of the worst-case residual. The LS solution, $x_{\mathrm{LS}} = A^\dagger b$, is intermediate (in the sense of accuracy and robustness). In fact, it can be shown that

$$r(A, b, x_{\mathrm{RLS}}, \rho) \leq r(A, b, x_{\mathrm{LS}}, \rho) \leq r(A, b, x_{\mathrm{TLS}}, \rho),$$
$$a(A, b, x_{\mathrm{TLS}}) \leq a(A, b, x_{\mathrm{LS}}) \leq a(A, b, x_{\mathrm{RLS}}),$$
$$\|x_{\mathrm{RLS}}\| \leq \|x_{\mathrm{LS}}\| \leq \|x_{\mathrm{TLS}}\|.$$

3.5. RLS and TLS. The RLS framework assumes that the data matrices (A, b) are the "nominal" values of the model, which are subject to unstructured perturbation, bounded in norm by ρ. Now, if we think of (A, b) as "measured" data, the assumption that (A, b) correspond to a nominal model may not be judicious. Also, in some applications, the norm bound ρ on the perturbation may be hard to estimate. The TLS solution, when it exists, can be used in conjunction with RLS to address this issue.

Assume that the TLS problem has a solution. Let ΔA_{TLS}, Δb_{TLS}, x_{TLS} be minimizers of the TLS problem

$$\text{minimize } \|\Delta A \ \Delta b\|_F \text{ subject to } (A + \Delta A)x = b + \Delta b,$$

and let

$$\rho_{\mathrm{TLS}} = \|\Delta A_{\mathrm{TLS}} \ \Delta b_{\mathrm{TLS}}\|_F, \quad A_{\mathrm{TLS}} = A + \Delta A_{\mathrm{TLS}}, \quad b_{\mathrm{TLS}} = A + \Delta b_{\mathrm{TLS}}.$$

TLS finds a consistent, linear system that is closest (in Frobenius norm sense) to the observed data (A, b). The underlying assumption is that the observed data (A, b) is the result of a consistent, linear system which, under the measurement process, has been subjected to unstructured perturbations, unknown but bounded in norm by ρ_{TLS}. With this assumption, any point of the ball

$$\{(A', b') \mid \|A' - A_{\mathrm{TLS}} \ b' - b_{\mathrm{TLS}}\|_F \leq \rho_{\mathrm{TLS}}\}$$

can be observed, just as well as (A, b). Thus, TLS computes an "uncertain linear system" representation of the observed phenomenon: $(A_{\mathrm{TLS}}, b_{\mathrm{TLS}})$ is the nominal model, and ρ_{TLS} is the perturbation level.

Once this uncertain system representation $(A_{\mathrm{TLS}}, b_{\mathrm{TLS}}, \rho_{\mathrm{TLS}})$ is computed, choosing x_{TLS} as a "solution" to $Ax \simeq b$ amounts to finding the exact solution to the nominal system. Doing so, we compute a very accurate solution (with zero residual), which does not take into account the perturbation level ρ_{TLS}. A more robust solution is given by the solution to the following RLS problem:

$$(23) \qquad \min_x \max_{\|\Delta A \ \Delta b\|_F \leq \rho_{\mathrm{TLS}}} \|(A_{\mathrm{TLS}} + \Delta A)x - (b_{\mathrm{TLS}} + \Delta b)\|.$$

The solution to the above problem coincides with the TLS one (that is, in our case, with x_{TLS}) when $\rho_{\mathrm{TLS}} \leq \rho_{\min}(A_{\mathrm{TLS}}, b_{\mathrm{TLS}})$. (Since $b_{\mathrm{TLS}} \in \mathbf{Range}(A_{\mathrm{TLS}})$, the latter quantity is strictly positive, except when $A_{\mathrm{TLS}} = 0$, $b_{\mathrm{TLS}} = 0$.)

With standard LS, the perturbations that account for measurement errors are structured (with $\Delta A = 0$). To be consistent with LS, one should consider the following RLS problem instead of (23):

$$(24) \qquad \min_x \max_{\|\Delta b\| \leq \rho_{\mathrm{LS}}} \|A_{\mathrm{LS}}x - (b_{\mathrm{LS}} + \Delta b)\|.$$

It turns out that the above problem yields the same solution as LS itself.

To summarize, RLS can be used in conjunction with TLS for "solving" a linear system $Ax \simeq b$. Solve the TLS problem to build an "uncertain linear system" representation $(A_{\mathrm{TLS}}, b_{\mathrm{TLS}}, \rho_{\mathrm{TLS}})$ of the observed data. Then, take the solution x_{RLS} to the RLS problem with the nominal matrices $(A_{\mathrm{TLS}}, b_{\mathrm{TLS}})$, and uncertainty size ρ_{TLS}. Note that computing the TLS solution (precisely, A_{TLS}, b_{TLS}, and ρ_{TLS}) only requires the computation of the smallest singular value and associated singular subspace [17].

4. Structured Robust Least Squares (SRLS). In this section, we consider the SRLS problem, which is to compute

$$(25) \qquad \phi_S(\mathbf{A}, \mathbf{b}, \rho) \triangleq \min_x \max_{\|\delta\| \leq \rho} \|\mathbf{A}(\delta)x - \mathbf{b}(\delta)\|,$$

where $\mathbf{A}, \mathbf{b}$ are defined in (2). As before, we assume with no loss of generality that $\rho = 1$ and denote $r_S(\mathbf{A}, \mathbf{b}, 1, x)$ by $r_S(\mathbf{A}, \mathbf{b}, x)$. Throughout the section, we use the following notation:

$$(26) \qquad M(x) \triangleq \begin{bmatrix} A_1 x - b_1 & \dots & A_p x - b_p \end{bmatrix}.$$

4.1. Computing the worst-case residual. We first examine the problem of computing the worst-case residual $r_S(\mathbf{A}, \mathbf{b}, x)$ for a given $x \in \mathbf{R}^m$. Define

$$(27) \qquad F \triangleq M(x)^T M(x), \quad g \triangleq M(x)^T(A_0 x - b_0), \quad h \triangleq \|A_0 x - b_0\|^2.$$

With the above notation, we have

$$(28) \qquad r_S(\mathbf{A}, \mathbf{b}, x)^2 = \max_{\delta^T \delta \leq 1} \begin{bmatrix} 1 \\ \delta \end{bmatrix} \begin{bmatrix} h & g^T \\ g & F \end{bmatrix} \begin{bmatrix} 1 \\ \delta \end{bmatrix}.$$

Now let $\lambda \geq 0$. Using the $\mathcal{S}$-procedure (Lemma 2.1), we have

$$\begin{bmatrix} 1 \\ \delta \end{bmatrix} \begin{bmatrix} h & g^T \\ g & F \end{bmatrix} \begin{bmatrix} 1 \\ \delta \end{bmatrix} \leq \lambda$$

for every δ, $\delta^T \delta \leq 1$ if and only if there exists a scalar $\tau \geq 0$ such that

$$\left[\begin{array}{c} 1 \\ \delta \end{array} \right] \left[\begin{array}{cc} \lambda - \tau - h & -g^T \\ -g & \tau I - F \end{array} \right] \left[\begin{array}{c} 1 \\ \delta \end{array} \right] \geq 0 \text{ for every } \delta \in \mathbf{R}^p.$$

Using the fact that $\tau \geq 0$ is implied by $\tau I \geq F$, we may rewrite the above condition as

$$(29) \qquad \mathcal{F}(\lambda, \tau) \triangleq \left[\begin{array}{cc} \lambda - \tau - h & -g^T \\ -g & \tau I - F \end{array} \right] \geq 0.$$

The consequence is that the worst-case residual is computed by solving an SDP with two scalar variables. A bit more analysis shows how to reduce the problem to a one-dimensional, convex differentiable problem and how to obtain the corresponding worst-case perturbation.

THEOREM 4.1. *For every x fixed, the squared worst-case residual (for $\rho = 1$) $r_S(\mathbf{A}, \mathbf{b}, x)^2$ can be computed by solving the SDP in two variables*

$$\text{minimize } \lambda \text{ subject to } (29),$$

or, alternatively, by minimizing a one-dimensional convex differentiable function

$$(30) \qquad r_S(\mathbf{A}, \mathbf{b}, x)^2 = h + \inf_{\tau \geq \lambda_{\max}(F)} f(\tau),$$

where

$$f(\tau) \triangleq \left\{ \begin{array}{ll} \tau + g^T(\tau I - F)^{-1} g & \text{if } \tau > \lambda_{\max}(F), \\ \infty & \text{if } \tau = \lambda_{\max}(F) \text{ is } (F, g)\text{-controllable}, \\ \lambda_{\max}(F) + g^T(\tau I - F)^{\dagger} g & \text{if } \tau = \lambda_{\max}(F) \text{ is not } (F, g)\text{-controllable}. \end{array} \right.$$

(31)
If τ is optimal for problem (30), the equations in δ

$$(\tau I - F)\delta = g, \quad \|\delta\| = 1$$

have a solution, any of which is a worst-case perturbation.

Proof. See Appendix A, where we also show how to compute a worst-case perturbation. ☐

4.2. Optimizing the worst-case residual. Using Theorem 4.1, the expression of F, g, h given in (27), and Schur complements, we obtain the following result.

THEOREM 4.2. *When $\rho = 1$, the Euclidean-norm SRLS can be solved by computing an optimal solution (λ, τ, x) of the SDP*

$$(32) \qquad \text{minimize } \lambda \text{ subject to } \left[\begin{array}{ccc} \lambda - \tau & 0 & (A_0 x - b_0)^T \\ 0 & \tau I & M(x)^T \\ A_0 x - b_0 & M(x) & I \end{array} \right] \geq 0,$$

where $M(x)$ is defined in (26).

REMARK 4.1. *Straightforward manipulations show that the results are coherent with the unstructured case.*

Although the above SDP is not directly amenable to the more efficient SOCP formulation, we may devise special interior-point methods for solving the problem. These special-purpose methods will probably have much greater efficiency than general-purpose SDP solvers. This study is left for the future.

REMARK 4.2. *The discussion of section 3.5 extends to the case when the perturbations are structured. TLS problems with (affine) structure constraints on perturbation matrices are discussed in [7]. While the structured version of the TLS problem becomes very hard to solve, the SRLS problem retains polynomial-time complexity.*

5. Linear-fractional SRLS. In this section, we examine a generalization of the SRLS problem. Our framework encompasses the case when the functions $\mathbf{A}(\delta)$, $\mathbf{b}(\delta)$ are rational. We show that the computation of the worst-case residual is NP-complete but that upper bounds can be computed (and optimized) using SDP. First, we need to motivate the problem and develop a formalism for posing it. This formalism was introduced by Doyle et al. [9] in the context of robust identification.

5.1. Motivations. In some structured robust least-squares problems such as (3), it may not be convenient to measure the perturbation size with Euclidean norm. Indeed, the latter implies a correlated bound on the perturbation. One may instead consider an SRLS problem in which the bounds are not correlated; that is, the perturbation size in (3) is measured by the maximum norm

$$(33) \qquad \min_{x} \max_{\|\delta\|_\infty \leq 1} \|\mathbf{A}(\delta)x - \mathbf{b}(\delta)\|.$$

Also, in some RLS problems, we may assume that some columns of $[A\ b]$ are perfectly known. For instance, the error $[\Delta A\ \Delta b]$ has the form $[\Delta A\ 0]$, where ΔA is bounded and otherwise unknown. More generally, we may be interested in SRLS problems, where the perturbed data matrices write

$$(34) \qquad \left[\begin{array}{cc} \mathbf{A}(\Delta) & \mathbf{b}(\Delta) \end{array}\right] = \left[\begin{array}{cc} A & b \end{array}\right] + L\Delta \left[\begin{array}{cc} R_A & R_b \end{array}\right],$$

where A, b, L, R_A, R_b are given matrices, and Δ is a (full) norm-bounded matrix. In such a problem, the perturbation is not structured, except via the matrices L, R_A, R_b. (Note that a special case of this problem is solved in [5].)

Finally, we may be interested in SRLS problems in which the matrix functions $\mathbf{A}(\delta)$, $\mathbf{b}(\delta)$ in (3) are rational functions of the parameter vector δ. One example is given in section 7.6.

It turns out that the extensions described in the three preceding paragraphs can be addressed using the same formalism, which we now detail.

5.2. Problem definition. Let $\mathcal{D}$ be a subspace of $\mathbf{R}^{N \times N}$, $A \in \mathbf{R}^{n \times m}$, $b \in \mathbf{R}^n$, $L \in \mathbf{R}^{n \times N}$, $R_A \in \mathbf{R}^{N \times m}$, $R_b \in \mathbf{R}^N$, $D \in \mathbf{R}^{N \times N}$. For every $\Delta \in \mathcal{D}$ such that $\det(I - D\Delta) \neq 0$, we define the matrix functions

$$\left[\begin{array}{cc} \mathbf{A}(\Delta) & \mathbf{b}(\Delta) \end{array}\right] = \left[\begin{array}{cc} A & b \end{array}\right] + L\Delta(I - D\Delta)^{-1}\left[\begin{array}{cc} R_A & R_b \end{array}\right].$$

For a given $x \in \mathbf{R}^m$, we define the worst-case residual by

$$(35) \quad r_{\mathcal{D}}(\mathbf{A}, \mathbf{b}, \rho, x) \triangleq \left\{ \begin{array}{ll} \displaystyle\max_{\Delta \in \mathcal{D},\ \|\Delta\| \leq \rho} \|\mathbf{A}(\Delta)x - \mathbf{b}(\Delta)\| & \text{if } \det(I - D\Delta) \neq 0, \\ \infty & \text{else.} \end{array} \right.$$

We say that x is an SRLS solution if x minimizes the worst-case residual above. As before, we assume $\rho = 1$ with no loss of generality and denote $r_{\mathcal{D}}(\mathbf{A}, \mathbf{b}, 1, x)$ by $r_{\mathcal{D}}(\mathbf{A}, \mathbf{b}, x)$.

The above formulation encompasses the three situations referred to in section 5.1. First, the maximum-norm SRLS problem (33) is readily transformed into problem (35) as follows. Let $L_i \in \mathbf{R}^{n \times N}$, $R_i \in \mathbf{R}^{N \times (m+1)}$ be such that $[A_i\ b_i] = L_i R_i$, $\mathbf{Rank}L_i = \mathbf{Rank}R_i = r_i$, where $r_i = \mathbf{Rank}[A_i\ b_i]$. Set $D = 0$, and let

$$(36) \qquad \begin{array}{c} L = \left[\begin{array}{ccc} L_1 & \dots & L_p \end{array}\right],\ R^T = \left[\begin{array}{ccc} R_1^T & \dots & R_p^T \end{array}\right], \\ \mathcal{D} = \{\bigoplus_{i=1}^{p} \delta_i I_{s_i} \mid \delta_i \in \mathbf{R},\ 1 \leq i \leq p\}. \end{array}$$

Problem (33) can be formulated as the minimization of (35), with $\mathcal{D}$ defined as above.

Also, we recover the case when the perturbed matrices write as in (34) when we allow Δ to be any full matrix (that is, $\mathcal{D} = \mathbf{R}^{N \times N}$). In particular, we recover the unstructured RLS problem of section 3 as follows. Assume $n > m$. We have

$$
\left[\begin{array}{cc} \Delta A & \Delta b \end{array} \right] = L \left[\begin{array}{ccc} \Delta A & \Delta b & \times \end{array} \right] R,
$$

where $L = I$, $R^T = [I\ 0]$. (The symbol $\times$ refers to dummy elements that are added to the perturbation matrix in order to make it a square, $n \times n$ matrix.) In this case, the perturbation set $\mathcal{D}$ is $\mathbf{R}^{n \times n}$.

Finally, the case when $\mathbf{A}(\delta)$ and $\mathbf{b}(\delta)$ are rational functions of a vector δ (well defined over the unit ball $\{\delta \mid \|\delta\|_\infty \leq 1\}$) can be converted (in polynomial time) into the above framework (see, e.g., [48] for a conversion procedure). We give an example of such a conversion in section 7.6.

5.3. Complexity analysis. In comparison with the SRLS problem of section 4, the linear-fractional SRLS problem offers two levels of increased complexity.

First, checking whether the worst-case residual is finite is NP-complete [6]. The linear-fractional dependence (that is, $D \neq 0$) is a first cause of increased complexity.

The SRLS problem above remains hard even when matrices $\mathbf{A}(\delta)$, $\mathbf{b}(\delta)$ depend affinely on the perturbation elements ($D = 0$). Consider, for instance, the SRLS problem with $D = 0$ and in which $\mathcal{D}$ is defined as in (36). In this case, the problem of computing the worst-case residual can be formulated as

$$
\max_{\|\delta\|_\infty \leq 1} \left[\begin{array}{cc} 1 & \delta \end{array} \right] \left[\begin{array}{cc} h & g^T \\ g & F \end{array} \right] \left[\begin{array}{c} 1 \\ \delta \end{array} \right]
$$

for appropriate F, g, h. The only difference with the worst-case residual defined in (28) is the norm used to measure perturbation. Computing the above quantity is NP-complete (it is equivalent to a MAX CUT problem [36, 38]). The following lemma, which we provide for the sake of completeness, is a simple corollary of a result by Nemirovsky [32].

LEMMA 5.1. *Consider the problem* $\mathcal{P}(\mathbf{A}, \mathbf{b}, \mathcal{D}, x)$ *defined as follows: given a positive rational number* λ, *matrices* A, b, L, R_A, R_b, D *of appropriate size, and an m-vector* x, *all with rational entries, and a linear subset* $\mathcal{D}$, *determine whether* $r_\mathcal{D}(\mathbf{A}, \mathbf{b}, x)$ $\leq \lambda$. *Problem* $\mathcal{P}(\mathbf{A}, \mathbf{b}, \mathcal{D}, x)$ *is NP-complete.*

Proof. See Appendix B. ☐

5.4. An upper bound on the worst-case residual. Although our problem is NP-complete, we can minimize upper bounds in polynomial time using SDP. Introduce the following linear subspaces:

(37)
$$
\begin{aligned}
\mathcal{B} &\triangleq \left\{ B \in \mathbf{R}^{N \times N} \mid B\Delta = \Delta B \text{ for every } \Delta \in \mathcal{D} \right\}, \\
\mathcal{S} &\triangleq \left\{ S \in \mathcal{B} \mid S = S^T \right\}, \ \mathcal{G} \triangleq \left\{ G \in \mathcal{B} \mid G = -G^T \right\}.
\end{aligned}
$$

Let $\lambda \in \mathbf{R}$. The inequality $\lambda > r_\mathcal{D}(\mathbf{A}, \mathbf{b}, x)$ holds if and only if, for every $\Delta \in \mathcal{D}$, $\|\Delta\| \leq 1$, we have $\det(I - D\Delta) \neq 0$ and

$$
\begin{aligned}
& \left[\begin{array}{cc} \lambda I & Ax - b \\ (Ax - b)^T & \lambda \end{array} \right] + \left[\begin{array}{c} L \\ 0 \end{array} \right] \Delta (I - D\Delta)^{-1} \left[\begin{array}{cc} 0 & R_A x - R_b \end{array} \right] \\
& + \left[\begin{array}{c} 0 \\ (R_A x - R_b)^T \end{array} \right] (I - D\Delta)^{-T} \Delta^T \left[\begin{array}{cc} L^T & 0 \end{array} \right] > 0.
\end{aligned}
$$

Using Lemma 2.3, we obtain that $\lambda > r_{\mathcal{D}}(\mathbf{A}, \mathbf{b}, x)$ holds if there exist $S \in \mathcal{S}$, $G \in \mathcal{G}$, such that

$$(38) \qquad \mathcal{F}(\lambda, S, G, x) = \left[\begin{array}{cc|c} & \Theta & \begin{array}{c} Ax - b \\ R_A x - R_b \end{array} \\ \hline (Ax - b)^T & (R_A x - R_b)^T & \lambda \end{array}\right] > 0,$$

where

$$(39) \qquad \Theta \triangleq \left[\begin{array}{cc} \lambda I - LSL^T & -LSD^T + LG \\ -DSL^T + G^T L^T & S + DG - GD^T - DSD^T \end{array}\right].$$

Minimizing λ subject to the above semidefinite constraint yields an upper bound for $r_{\mathcal{D}}(\mathbf{A}, \mathbf{b}, x)$. It turns out that the above estimate of the worst-case residual is actually *exact* in some "generic" sense.

THEOREM 5.2. *When $\rho = 1$, an upper bound on the worst-case residual $r_{\mathcal{D}}(\mathbf{A}, \mathbf{b}, x)$ can be obtained by solving the SDP*

$$(40) \qquad \inf_{S,G,\lambda} \lambda \text{ subject to } S \in \mathcal{S}, \ G \in \mathcal{G}, \ (38).$$

The upper bound is exact when $\mathcal{D} = \mathbf{R}^{N \times N}$. If $\Theta > 0$ at the optimum, the upper bound is also exact.

Proof. See Appendix C.   □

5.5. Optimizing the worst-case residual. Since x appears linearly in the constraint (38), we may optimize the worst-case residual's upper bound using SDP. We may reduce the number of variables appearing in the previous problem, using the elimination Lemma 2.4. Inequality in (38) can be written as in (11) with

$$W = \left[\begin{array}{c|c} \Theta & \begin{array}{c} -b \\ -R_b \end{array} \\ \hline -b \quad -R_b & \lambda \end{array}\right], \quad U = \left[\begin{array}{c} A \\ R_A \\ 0 \end{array}\right], \quad V = \left[\begin{array}{c} 0 \\ 0 \\ 1 \end{array}\right],$$

where Θ is defined in (39).

Denote by $\mathcal{N}$ the orthogonal complement of $[A^T \ R_A^T]^T$. Using the elimination Lemma 2.4, we obtain an equivalent condition for (38) to hold for some $x \in \mathbf{R}^m$; namely,

$$(41) \ S \in \mathcal{S}, \ G \in \mathcal{G}, \ \Theta > 0, \ (\mathcal{N} \oplus 1)^T \left[\begin{array}{c|c} \Theta & \begin{array}{c} -b \\ -R_b \end{array} \\ \hline -b \quad -R_b & \lambda \end{array}\right] (\mathcal{N} \oplus 1) > 0.$$

For every λ, S, G that are strictly feasible for the above constraints, an x that satisfies (38) is given, when R_A is full rank, by

$$(42) \qquad x = \left([\ A^T \ \ R_A^T \ ] \Theta^{-1} \left[\begin{array}{c} A \\ R_A \end{array}\right]\right)^{-1} [\ A^T \ \ R_A^T \ ] \Theta^{-1} \left[\begin{array}{c} b \\ R_b \end{array}\right].$$

(To prove this, we applied formula (13) and took $\sigma \to \infty$.)

THEOREM 5.3. *When $\rho = 1$, an upper bound on the optimal worst-case residual can be obtained by solving the SDP*

$$(43) \qquad \inf_{S,G,\lambda,x} \lambda \text{ subject to } S \in \mathcal{S}, \ G \in \mathcal{G}, \ (38),$$

or, alternatively, the SDP

$$(44) \qquad\qquad \inf_{S,G,\lambda} \lambda \ \textit{subject to (41)}.$$

The upper bound is always exact when $\mathcal{D} = \mathbf{R}^{N \times N}$. If $\Theta > 0$ at the optimum, the upper bound is also exact. The optimal x is then unique and given by (42) when R_A is full rank.

Proof. See Appendix C. □

REMARK 5.1. *In parallel to the unstructured case (see Remark 3.1), the linear-fractional SRLS can be interpreted as a weighted LS for an augmented system. Precisely, when $\Theta > 0$, the linear-fractional SRLS solution can be interpreted as the solution of a weighted LS problem*

$$x_{\mathrm{SRLS}} \in \arg\min \left\| \begin{bmatrix} A \\ R_A \end{bmatrix} x - \begin{bmatrix} b \\ R_b \end{bmatrix} \right\|_{\Theta}.$$

The SRLS method amounts to computing the weighting matrix Θ that is optimal for robustness.

REMARK 5.2. *Our results are coherent with the unstructured case: replace L by I, R by $[I\ 0]^T$, variable S by τI, and set $G = 0$. The parameter μ of Theorem 3.2 can be interpreted as the Schur complement of $\lambda I - LSL^T$ in the matrix Θ.*

REMARK 5.3. *We emphasize that the above results are exact (nonconservative) when the perturbation structure is full. In particular, we recover (and generalize) the results of [5] in the case when only some columns of A are affected by otherwise unstructured perturbations.*

REMARK 5.4. *When $D = 0$, it is possible to use the approximation method of [16] to obtain solutions (based on the SDP relaxations given in Theorem 5.3) that have expected value within 14% of the true value.*

6. Link with regularization. The standard LS solution x_{LS} is very sensitive to errors in A, b when A is ill conditioned. In fact, the LS solution might not be a continuous function of A, b when A is near deficient. This has motivated many researchers to look for ways to regularize the LS problem, which is to make the solution x unique and continuous in the data matrices (A, b). In this section, we briefly examine the links of our RLS and SRLS solution with regularization methods for standard LS.

Beforehand, we note that since all our problems are formulated as SDPs, we could invoke the quite complete sensitivity analysis results obtained by Bonnans, Cominetti, and Shapiro [3]. The application of these general results to our SDPs is considered in [35].

6.1. Regularization methods for LS. Most regularization methods for LS require imposing an additional bound on the solution vector x. One way is to minimize $\|Ax - b\|^2 + \Omega(x)$, where Ω is some squared norm (see [23, 43, 8]). Another way is to use constrained least squares (see [18, pp. 561–571]).

In a classical Tikhonov regularization method, $\Omega(x) = \mu\|x\|^2$, where $\mu > 0$ is some "regularization" parameter. The modified value of x is obtained by solving an augmented LS problem

$$(45) \qquad\qquad \text{minimize } \|Ax - b\|^2 + \mu\|x\|^2$$

and is given by

$$(46) \qquad x(\mu) = (\mu I + A^T A)^{-1} A^T b.$$

(Note that for every $\mu > 0$, the above x is continuous in (A, b).)

The above expression also arises in the Levenberg–Marquardt method for optimization or in the Ridge regression problem [17]. As mentioned in [18], the choice of an appropriate μ is problem dependent and in many cases not obvious.

In more elaborate regularization schemes of the Tikhonov type, the identity matrix in (46) is replaced with a positive semidefinite weighting matrix (see for instance [31, 8]). Again, this can be interpreted as a (weighted) least-squares method for an augmented system.

6.2. RLS and regularization. Noting the similarity between (17) and (46), we can interpret the (unstructured) RLS method as that of Tikhonov regularization. The following theorem yields an estimate of the "smoothing effect" of the RLS method. Note that improved regularity results are given in [35].

THEOREM 6.1. *The (unique) RLS solution x_{RLS} and the optimal worst-case residual are continuous functions of the data matrices A, b. Furthermore, if $\mathcal{K}$ is a compact set of $\mathbf{R}^n$, and $d_{\mathcal{K}} = \max\{\|b\| \mid b \in \mathcal{K}\}$, then for every uncertainty size $\rho > 0$, the function*

$$\begin{aligned} \mathbf{R}^{n \times m} \times \mathcal{K} &\longrightarrow [1\ d_{\mathcal{K}} + 1], \\ (A, b) &\longmapsto \phi(A, b, \rho) \end{aligned}$$

is Lipschitzian, with Lipschitz constant $1 + d_{\mathcal{K}}/\rho$.

Theorem 6.1 shows that any level of robustness (that is, any norm bound on perturbations $\rho > 0$) guarantees regularization. We describe in section 7 some numerical examples that illustrate our results.

REMARK 6.1. *In the RLS method, the Tikhonov regularization parameter μ is chosen by solving a second-order cone problem in such a way that μ is optimal for robustness. The cost of the RLS solution is equal to the cost of solving a small number of least-squares problems of the same size as the classical Tikhonov regularization problem (45).*

REMARK 6.2. *The equation that determines μ in the RLS method is*

$$\mu = \frac{\|Ax(\mu) - b\|}{\rho\sqrt{\|x(\mu)\|^2 + 1}}.$$

This choice resembles Miller's choice [30], where μ is determined recursively by the equations

$$\mu = \frac{\|Ax(\mu) - b\|}{\rho\|x(\mu)\|}.$$

This formula arises in RLS when there is no perturbation in b (see Remark 3.2). Thus, Miller's solution corresponds to an RLS problem in which the perturbation affects only the columns of A. We note that this solution is not necessarily regular (continuous).

TLS deserves a special mention here. When the TLS problem has a solution, it is given by $x_{\mathrm{TLS}} = (A^T A - \sigma^2 I)^{-1} A^T b$, where σ is the smallest singular value of $[A\ b]$. This corresponds to $\mu = -\sigma^2$ in (46). The negative value of μ implies that the TLS is a "deregularized" LS, a fact noted in [17]. In view of our link between regularization and robustness, the above is consistent with the fact that RLS trades off the accuracy of TLS with robustness and regularity, at the expense of introducing bias in the solution. See also Remark 3.3.

6.3. SRLS and regularization. Similarly, we may ask whether the solution to the SRLS problem of section 4 is continuous in the data matrices A_i, b_i, as was the case for unstructured RLS problems. We only discuss continuity of the optimal worst-case residual with respect to (A_0, b_0) (in many problems, the coefficient matrices A_i, b_i for $i = 1, \ldots, p$ are fixed).

In view of Theorem 4.2, continuity holds if the feasible set of the SDP (32) is bounded. Obviously, the objective λ is bounded above by

$$\max_{\delta^T \delta \leq 1} \left\| b_0 + \sum_{i=1}^p \delta_i b_i \right\| \leq \|b_0\| + \sum_{i=1}^p \|b_i\|.$$

Thus the variable τ is also bounded, as (32) implies $0 \leq \tau \leq \lambda$. With λ, τ bounded above, we see that (32) implies that x is bounded if

$$\left\| \begin{array}{cccc} A_0 x - b_0 & A_1 x - b_1 & \ldots & A_p x - b_p \end{array} \right\| \text{ bounded implies } x \text{ bounded.}$$

The above property holds if and only if $[A_0^T \ A_1^T \ldots A_p^T]^T$ is full rank.

THEOREM 6.2. *A sufficient condition for continuity of the optimal worst-case residual (as a function of (A_0, b_0)) is that $[A_1^T \ldots A_p^T]^T$ is full rank.*

6.4. Linear-fractional SRLS and regularization. Precise conditions for continuity of the optimal upper bound on worst-case residual in the linear-fractional case are not known. We may, however, regularize this quantity using a method described in [29] for a related problem. For a given $\epsilon > 0$, define the bounded set

$$\mathcal{S}_\epsilon \triangleq \left\{ S \in \mathcal{S} \ \bigg| \ \epsilon I \leq S \leq \frac{1}{\epsilon} I \right\},$$

where $\mathcal{S}$ is defined in (37). It is easy to show that restricting the condition number of variable S also bounds the variable G in the SDP (44). This yields the following result.

THEOREM 6.3. *An upper bound on the optimal worst-case residual can be obtained by computing the optimal value $\lambda(\epsilon)$ of the SDP*

$$(47) \qquad \min_{S, G, \lambda} \lambda \text{ subject to } S \in \mathcal{S}_\epsilon, \ G \in \mathcal{G}, \ (41).$$

The corresponding upper bound is a continuous function of $[A \ b]$. As $\epsilon \to 0$, the corresponding optimal value $\lambda(\epsilon)$ has a limit, equal to the optimal value of SDP (44).

As noted in Remark 5.1, the linear-fractional SRLS can be interpreted as a weighted LS and so can the above regularization method. Thus, the above method belongs to the class of Tikhonov (or weighted LS) regularization methods referred to in section 6.1, the weighting matrix being optimal for robustness.

7. Numerical examples. The following numerical examples were obtained using two different codes: for SDPs, we used the code **SP** [45], and a MATLAB interface to **SP** called **LMITOOL** [10]. For the (unstructured) RLS problems, we used the SOCP described in [28].

7.1. Complexity estimates of RLS. We first did "large-scale" experiments for the RLS problem in section 3. As mentioned in section 2.1, the number of iterations is almost independent of the size of the problem for SOCPs. We have solved problem (15) for uniformly generated random matrices A and vectors b with various sizes of n, m. Figure 1 shows the average number of iterations as well as the minimum and maximum number of iterations for various values of n, m. The experiments confirm the fact that the number of iterations is almost independent of problem size for the RLS problem.
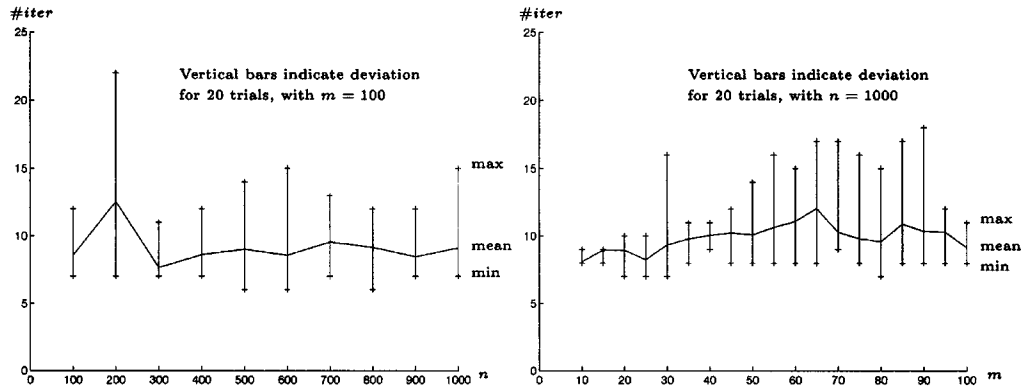
FIG. 1. *Average, minimum, and maximum number of iterations for various RLS problems using the SOCP formulation. In the left figure, we show these numbers for values of n ranging from* 100 *to* 1000. *For each value of n, the vertical bar indicates the minimum and maximum values obtained with* 20 *trials of A, b, with m =* 100. *In the right figure, we show these numbers for values of m ranging from* 11 *to* 100. *For each value of n, the vertical bar indicates the minimum and maximum values obtained with* 20 *trials of A, b, with n =* 1000. *For both plots, the plain curve is the mean value.*

7.2. LS, TLS, and RLS. We now compare the LS, TLS, and RLS solutions for

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}^T, \ b = \begin{bmatrix} 3 & 7 & 1 & 3 \end{bmatrix}^T.$$

On the left and right plots in Fig. 2, we show the four points (A_i, b_i) indicated with "+" signs, and the corresponding linear fits for LS problems (solid line), TLS problems (dotted line), and RLS problems for $\rho = 1, 2$ (dashed lines). The left plot gives the RLS solution with perturbations $[A + \Delta A, b + \Delta b]$, whereas the right plot considers perturbation in A only, $[A + \Delta A, b]$. In both plots, the worst-case points for the RLS solution are indicated by "o" for $\rho = 1$ and "$*$" for $\rho = 2$. As ρ increases, the slope of the RLS solution decreases and goes to zero when $\rho \to \infty$. The plot confirms Remark 3.3: the TLS solution is the most accurate and the least robust, and LS is intermediate.

In the case when we have perturbations in A only (right plot), we obtain an instance of a linear-fractional SRLS (with a full perturbation matrix), as mentioned in section 5.1. (It is also possible to solve this problem directly, as in section 3.) In this last case, of course, the worst-case perturbation can only move along the A-axis.

7.3. RLS and regularization. As mentioned in section 6, we may use RLS to regularize an ill-conditioned LS problem. Consider the RLS problem for

$$A = \begin{bmatrix} 3 & 1 & 4 \\ 0 & 1 & 1 \\ -2 & 5 & 3 \\ 1 & 4 & \alpha \end{bmatrix}, \ b = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 3 \end{bmatrix}.$$

The matrix A is singular when $\alpha = 5$.

Figure 3 shows the regularizing effect of the RLS solution. The left (resp., right) figure shows the optimal worst-case residual (resp., norm of RLS solution) as a function of the parameter α for various values of ρ. When $\rho = 0$, we obtain the LS solution. The latter is not a continuous function of α, and both the solution norm
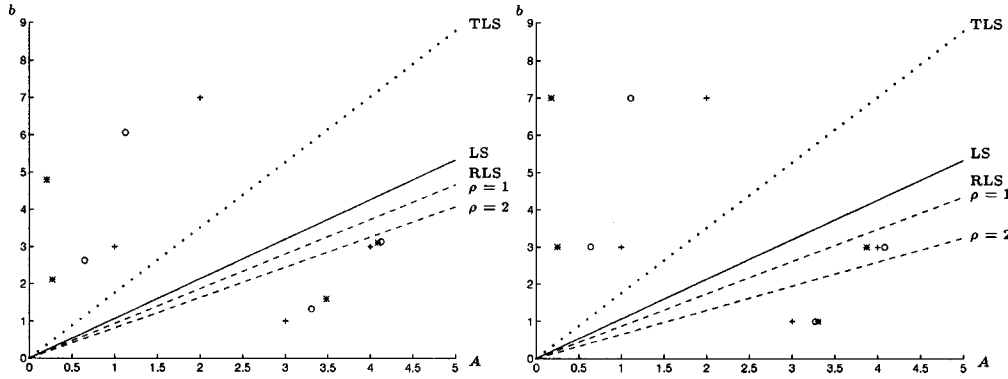
FIG. 2. *Least-squares (solid), total least-squares (dotted), and robust least-squares (dashed) solutions. The + signs correspond to the nominal* [A b]. *The left plot gives the RLS solution with perturbations* [A + ΔA, b + Δb], *whereas the right plot considers perturbation in A only,* [A + ΔA, b]. *The worst-case perturbed points for the RLS solution are indicated by "o" for ρ = 1 and "∗" for ρ = 2.*
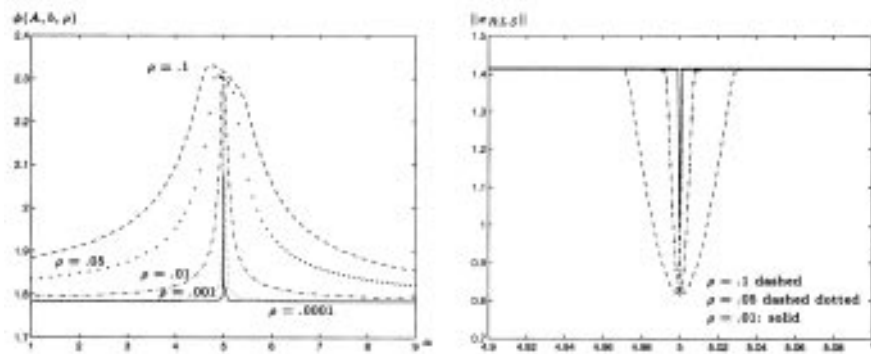

FIG. 3. *Optimal worst-case residual and norm of RLS solution versus α for various values of perturbation level ρ. For ρ = 0 (standard LS), the optimal residual and solution are discontinuous. The spike is smoothed as more robustness is asked for (that is, when ρ increases). On the right plot the curves for ρ = .001 and .0001 are not visible.*

and residual exhibit a spike for $\alpha = 5$ (when A becomes singular). For $\rho > 0$, the RLS solution is smooth. The spike is more and more flattened as ρ grows, which illustrates Theorem 6.1. For $\rho = \infty$, the optimal worst-case residual becomes flat (independent of α), and equal to $\|b\| + 1$, with $x_{\mathrm{RLS}} = 0$.

7.4. Robustness of LS solution. The next example illustrates that sometimes (precisely, if $b \in \mathbf{Range}(A)$) the LS solution is robust up to the perturbation level $\rho_{\min}$ defined in (22). This "natural" robustness of the LS solution degradates as the condition number of A grows. For $\varepsilon_A > 0$, consider the RLS problem for

$$A = \begin{bmatrix} 1 & 0 \\ 0 & \varepsilon_A \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ .1 \end{bmatrix}.$$

We have considered six values of ε_A (which equals the inverse of the condition number of A) from .05 to .55. Table 1 shows the values of $\rho_{\min}$ (as defined in (22))

TABLE 1
Values of $\rho_{\min}$ for various ε_A.

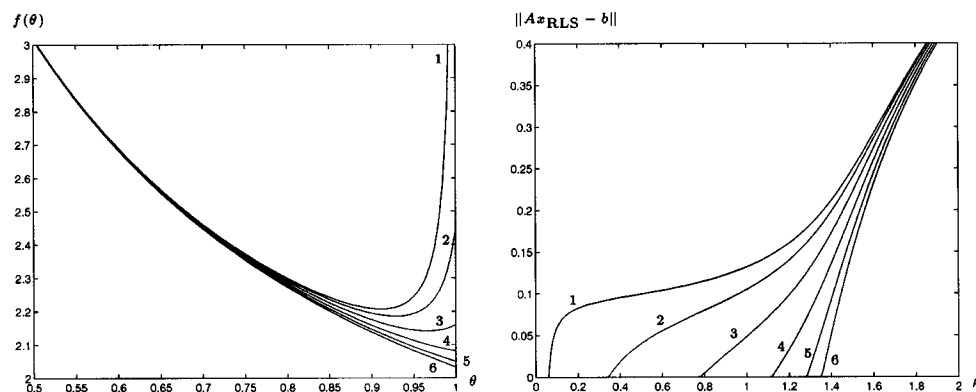| curve #	1	2	3	4	5	6
ε_A	.05	.15	.25	.35	.45	.55
$\rho_{\min}$	0.06	0.34	0.78	1.12	1.28	1.35

FIG. 4. *The left plot shows function $f(\theta)$ (as defined in (20)) for the six values of ε_A (for $\rho = 1$). The right plot gives the optimal RLS residuals versus ρ for the same values of ε_A. The labels $1, \ldots, 6$ correspond to values of ε_A given in Table 1.*

for the six values of ε_A. When the condition number of A grows, the robustness of the LS solution (measured by $\rho_{\min}$) decreases.

The right plot of Fig. 4 gives the worst-case residual versus the robustness parameter ρ for the six values of ε_A. The plot illustrates that for $\rho > \rho_{\min}$, the LS solution (in our case, $A^{-1}b$) differs from the RLS one. Indeed, for each curve, the residual remains equal to zero as long as $\rho \leq \rho_{\min}$. For example, the curve labeled "1" (corresponding to $\varepsilon_A = 0.05$) quits the x-axis for $\rho \geq \rho_{\min} = 0.06$.

The left plot of Fig. 4 corresponds to the RLS problem with $\rho = 1$ for various values of ε_A. This plot shows the various functions $f(\theta)$ as defined in (20). For each value of ε_A, the optimal θ (hence the RLS solution) is obtained by minimizing the function f. The three smallest values of ε_A induce functions f (as defined in (20)) that are minimal for $\theta < 1$. For the three others, the optimal θ is 1. This means that $\rho_{\min}$ is smaller than 1 in the first three cases and larger than 1 in the other cases. This is confirmed in Table 1.

7.5. Robust identification. Consider the following system identification problem. We seek to estimate the impulse response h of a discrete-time system from its input u and output y. Assuming that the system is single input and single output, linear, and of order m and that u is zero for negative time indices, y, u, and h are related by the convolution equations $Uh = y$, where

$$h = \begin{bmatrix} h(1) \\ \vdots \\ h(m) \end{bmatrix}, \quad y = \begin{bmatrix} y(1) \\ \vdots \\ y(m) \end{bmatrix}, \quad u = \begin{bmatrix} u(1) \\ \vdots \\ u(m) \end{bmatrix},$$

and U is a lower-triangular Toeplitz matrix whose first column is u. Assuming y, U are known exactly leads to a linear equation in h, which can be computed with standard

LS.

In practice, however, both y and u are subject to errors. We may assume, for instance, that the actual value of y is $y + \delta y$ and that of u is $u + \delta u$, where $\delta u, \delta y$ are unknown-but-bounded perturbations. For the perturbed matrices U, y write

$$\mathbf{U}(\delta) = U + \sum_{i=1}^{m} \delta u_i U_i, \quad \mathbf{y}(\delta) = y + \sum_{i=1}^{m} \delta y_i e_i,$$

where e_i, $i = 1, \ldots, m$ is the ith column of the $m \times m$ identity matrix and U_i are lower-triangular Toeplitz matrices with first column equal to e_i.

We first assume that the sum of the input and output energies is bounded, that is, $\|\delta\| \leq \rho$, where $\delta = [\delta u^T \ \delta y^T]^T \in \mathbf{R}^{2m}$, and $\rho \geq 0$ is given. We address the following SRLS problem:

$$(48) \qquad\qquad \min_{h \in \mathbf{R}^m} \max_{\|\delta\| \leq \rho} \|\mathbf{U}(\delta)h - \mathbf{y}(\delta)\|.$$

As an example, we consider the following nominal values for y, u:

$$u = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T, \quad y = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix}^T.$$

In Fig. 5, we have shown the optimal worst-case residual and that corresponding to the LS solution as given by solving problems (30) and (32), respectively. Since the LS solution has zero residual (U is invertible), we can prove (and check on the figure) that the worst-case residual grows linearly with ρ. In contrast, the RLS optimal worst-case residual has a finite limit as $\rho \to \infty$.
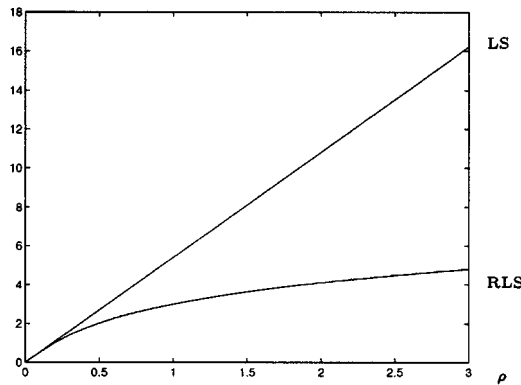

FIG. 5. *Worst-case residuals of LS and Euclidean-norm SRLS solutions for various values of perturbation level ρ. The worst-case residual for LS has been computed by solving problem (30) with $x = x_{\mathrm{LS}}$ fixed.*

We now assume that the perturbation bounds on y, u are not correlated. For instance, we consider problem (48), with the bound $\|\delta\| \leq \rho$ replaced with

$$\|\delta y\| \leq \rho, \quad \|\delta u\|_\infty \leq \rho.$$

Physically, the above bounds mean that the output energy and peak input are bounded.

This problem can be formulated as minimizing the worst-case residual (35), with

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix},$$

$$[A\ b] = \begin{bmatrix} 1 & 0 & 0 & 4 \\ 2 & 1 & 0 & 5 \\ 3 & 2 & 1 & 6 \end{bmatrix},$$

$$R^T = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

and Δ has the following structure:

$$\Delta = \mathbf{diag}\left(\delta u_1 I_3, \delta u_2 I_2, \delta u_3, \begin{bmatrix} \delta y_1 & \times & \times \\ \delta y_2 & \times & \times \\ \delta y_3 & \times & \times \end{bmatrix} \right).$$

Here, the symbols $\times$ denote dummy elements of Δ that were added in order to work with a square perturbation matrix. The above structure corresponds to the set $\mathcal{D}$ in (36), with $s = [3\ 2\ 1]$.
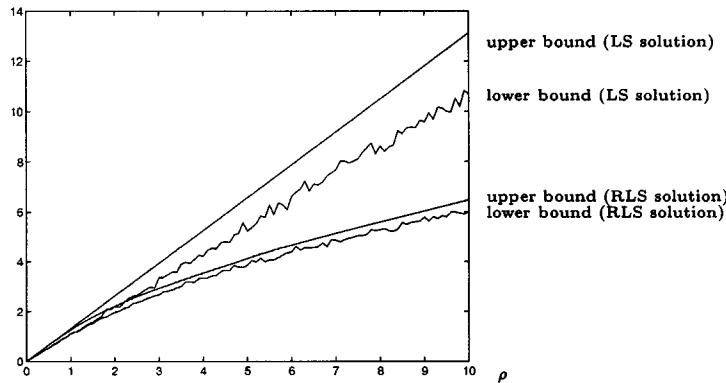

FIG. 6. *Upper and lower bounds on worst-case residuals for LS and RLS solutions. The upper bound for LS has been computed by solving the SDP* (38) *with* $x = x_{\mathrm{LS}}$ *fixed. The lower bounds correspond to the largest residuals* $\|\mathbf{U}(\delta^{\mathrm{trial}})x - \mathbf{y}(\delta^{\mathrm{trial}})\|$ *among* 100 *trial points* δ^{trial} *with* $x = x_{\mathrm{LS}}$ *and* $x = x_{\mathrm{RLS}}$.

In Fig. 6, we show the worst-case residual versus ρ, the uncertainty size. We show the curves corresponding to the values predicted by solving the SDP (43), with x variable (RLS solution), and x fixed to the LS solution x_{LS}. We also show lower bounds on the worst case, obtained using 100 trial points. This plot shows that, for the LS solution, our estimate of the worst-case residual is not exact, and the discrepancy grows linearly with uncertainty size. In contrast, for the RLS solution the estimate appears to be exact for every value of ρ.

7.6. Robust interpolation. The following example is a robust interpolation problem that can be formulated as a linear-fractional SRLS problem. For given integers $n \geq 1$, k, we seek a polynomial of degree $n - 1$, $p(t) = x_1 + \cdots + x_n t^{n-1}$ that interpolates given points (a_i, b_i), $i = 1, \ldots, k$; that is,

$$p(a_i) = b_i, \quad i = 1, \ldots, k.$$

If we assume that (a_i, b_i) are known exactly, we obtain a linear equation in the unknown x, with a Vandermonde structure

$$
\begin{bmatrix}
1 & a_1 & \cdots & a_1^{n-1} \\
\vdots & \vdots & & \vdots \\
1 & a_k & \cdots & a_k^{n-1}
\end{bmatrix}
\begin{bmatrix}
x_1 \\
\vdots \\
x_n
\end{bmatrix}
=
\begin{bmatrix}
b_1 \\
\vdots \\
b_n
\end{bmatrix},
$$

which can be solved via standard LS.

Now assume that the interpolation points are not known exactly. For instance, we may assume that the b_i's are known, while the a_i's are parameter dependent:

$$
a_i(\delta) = a_i + \delta_i, \quad i = 1, \ldots, k,
$$

where the δ_i's are unknown but bounded, $|\delta_i| \le \rho$, $i = 1, \ldots, k$, where $\rho \ge 0$ is given. We seek a robust interpolant, that is, a solution x that minimizes

$$
\max_{\|\delta\|_\infty \le \rho} \|\mathbf{A}(\delta)x - b\|,
$$

where

$$
\mathbf{A}(\delta) =
\begin{bmatrix}
1 & a_1(\delta) & \cdots & a_1(\delta)^{n-1} \\
\vdots & \vdots & & \vdots \\
1 & a_k(\delta) & \cdots & a_k(\delta)^{n-1}
\end{bmatrix}.
$$

The above problem is a linear-fractional SRLS problem. Indeed, it can be shown that

$$
\begin{bmatrix} \mathbf{A}(\delta) & b \end{bmatrix} = \begin{bmatrix} \mathbf{A}(0) & b \end{bmatrix} + L\Delta(I - D\Delta)^{-1} \begin{bmatrix} R_A & 0 \end{bmatrix},
$$

where

$$
L = \bigoplus_{i=1}^{k} \begin{bmatrix} 1 & a_i & \cdots & a_i^{n-2} \end{bmatrix}, R_A = \begin{bmatrix} R_1 \\ \vdots \\ R_k \end{bmatrix}, \quad D = \bigoplus_{i=1}^{k} D_i, \quad \Delta = \bigoplus_{i=1}^{k} \delta_i I_{n-1},
$$

and, for each i, $i = 1, \ldots, k$,

$$
R_i =
\begin{bmatrix}
0 & 1 & a_i & \cdots & a_i^{n-2} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & a_i \\
0 & \cdots & \cdots & 0 & 1
\end{bmatrix}
\in \mathbf{R}^{(n-1)\times n},
$$

$$
D_i =
\begin{bmatrix}
0 & 1 & a_i & \cdots & a_i^{n-3} \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
\vdots & & \ddots & \ddots & a_i \\
\vdots & & & \ddots & 1 \\
0 & \cdots & \cdots & \cdots & 0
\end{bmatrix}
\in \mathbf{R}^{(n-1)\times(n-1)}.
$$

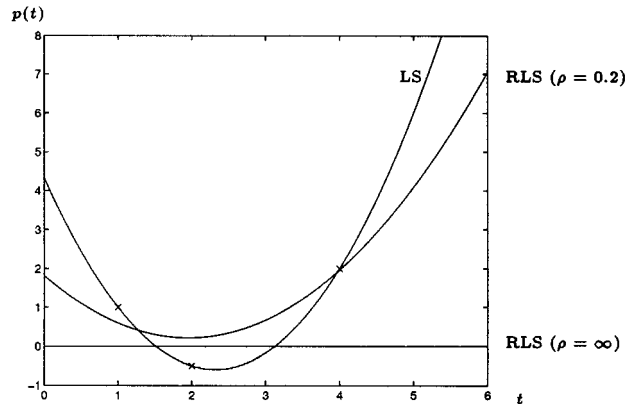(Note that $\det(I - D\Delta) \ne 0$, since D is strictly upper triangular.)

FIG. 7. *Interpolation polynomials: LS and RLS solutions for $\rho = 0.2$. The LS solution interpolates the points exactly, while the RLS one guarantees a worst-case residual error less than 1.1573. For $\rho = \infty$, the RLS solution is the zero polynomial.*

In Fig. 7, we have shown the result $n = 3$, $k = 1$, and

$$a_1 = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}, \quad b_1 = \begin{bmatrix} 1 \\ -0.5 \\ 2 \end{bmatrix}, \quad \rho = 0.2.$$

The LS solution is very accurate (zero nominal residual: every point is interpolated exactly) but has a (predicted) worst-case residual of 1.7977. The RLS solution trades off this accuracy (only one point interpolated and nominal residual of 0.8233) for robustness (with a worst-case residual less than 1.1573). As $\rho \to \infty$, the RLS interpolation polynomial becomes more and more horizontal. (This is consistent with the fact that we allow perturbations on vector a only.) In the limit, the interpolation polynomial is the solid line $p(t) = 0$.

8. Conclusions. This paper shows that several RLS problems with unknown-but-bounded data matrices are amenable to (convex) SOCP or SDP. The implication is that these RLS problems can be solved in polynomial time and efficiently in practice.

When the perturbation enters linearly in the data matrices, and its size is measured by Euclidean norm, or in a linear-fractional problem with full perturbation matrix Δ, the method yields the *exact* value of the optimal worst-case residual. In the other cases we have examined (such as arbitrary rational dependence of data matrices on the perturbation parameters), computing the worst-case residual is NP-complete. We have shown how to compute and optimize, using SDP, an upper bound on the worst-case residual that takes into account structure information.

In the unstructured case, we have shown that both the worst-case residual and the (unique) RLS solution are continuous. The unstructured RLS can be interpreted as a regularization method for ill-conditioned problems. A striking fact is that the cost of the RLS solution is equal to a small number of least-squares problems arising in classical Tikhonov regularization approaches. This method provides a rigorous way to compute the optimal parameter from the data and associated perturbation bounds. Similar (weighted) least-squares interpretations and continuity results were given for the structured case.

In our examples, we have demonstrated the use of an SOCP code [27] and a general-purpose semidefinite programming code **SP** [45]. Future work could be devoted to writing special code that exploits the structure of these problems in order to further increase the efficiency of the method. For instance, it seems that in many problems the perturbation matrices are sparse and/or have special (e.g., Toeplitz) structure.

The method can be used for several related problems.

- *Constrained RLS.* We may consider problems where additional (convex) constraints are added on the vector x. (Such constraints arise naturally in, e.g., image processing.) For instance, we may consider problem (1) with an additional linear (resp., quadratic convex) constraint $(Cx)_i \geq 0$, $i = 1, \ldots, q$ (resp., $x^T Q x \leq 1$), where C (resp., $Q \geq 0$) is given. To solve such a problem, it suffices to add the related constraint to corresponding SOCP or SDP formulation. (Note that the SVD approach of section 3.3 fails in this case.)
- *RLS problems with other norms.* We may consider RLS problems in which the worst-case residual errors measured in other norms such as the maximum (l_∞) norm.
- *Matrix RLS.* We may, of course, derive similar results when the constant term b is a matrix. The worst-case error can be evaluated in a variety of norms.
- *Error-in-variables RLS.* We may consider problems where the solution x is also subject to uncertainty (due to implementation and/or quantization errors). That is, we may consider a worst-case residual of the form

$$\max_{\|\Delta x\| \leq \rho_1} \max_{\|\Delta A\ \Delta b\|_F \leq \rho_2} \|(A + \Delta A)(x + \Delta x) - (b + \Delta b)\|,$$

where ρ_i, $i = 1, 2$, are given. We may compute (and optimize) upper bounds on the above quantity using SDP. This subject is examined in [25].

Appendix A. Proof of Theorem 4.1. Introduce the eigendecomposition of F and a related decomposition for g:

$$F = \tau I - U \begin{bmatrix} \tau - \lambda_{\max}(F) & 0 \\ 0 & \tau I - \Sigma \end{bmatrix} U^T, \ \ U^T g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix},$$

where $\tau > \|\Sigma\|$, $\Sigma \in \mathbf{R}^{r \times r}$, $\Sigma > 0$, and $g_2 \in \mathbf{R}^r$. When $\tau > \lambda_{\max}(F)$, inequality (29) writes

(A.49) $$\lambda \geq h + \tau + \frac{g_1^T g_1}{\tau - \lambda_{\max}(F)} + g_2^T (\tau I - \Sigma)^{-1} g_2.$$

If $\tau = \lambda_{\max}(F)$ at the optimum, then $g_1 = 0$, and there exists a nonzero vector u such that $(\tau I - F)u = 0$. From inequality (29), we conclude that $g^T u = 0$. In other words, $\lambda_{\max}(F)$ is not (F, g)-controllable, and u is an eigenvector that proves this uncontrollability. Using $g_1 = 0$ in (A.49), we obtain the optimal value of λ in this case:

$$\lambda = h + \tau + g_2^T (\tau I - \Sigma)^{-1} g_2.$$

Thus, the worst-case residual can be computed as claimed in the theorem.

For every pair (λ, τ) that is optimal for problem (29), we can compute a worst-case perturbation as follows. Define

$$\delta_0 = (\tau I - F)^\dagger g.$$

We have $\tau > \lambda_{\max}(F)$ at the optimum if and only if $\lambda_{\max}(F)$ is (F, g)-controllable (that is, $g_2 \neq 0$) or if $\lambda_{\max}(F)$ is not (F, g)-controllable and the function f defined in (31) satisfies

$$\frac{df}{d\tau}(\lambda_{\max}(F)) = 1 - g^T(\lambda_{\max}(F)I - F)^{2\dagger}g < 0.$$

In this case, the optimal τ satisfies

(A.50) $$1 = g^T(\tau I - F)^{-2}g;$$

that is, $\|\delta_0\| = 1$. Using this and (A.50), we obtain

$$\begin{bmatrix} 1 \\ \delta_0 \end{bmatrix}^T \begin{bmatrix} h & g^T \\ g & F \end{bmatrix} \begin{bmatrix} 1 \\ \delta_0 \end{bmatrix} = \lambda.$$

This proves that δ_0 is a worst-case perturbation.

If $\tau = \lambda_{\max}(F)$ at the optimum, then

$$\frac{df}{d\tau}(\lambda_{\max}(F)) = 1 - g^T(\lambda_{\max}(F)I - F)^{2\dagger}g \geq 0,$$

which implies that $\|\delta_0\| \leq 1$. Since $\tau = \lambda_{\max}(F)$, there exists a vector u such that $(\tau I - F)u = 0$, $g^T u = 0$. Without loss of generality, we may assume that the vector $\delta = \delta_0 + u$ satisfies $\|\delta\| = 1$. We have

$$\begin{bmatrix} 1 \\ \delta \end{bmatrix}^T \begin{bmatrix} h & g^T \\ g & F \end{bmatrix} \begin{bmatrix} 1 \\ \delta \end{bmatrix} = \tau\delta^T\delta - \delta^T(\tau I - F)\delta + 2\delta_0^T g + h$$
$$= h + \tau + g^T(\tau I - F)^\dagger g - 2u^T(\tau I - F)\delta_0 - u^T(\tau I - F)u = \lambda.$$

This proves that δ defined above is a worst-case perturbation.

In both cases seen above (τ equals $\lambda_{\max}(F)$ or not), a worst-case perturbation is any vector δ such that

$$(\tau I - F)\delta = g, \quad \|\delta\| = 1.$$

(We have just shown that the above equations always have a solution δ when τ is optimal.) This ends our proof. $\square$

Appendix B. Proof of Lemma 5.1. We use the following result, due to Nemirovsky [32].

LEMMA B.1. *Let $\Gamma(p, a)$ be a scalar function of positive integer p and p-dimensional vector a such that, first, Γ is well defined and takes rational values from $(0, \|a\|^{-2})$ for all positive integers p and all p-dimensional vectors a with $\|a\| \leq 0.1$ and, second, the value of this function at a given pair (p, a) can be computed in time polynomial in p and the length of the standard representation of the (rational) vector a. Then the problem $\mathcal{P}_\Gamma(p, a)$: given an integer $p \geq 0$ and $a \in \mathbf{R}^p$, $\|a\| \leq 0.1$, with rational positive entries, determine whether*

(B.51) $$p \leq \max_{\|\delta\|_\infty \leq 1} \delta^T(I - \Gamma(p, a)aa^T)\delta$$

is NP-complete. Besides this, either (B.51) *holds, or*

$$p - \frac{\Gamma(p, a)}{d(a)^2} \geq \max_{\|\delta\|_\infty \leq 1} \delta^T(I - aa^T)\delta,$$

where $d(a)$ is the smallest common denominator of the entries of a.

To prove our result, it suffices to show that for some appropriate function Γ satisfying the conditions of Lemma B.1, for any given p, a, we can reduce the problem $\mathcal{P}_\Gamma(p, a)$ to problem $\mathcal{P}(\mathbf{A}, \mathbf{b}, \mathcal{D}, x)$ in polynomial time. Set

$$\Gamma(p, a) = \frac{2a^T a + 1}{(a^T a + 1)^2}.$$

This function satisfies all requirements of Lemma B.1, so problem $\mathcal{P}_\Gamma(p, a)$ is NP-hard.

Given p, a, $\|a\| \leq 0.1$ with rational positive entries, set $\mathbf{A}$, $\mathbf{b}$, $\mathcal{D}$ and x as follows. First, set $\mathcal{D}$ to be the set of diagonal matrices of $\mathbf{R}^{p \times p}$. Set $A = 0$, $b = 0$, $R_A = 0$, $R_b = [1 \ldots 1]^T$, $D = 0$, $x = 0$, and

$$L = I - \frac{aa^T}{1 + a^T a}.$$

Finally, set $\mathbf{A}$, $\mathbf{b}$ as in (34) and $\lambda = p - \Gamma(p, a)/d(a)^2$. When $\rho = 1$, the worst-case residual for this problem is

$$r_\mathcal{D}(\mathbf{A}, \mathbf{b}, 1, x)^2 = \max_{\|\delta\|_\infty \leq 1} \|L\delta\|^2 = \max_{\|\delta\|_\infty \leq 1} \delta^T (I - \Gamma(p, a)aa^T)\delta.$$

Our proof is now complete.

Appendix C. Proof of Theorem 5.3. In this section, we only prove Theorem 5.3. The proof of Theorem 5.2 follows the same lines. We start from problem (43), the dual of which is the maximization of $2(b^T w + R_b^T u)$ subject to

(C.52)
$$\mathcal{Z} = \begin{bmatrix} Z & Y & w \\ Y^T & V & u \\ w^T & u^T & t \end{bmatrix} \geq 0$$

and the linear constraints

(C.53) $$\mathbf{Tr} Z = 1 - t,$$
(C.54) $$\forall S \in \mathcal{S}, \quad \mathbf{Tr} S(V - L^T ZL - D^T Y^T L - L^T YD - D^T VD) = 0,$$
(C.55) $$A^T w + R_A^T u = 0,$$
(C.56) $$\forall G \in \mathcal{G}, \quad \mathbf{Tr} G(YL - L^T Y^T - D^T V + VD) = 0.$$

Since both primal and dual problems are strictly feasible, all primal and dual feasible points are optimal if and only if $\mathcal{Z}\mathcal{F}(\lambda, S, G, x) = 0$, where $\mathcal{F}$ is defined in (38) (see [46]). One obtains, in particular,

(C.57) $$Jw + t(Ax - b) - L\Gamma u = 0,$$
(C.58) $$(Ax - b)^T w + t\lambda + z^T R^T u = 0,$$
(C.59) $$-\Gamma^T L^T w + Rz + \Sigma u = 0,$$

where $z = [x^T \ -1]^T$, $J = \lambda I - LSL^T$, $\Sigma = S + DG - GD^T - DSD^T$, and $\Gamma = SD^T - G$.

Using equation (C.58) and (C.55), we obtain

(C.60) $$t\lambda = -(Ax - b)^T w - z^T R^T u = b^T w + R_b^T u,$$

which implies that $t = 1/2$ from equality of the primal and dual objectives (the trivial case $\lambda = 0$ can be easily ruled out).

Assume that the matrix Θ defined in (39) is positive definite at the optimum. From equations (C.57)–(C.59), we deduce that the dual variable $\mathcal{Z}$ is rank one:

$$(C.61) \qquad \mathcal{Z} = 2vv^T \text{ with } v = \begin{bmatrix} w & u & 1/2 \end{bmatrix}^T.$$

Using (C.57) and (C.59), we obtain

$$\Theta \begin{bmatrix} w \\ u \end{bmatrix} = \frac{1}{2} \begin{bmatrix} Ax - b \\ R_A x - R_b \end{bmatrix}.$$

From (C.55), it is easy to derive the expression (42) for the optimal x in the case when $\Theta > 0$ at the optimum and R_A is full rank.

We now show that the upper bound is exact at the optimum in this case. If we use condition (C.54) and the expression for Z, V deduced from (C.53), we obtain

$$u^T S u = (L^T w + D^T u)^T S (L^T w + D^T u) \text{ for every } S \in \mathcal{S}.$$

This implies that there exists $\Delta \in \mathcal{D}$, $\Delta^T \Delta = I$, such that $u = \Delta^T (L^T w + D^T u)$. Since $\Theta > 0$, a straightforward application of Lemma 2.3 shows that $\det(I - D\Delta) \neq 0$, so we obtain

$$u^T = w^T L\Delta(I - D\Delta)^{-1}.$$

Define $M = [A \ b]$ and recall $z = [x^T \ -1]^T$. Since $Z = 2ww^T$ (from (C.61)) and $\mathbf{Tr}Z = 1 - t = 1/2$ (from (C.53)), we have $\|w\| = 1/2$. We can now compute

$$\begin{aligned} w^T(M + L\Delta(I - D\Delta)^{-1}R)z &= w^T(Ax - b) + w^T L\Delta(I - D\Delta)^{-1}Rz \\ &= w^T(Ax - b) + u^T Rz \\ &= -\frac{\lambda}{2} \text{ (from (C.55) and (C.60)).} \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{\lambda}{2} = \left| w^T(M + L\Delta(I - D\Delta)^{-1}R)z \right| &\leq \|w\| \left\| (M + L\Delta(I - D\Delta)^{-1}R)z \right\| \\ &\leq \|w\|\lambda \text{ (since } \Delta \in \mathcal{D}, \|\Delta\| \leq 1) \\ &= \frac{\lambda}{2} \left(\text{from} \|w\| = \frac{1}{2} \right). \end{aligned}$$

We obtain $\lambda = \left\| (M + L\Delta(I - D\Delta)^{-1}R)z \right\|$, which proves that the matrix Δ is a worst-case perturbation.

REFERENCES

[1] K. D. ANDERSEN, *An efficient Newton barrier method for minimizing a sum of Euclidean norms*, SIAM J. Optim., 6 (1996), pp. 74–95.

[2] A. BJÖRCK, *Component-wise perturbation analysis and error bounds for linear least squares solutions*, BIT, 31 (1991), pp. 238–244.

[3] J. F. BONNANS, R. COMINETTI, AND A. SHAPIRO, *Sensitivity analysis of optimization problems under abstract constraints*, SIAM J. Optim., submitted.

[4] S. BOYD, L. EL GHAOUI, E. FERON, AND V. BALAKRISHNAN, *Linear Matrix Inequalities in System and Control Theory*, in Studies in Applied Mathematics, SIAM, Philadelphia, PA, 1994.

[5] S. CHANDRASEKARAN, G. H. GOLUB, M. GU, AND A. H. SAYED, *A new linear least-squares type model for parameter estimation in the presence of data uncertainties*, SIAM J. Matrix Anal. Appl., submitted.

[6] G. E. COXSON AND C. L. DEMARCO, *Computing the Real Structured Singular Value is NP-Hard*, Tech. report ECE-92-4, Dept. of Elec. and Comp. Eng., University of Wisconsin-Madison, Madison, WI, June 1992.

[7] B. DE MOOR, *Structured total least squares and $\mathbf{L}_2$ approximation problems*, Linear Algebra Appl., 188–189 (1993), pp. 163–207.

[8] G. DEMOMENT, *Image reconstruction and restoration: Overview of common estimation problems*, IEEE Trans. Acoustic Speech and Signal Processing, 37 (1989), pp. 2024–2036.

[9] J. DOYLE, M. NEWLIN, F. PAGANINI, AND J. TIERNO, *Unifying robustness analysis and system ID*, in Proc. IEEE Conf. on Decision and Control, December 1994, pp. 3667–3672.

[10] L. EL GHAOUI, R. NIKOUKHAH, AND F. DELEBECQUE, `LMITOOL:` *A Front-End for LMI Optimization, User's Guide*, February 1995. Available via anonymous ftp from ftp.ensta.fr/pub/elghaoui/lmitool.

[11] L. ELDEN, *Algorithms for the regularization of ill conditioned least-squares problems*, BIT, 17 (1977), pp. 134–145.

[12] L. ELDEN, *Perturbation theory for the least-squares problem with linear equality constraints*, BIT, 24 (1985), pp. 472–476.

[13] M. K. H. FAN, A. L. TITS, AND J. C. DOYLE, *Robustness in the presence of mixed parametric uncertainty and unmodeled dynamics*, IEEE Trans. Automat. Control, 36 (1991), pp. 25–38.

[14] R. D. FIERRO AND J. R. BUNCH, *Collinearity and total least squares*, SIAM J. Matrix Anal. Appl., 15 (1994), pp. 1167–1181.

[15] M. FURUYA, H. OHMORI, AND A. SANO, *Optimization of weighting constant for regularization in least squares system identification*, Trans. Inst. Elec. Inform. Comm. Eng. A, J72A (1989), pp. 1012–1015.

[16] M. X. GOEMANS AND D. P. WILLIAMSON, *.878-approximation for MAX CUT and MAX 2SAT*, in Proc. 26th ACM Symp. Theor. Computing, 1994, pp. 422–431.

[17] G. H. GOLUB AND C. F. VAN LOAN, *An analysis of the total least squares problem*, SIAM J. Numer. Anal., 17 (1980), pp. 883–893.

[18] G. H. GOLUB AND C. F. VAN LOAN, *Matrix Computations*, 2nd ed., Johns Hopkins University Press, Baltimore, MD, 1989.

[19] G. H. GOLUB AND U. VON MATT, *Quadratically constrained least squares and quadratic problems*, Numer. Math., 59 (1991), pp. 561–580.

[20] M. L. HAMBABA, *The robust generalized least-squares estimator*, Signal Processing, 26 (1992), pp. 359–368.

[21] M. HANKE AND P. C. HANSEN, *Regularization methods for large-scale problems*, Surveys on Mathematics for Industry, 3 (1993), pp. 253–315.

[22] D. J. HIGHAM AND N. J. HIGHAM, *Backward error and condition of structured linear systems*, SIAM J. Matrix Anal. Appl., 13 (1992), pp. 162–175.

[23] B. R. HUNT, *The application of constrained least-squares estimation to image restoration by digital computer*, IEEE Trans Comput., C-22 (1973), pp. 805–812.

[24] T. IWASAKI AND R. E. SKELTON, *All controllers for the general $\mathbf{H}_\infty$ control problem: LMI existence conditions and state space formulas*, Automatica, 30 (1994), pp. 1307–1317.

[25] C. JACQUEMONT, *Error-in-Variables Robust Least-Squares*, Tech. rep., Ecole Nat. Sup. Techniques Avancies, 32, Bd. Victor, 75739 Paris, France, December 1995.

[26] G. LADY AND J. MAYBEE, *Qualitatively invertible matrices*, J. Math. Social Sciences, 6 (1983), pp. 397–407.

[27] H. LEBRET, *Synthèse de diagrammes de réseaux d'antennes par optimisation convexe*, Ph.D. thesis, UFR Structure et Propriétés de la Matière, mention Electronique, Université de

Rennes I, France, 1994.

[28] H. LEBRET, *Antenna pattern synthesis through convex optimization*, in Advanced Signal Processing Algorithms, Proc. SPIE 2563, F. T. Luk, ed., 1995, pp. 182–192.

[29] L. LEE AND A. TITS, *On continuity/discontinuity in robustness indicators*, IEEE Trans. Automat. Control, 38 (1993), pp. 1551–1553.

[30] K. MILLER, *Least squares methods for ill-posed problems with a prescribed bound*, SIAM J. Math. Anal., 1 (1970), pp. 52–74.

[31] M. Z. NASHED, *Operator-theoretic and computational approaches to ill-posed problems with applications to antenna theory*, IEEE Trans. Antennas and Propagation, 29 (1981), pp. 220–231.

[32] A. NEMIROVSKY, *Several NP-hard problems arising in robust stability analysis*, Mathematics of Control, Signals, and Systems, 6 (1993), pp. 99–105.

[33] Y. NESTEROV AND A. NEMIROVSKY, *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*, SIAM, Philadelphia, PA, 1994.

[34] J. NORTON, *Identification and application of bounded parameter models*, Automatica, 31 (1987), pp. 497–507.

[35] F. OUSTRY, L. EL GHAOUI, AND H. LEBRET, *Robust solutions to uncertain semidefinite programs*, SIAM J. Optim., submitted, 1996.

[36] C. PAPADIMITRIOU AND M. YANNAKAKIS, *Optimization, approximation and complexity classes*, J. Comput. System Sci., 43 (1991), pp. 425–440.

[37] J. R. PARTINGTON AND P. M. MÄKILÄ, *Worst-case analysis of the least-squares method and related identification methods*, Systems Control Lett., 24 (1995), pp. 193–200.

[38] S. POLJAK AND J. ROHN, *Checking robust nonsingularity is NP-hard*, Math. Control Signals Systems, 6 (1993), pp. 1–9.

[39] B. L. SHADER, *Least squares sign-solvability*, SIAM J. Matrix Anal. Appl., 16 (1995), pp. 1056–1073.

[40] R. SMITH AND J. DOYLE, *Model validation: A connection between robust control and identification*, IEEE Trans. Automat. Control, 37 (1992), pp. 942–952.

[41] R. J. STERN AND H. WOLKOWICZ, *Indefinite trust region subproblems and nonsymmetric eigenvalue perturbations*, SIAM J. Optim., 5 (1995), pp. 286–313.

[42] R. TEMPO, *Worst-case optimality of smoothing algorithms for parametric system identification*, Automatica, 31 (1995), pp. 759–764.

[43] A. TIKHONOV AND V. ARSENIN, *Solutions of Ill-Posed Problems*, Wiley, New York, 1977.

[44] S. VAN HUFFEL AND J. VANDEWALLE, *The total least squares problem: Computational aspects and analysis*, in Frontiers in Applied Mathetics 9, SIAM, Philadelphia, PA, 1991.

[45] L. VANDENBERGHE AND S. BOYD, **SP,** *Software for Semidefinite Programming, User's Guide*, Dec. 1994. Available via anonymous ftp from isl.stanford.edu under /pub/boyd/semidef_prog.

[46] L. VANDENBERGHE AND S. BOYD, *Semidefinite programming*, SIAM Rev., 38 (1996), pp. 49–95.

[47] M. E. ZERVAKIS AND T. M. KWON, *Robust estimation techniques in regularized image restoration*, Op. Eng., 31 (1992), pp. 2174–2190.

[48] K. ZHOU, J. DOYLE, AND K. GLOVER, *Robust and Optimal Control*, Prentice–Hall, Englewood Cliffs, NJ, 1995.