SIAM J. MATRIX ANAL. APPL.
Vol. 29, No. 2, pp. 606–625

REGULARIZATION IN REGRESSION WITH BOUNDED NOISE: A CHEBYSHEV CENTER APPROACH*

AMIR BECK† AND YONINA C. ELDAR‡

Abstract. We consider the problem of estimating a vector $\mathbf{z}$ in the regression model $\mathbf{b} = \mathbf{A}\mathbf{z}+\mathbf{w}$, where $\mathbf{w}$ is an unknown but bounded noise. As in many regularization schemes, we assume that an upper bound on the norm of $\mathbf{z}$ is available. To estimate $\mathbf{z}$ we propose a relaxation of the Chebyshev center, which is the vector that minimizes the worst-case estimation error over all feasible vectors $\mathbf{z}$. Relying on recent results regarding strong duality of nonconvex quadratic optimization problems with two quadratic constraints, we prove that in the *complex domain* our approach leads to the exact Chebyshev center. In the real domain, this strategy results in a "pretty good" approximation of the true Chebyshev center. As we show, our estimate can be viewed as a Tikhonov regularization with a special choice of parameter that can be found efficiently by solving a convex optimization problem with two variables or a semidefinite program with three variables, regardless of the problem size. When the norm constraint on $\mathbf{z}$ is a Euclidean one, the problem reduces to a single-variable convex minimization problem. We then demonstrate via numerical examples that our estimator can outperform other conventional methods, such as least-squares and regularized least-squares, with respect to the estimation error. Finally, we extend our methodology to other feasible parameter sets, showing that the total least-squares (TLS) and regularized TLS can be obtained as special cases of our general approach.

Key words. Chebyshev center, nonconvex quadratic optimization, strong duality, bounded error estimation

AMS subject classifications. 90C20, 90C22, 90C26, 65F30

DOI. 10.1137/060656784

1. Introduction.

Many problems in data fitting and estimation give rise to a system of linear equations $\mathbf{A}\mathbf{z} \approx \mathbf{b}$ where the right-hand side $\mathbf{b}$ is contaminated by noise. More specifically, we consider the linear model

$$\mathbf{b} = \mathbf{A}\mathbf{z} + \mathbf{w}, \tag{1}$$

where $\mathbf{A} \in \mathbb{F}^{m\times n}$ is the model matrix, $\mathbf{b} \in \mathbb{F}^m$ is the observation vector, $\mathbf{w} \in \mathbb{F}^m$ is the unknown noise (or "error"), and $\mathbf{z} \in \mathbb{F}^n$ is the unknown parameter vector. Here $\mathbb{F}$ denotes either the real number field $\mathbb{R}$ or the complex number field $\mathbb{C}$. Given the observation $\mathbf{b}$, we seek an estimator $\hat{\mathbf{z}}$ of $\mathbf{z}$ that is close in some sense to $\mathbf{z}$. This estimation problem arises in a large variety of areas in science and engineering, e.g., communication, economics, signal processing, seismology, and control.

The celebrated least-squares (LS) approach [5, 18] to estimating $\mathbf{z}$ in the model (1) is to seek the vector $\hat{\mathbf{z}}_{\mathrm{LS}}$ that minimizes the norm of the data error $\|\mathbf{A}\hat{\mathbf{z}} - \mathbf{b}\|^2$, where $\|\mathbf{v}\|$ stands for the Euclidean norm of the vector $\mathbf{v}$. When $\mathbf{A}$ has full column rank, $\hat{\mathbf{z}}_{\mathrm{LS}}$ is given by

$$\hat{\mathbf{z}}_{\mathrm{LS}} = (\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{b}. \tag{2}$$

*Received by the editors April 8, 2006; accepted for publication (in revised form) by J. G. Nagy December 7, 2006; published electronically May 1, 2007.
http://www.siam.org/journals/simax/29-2/65678.html

†Department of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa 32000, Israel (becka@ie.technion.ac.il).
‡Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (yonina@ee.technion.ac.il).

In practical situations the matrix $\mathbf{A}$ is often ill-conditioned—for example, when the system is obtained via discretization of ill-posed problems such as integral equations of the first kind (see, e.g., [16] and references therein). In these cases the LS solution might give poor results with respect to the estimation error. A well-established approach for stabilizing the LS estimate is to incorporate prior information on the true parameter vector $\mathbf{z}$ into the optimization problem (2) by adding a quadratic constraint:

$$\hat{\mathbf{z}}_{\mathrm{RLS}} \in \operatorname*{argmin}_{\mathbf{z}\in\mathbb{F}^n}\{\|\mathbf{A}\mathbf{z}-\mathbf{b}\|^2 : \|\mathbf{L}\mathbf{z}\|^2 \le \eta\}. \tag{3}$$

The matrix $\mathbf{L}$ is often chosen as the identity, or as a discrete approximation of some derivative operator (see [5, 16]). The resulting estimator is referred to as the *regularized* LS (RLS) estimator [5]. It is well known that $\hat{\mathbf{z}}_{\mathrm{RLS}}$ is either equal to the LS solution when $\|\mathbf{L}\mathbf{z}_{\mathrm{LS}}\|^2 \le \eta$ or given by $\hat{\mathbf{z}}_{\mathrm{RLS}} = \mathbf{z}_\lambda$, where $\mathbf{z}_\lambda$ satisfies the generalized normal equations [5]

$$(\mathbf{A}^*\mathbf{A} + \lambda\mathbf{L}^*\mathbf{L})\mathbf{z}_\lambda = \mathbf{A}^*\mathbf{b}. \tag{4}$$

The parameter λ is determined by the secular equation $\|\mathbf{L}\mathbf{z}_\lambda\|^2 = \eta$. Therefore, the RLS solution is a *Tikhonov* estimator [28] with a choice of regularization parameter λ that takes into account the norm constraint $\|\mathbf{L}\mathbf{z}\|^2 \le \eta$.

It is important to note that both the LS and the RLS strategies are based on minimizing the data error. However, in an estimation context, typically we would like to minimize the squared *estimation error* $\|\hat{\mathbf{z}} - \mathbf{z}\|^2$. When the noise $\mathbf{w}$ in (1) is assumed to be random with zero mean and known covariance matrix, the squared estimation error will also be a random variable. Using the known statistics of $\mathbf{w}$, the average squared estimation error, referred to as the mean-squared error (MSE), can be computed. Several different strategies based on the MSE have been recently proposed [25, 9, 8, 2, 7]. These methods consider *linear* estimates of $\mathbf{z}$ and assume knowledge of the statistics of $\mathbf{w}$.

1.1. Bounded error estimation. In some scenarios, the distribution of the noise might not be known exactly (or at all). There are also cases where the noise is not inherently random (for example, in problems resulting from quantizing a continuous-time signal). This leads to the *bounded error estimation* approach which deals with unknown but bounded noise (see, e.g., [19] and the survey papers [21, 24]). In this paper we adopt the bounded error methodology and assume that the noise is norm-bounded $\|\mathbf{w}\|^2 \le \rho$. As in the RLS strategy, in order to obtain a stable solution, we further restrict $\mathbf{z}$ to have weighted bounded norm.

The first stage in the deterministic bounded error approach is to construct all admissible solutions to the linear system (1); for this reason this approach is also referred to as *set-membership estimation* [21]. In our setting, the feasible parameter set (FPS) is given by the intersection of two ellipsoids[1]:

$$\mathrm{FPS} = \{\mathbf{z} \in \mathbb{F}^n : \|\mathbf{L}\mathbf{z}\|^2 \le \eta, \|\mathbf{A}\mathbf{z}-\mathbf{b}\|^2 \le \rho\}. \tag{5}$$

The second step is to choose a central representative of the FPS. A popular choice is the Chebyshev center [29], which is defined as the solution $\hat{\mathbf{z}}$ to the following min-max problem:

$$\min_{\hat{\mathbf{z}}\in\mathbb{F}^n} \max_{\mathbf{z}\in\mathrm{FPS}} \|\mathbf{z}-\hat{\mathbf{z}}\|^2. \tag{6}$$

[1]Note that here the norm bound $\|\mathbf{w}\|^2 \le \rho$ translates to $\|\mathbf{A}\mathbf{z}-\mathbf{b}\|^2 \le \rho$.

Geometrically, the Chebyshev center is the center of the minimum radius ball enclosing the FPS; the optimal value of (6) is the squared radius of the minimal ball enclosing the set. This is illustrated in Figure 1 with the filled area being the intersection of two ellipsoids. The dotted circle is the minimum inscribing circle of the intersection of the ellipsoids.

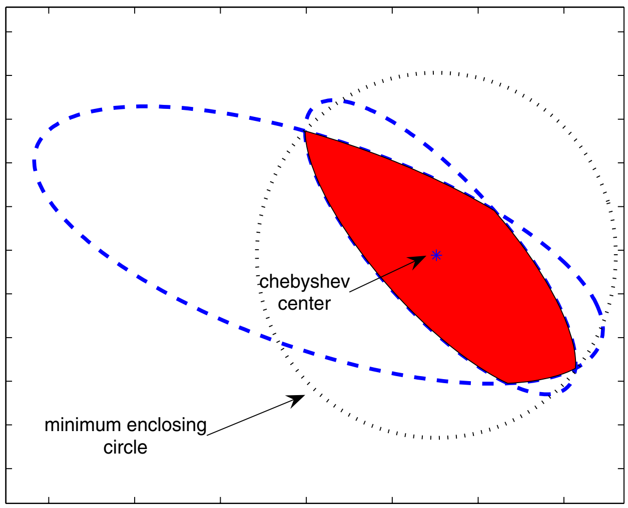

FIG. 1. *The Chebyshev center of the intersection of two ellipsoids.*

The Chebyshev center of the FPS gives the best worst-case estimation error over the set. Thus, it is aimed at optimizing an objective that depends on the estimation error rather than the data error. In section 5 we demonstrate by simulations that an estimator based on the Chebyshev center typically performs worse than the LS and RLS approaches with respect to the data error; however, it appears to perform *significantly* better in terms of the estimation error even when only loose bounds on the norm of the noise (ρ) are known. Thus, this approach can improve the estimation error without requiring much more knowledge than the RLS strategy.

Finding a Chebyshev center of a convex set is, in general, a hard problem. Two exceptions are the case where the set is polyhedral and the enclosing ball is the l_∞ ball [20], and the case when the set is finite (see, e.g., [30] and references therein).

The Chebyshev center problem (6) we tackle in this paper is seemingly hard. To better understand the intrinsic difficulty of this min-max problem, note that the inner maximization problem is a *nonconvex* quadratic optimization problem. However, relying on some recent strong duality results derived in the context of quadratic optimization [1], we will show that despite the nonconvexity of the problem, it can be solved efficiently when $\mathbb{F} = \mathbb{C}$. The same approach can be used when $\mathbb{F} = \mathbb{R}$ to develop an approximation of the Chebyshev center. Simulation results show that this approximation is pretty good in the sense that it yields favorable estimation performance.

1.2. Paper layout and main results. A review of the relevant optimization concepts and the strong duality results of [1] is given in section 2. These results are then used in section 3 to reduce the problem of finding the Chebyshev center of the intersection of two level sets of quadratic functions[2] with $\mathbb{F} = \mathbb{C}$ to a convex optimization problem in only two variables. This problem can also be recast as a semidefinite program (SDP) involving linear matrix inequality (LMI) constraints, with three variables.

In section 4 we present the relaxed Chebyshev center (RCC) estimator, which is exactly the Chebyshev center of the FPS in the case $\mathbb{F} = \mathbb{C}$ under strict feasibility constraints. We show that the RCC, like the RLS solution, is a Tikhonov estimator. However, in the RCC approach, as opposed to the RLS method, the regularization parameter is chosen to account for *both* constraints defining the FPS. Furthermore, it is designed to minimize an estimation error rather than a data error. We also show that when considering the FPS with a Euclidean norm constraint on $\mathbf{z}$ (i.e., $\mathbf{L} = \mathbf{I}$), the problem reduces to a convex optimization problem with a single variable.

Section 5 presents numerical examples demonstrating the effectiveness of the RCC strategy. We also compare two methods for evaluating the RCC estimator: an implementation of the ellipsoid method [3] (described in full detail in Appendix A) and a standard interior point method applied to the resulting SDP. We show both theoretically and numerically that in our problem the ellipsoid method is more efficient.

Finally, in section 6, we extend our approach to several related problems, and show that the total LS (TLS) [13, 17] and regularized TLS (RTLS) estimators [12] can be viewed as special cases of our general methodology.

1.3. Notation. Throughout the paper, the following notation is used: vectors are denoted by boldface lowercase letters, e.g., $\mathbf{y}$, and matrices by boldface uppercase letters, e.g., $\mathbf{A}$. The ith component of a vector $\mathbf{y}$ is written as y_i, and $\hat{(\cdot)}$ is an estimated vector. The identity matrix is denoted by $\mathbf{I}$. The real and imaginary parts of scalars, vectors, or matrices are written as $\Re(\cdot)$ and $\Im(\cdot)$. For a matrix $\mathbf{A}$, $\mathbf{A}^*, \mathbf{A}^T, \mathbf{A}^\dagger$, and $\mathcal{R}(\mathbf{A})$ are the Hermitian conjugate, transpose, Moore-Penrose generalized inverse [14], and image space. For a square symmetric matrix, $\lambda_{\min}(\mathbf{A})$ is the minimum eigenvalue of $\mathbf{A}$. Given two matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \succ \mathbf{B}$ ($\mathbf{A} \succeq \mathbf{B}$) means that $\mathbf{A} - \mathbf{B}$ is positive definite (semidefinite). The value of the optimal objective function of an optimization problem

$$(P):\ \min/\max\{f(\mathbf{x}) : \mathbf{x} \in C\}$$

is denoted by $\mathrm{val}(P)$. For simplicity, instead of inf/sup we use min/max; however this does not mean that we assume that the optimum is attained and/or finite.

2. Quadratically constrained quadratic programs: A review. Our goal is to find the Chebyshev center of the FPS (5). The difficulty is that the inner maximization in (6)

$$\max_{\mathbf{z} \in \mathrm{FPS}} \|\mathbf{z} - \hat{\mathbf{z}}\|^2$$

is not convex. In this section, we summarize prior results concerning the minimization of a general quadratic form subject to quadratic constraints. We will then show, in sections 3 and 4, how these results can be applied in order to solve (6).

[2]This is a more general form than a set that is an intersection of two ellipsoids.

Consider the general form quadratically constrained quadratic problem

$$(\mathrm{QP}_m) \quad \min_{\mathbf{z}\in\mathbb{F}^n}\{f_0(\mathbf{z}) : f_i(\mathbf{z}) \le 0, i = 1,\dots,m\},$$

where m denotes the number of constraints, and

$$f_i(\mathbf{z}) = \mathbf{z}^*\mathbf{A}_i\mathbf{z} + 2\Re(\mathbf{b}_i^*\mathbf{z}) + c_i$$

with $\mathbf{A}_i = \mathbf{A}_i^* \in \mathbb{F}^{n\times n}, \mathbf{b}_i \in \mathbb{F}^n$, and $c_i \in \mathbb{R}$ for $i = 0,\dots,m$. Note that in the case $\mathbb{F} = \mathbb{R}$, the quadratic functions $f_i(\mathbf{z})$ can be written as $\mathbf{z}^T\mathbf{A}_i\mathbf{z} + 2\mathbf{b}_i^T\mathbf{z} + c_i$.

The problem (QP_m) is in general not convex since $\mathbf{A}_i$ are not necessarily positive semidefinite. The Lagrangian dual of (QP_m) is the maximization problem [4, 26]

$$\max_{\boldsymbol{\alpha}}\{q(\boldsymbol{\alpha}) : \boldsymbol{\alpha} \ge 0\}, \tag{7}$$

where $q(\boldsymbol{\alpha})$ is the dual objective function defined by

$$q(\boldsymbol{\alpha}) = \min_{\mathbf{z}\in\mathbb{F}^n}\left\{f_0(\mathbf{z}) + \sum_{i=1}^m \alpha_i f_i(\mathbf{z})\right\}.$$

The function $q(\boldsymbol{\alpha})$ can also be written in the form

$$q(\boldsymbol{\alpha}) = \max_{\lambda}\left\{\lambda : f_0(\mathbf{z}) + \sum_{i=1}^m \alpha_i f_i(\mathbf{z}) \ge \lambda \text{ for every } \mathbf{z} \in \mathbb{F}^n\right\}. \tag{8}$$

To obtain a more convenient representation of $q(\boldsymbol{\alpha})$ we exploit the following well-known lemma.

LEMMA 2.1 (see [3, p. 163]). *Let $g : \mathbb{F}^n \to \mathbb{R}$ be given by $g(\mathbf{z}) = \mathbf{z}^*\mathbf{A}\mathbf{z} + 2\Re(\mathbf{b}^*\mathbf{z}) + c$, where $\mathbf{A} = \mathbf{A}^* \in \mathbb{F}^{n\times n}, \mathbf{b} \in \mathbb{F}^n$, and $c \in \mathbb{R}$. Then the two statements below are equivalent:*

(i) *$g(\mathbf{z}) \ge 0$ for every $\mathbf{z} \in \mathbb{F}^n$.*

(ii) $\begin{pmatrix} \mathbf{A} & \mathbf{b} \\ \mathbf{b}^* & c \end{pmatrix} \succeq \mathbf{0}$.

Applying Lemma 2.1 to (8), we can represent $q(\boldsymbol{\alpha})$ as

$$q(\boldsymbol{\alpha}) = \max_{\lambda}\left\{\lambda : \begin{pmatrix} \mathbf{A}_0 & \mathbf{b}_0 \\ \mathbf{b}_0^* & c_0 - \lambda \end{pmatrix} + \sum_{i=1}^m \alpha_i \begin{pmatrix} \mathbf{A}_i & \mathbf{b}_i \\ \mathbf{b}_i^* & c_i \end{pmatrix} \succeq \mathbf{0}\right\}.$$

The dual problem (7) then becomes

$$(\mathrm{D}_m) \quad \max_{\alpha_i \ge 0,\lambda}\left\{\lambda : \begin{pmatrix} \mathbf{A}_0 & \mathbf{b}_0 \\ \mathbf{b}_0^* & c_0 - \lambda \end{pmatrix} + \sum_{i=1}^m \alpha_i \begin{pmatrix} \mathbf{A}_i & \mathbf{b}_i \\ \mathbf{b}_i^* & c_i \end{pmatrix} \succeq \mathbf{0}\right\}.$$

Note that (D_m), also called Shor's relaxation, is an SDP [3], i.e., a problem involving the minimization of a linear function subject to LMIs.

The *weak duality theorem* [4] states that one always has $\mathrm{val}(\mathrm{D}_m) \le \mathrm{val}(\mathrm{QP}_m)$. A fundamental question is whether or not there is *strong duality*, i.e., is $\mathrm{val}(\mathrm{QP}_m) = \mathrm{val}(\mathrm{D}_m)$? When all the functions $f_i, i = 0,\dots,m$, are convex and strict feasibility holds, the answer is affirmative (this follows from the well-known strong duality theorem for convex programming [26]). However, if even one of the functions is not convex, then strong duality can be violated. Two exceptions are (i) the case of a single quadratic constraint ($m = 1$) (see, e.g., [10, 22]) and (ii) the case of two quadratic constraints ($m = 2$) in which the underlying number field is complex (strong duality is *not* guaranteed when $\mathbb{F} = \mathbb{R}$). The latter result was recently derived in [1] and is recalled in Theorem 2.1.

THEOREM 2.1 (see [1]). *Suppose that* $\mathbb{F} = \mathbb{C}$ *and that problem* (QP_2) *is strictly feasible,* i.e., *there exists* $\tilde{\mathbf{z}} \in \mathbb{F}^n$ *such that* $f_1(\tilde{\mathbf{z}}) < 0, f_2(\tilde{\mathbf{z}}) < 0$. *Further assume that*

$$\exists \gamma_1 \geq 0, \gamma_2 \geq 0 : \gamma_1 \mathbf{A}_1 + \gamma_2 \mathbf{A}_2 \succ \mathbf{0}. \tag{9}$$

Then the minimum and maximum of problems (QP_2) *and* (D_2)*, respectively, are attained and* $val(QP_2) = val(D_2)$.

3. The two-quadratic Chebyshev center. We now apply the results of the previous section to the problem of finding the Chebyshev center of the intersection of two level sets of quadratic functions. Specifically, we show that if the underlying number field is complex ($\mathbb{F} = \mathbb{C}$), then the Chebyshev center can be found by solving a convex optimization problem with two variables, or an SDP with three variables, thus rendering the problem tractable. In the case $\mathbb{F} = \mathbb{R}$ the proposed methodology results in an approximation of the exact Chebyshev center.

Consider the set Ω given as the intersection of level sets of two quadratic functions:

$$\Omega = \{\mathbf{z} \in \mathbb{F}^n : f_i(\mathbf{z}) \leq 0, i = 1, 2\}, \tag{10}$$

where $f_i(\mathbf{z}) = \mathbf{z}^* \mathbf{A}_i \mathbf{z} + 2\Re(\mathbf{b}_i^* \mathbf{z}) + c_i$ with $\mathbf{A}_i = \mathbf{A}_i^* \in \mathbb{F}^{n \times n}, \mathbf{b}_i \in \mathbb{F}^n$, and $c_i \in \mathbb{R}$ for $i = 1, 2$. We assume that condition (9) holds true. This is the case, for example, when at least one of the functions is strictly convex, which is equivalent to saying that the corresponding level set is a nondegenerate ellipsoid.

The Chebyshev center of Ω is the vector $\hat{\mathbf{z}} \in \mathbb{F}^n$ which is the solution to

$$\min_{\hat{\mathbf{z}} \in \mathbb{F}^n} \max_{\mathbf{z} \in \Omega} \|\mathbf{z} - \hat{\mathbf{z}}\|^2. \tag{11}$$

Theorem 3.1 below shows that finding the Chebyshev center of Ω can be recast as a convex optimization problem with only two variables. In order to prove the theorem, we will require the following lemma on Schur complements of singular matrices.

LEMMA 3.1 (see [6, Appendix A.5]). *Let*

$$\mathbf{X} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{C} \end{pmatrix},$$

where $\mathbf{A} = \mathbf{A}^* \in \mathbb{F}^{k \times k}, \mathbf{B} \in \mathbb{F}^{k \times p}$, *and* $\mathbf{C} = \mathbf{C}^* \in \mathbb{F}^{p \times p}$. *Then* $\mathbf{X} \succeq \mathbf{0}$ *if and only if*

$$\mathbf{A} \succeq \mathbf{0}, \quad \mathbf{C} - \mathbf{B}^* \mathbf{A}^\dagger \mathbf{B} \succeq \mathbf{0}, \quad (\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{B} = \mathbf{0}.$$

Remark 3.1. Note that the condition $(\mathbf{I} - \mathbf{A}\mathbf{A}^\dagger)\mathbf{B} = \mathbf{0}$ is equivalent to saying that $\mathbf{A}\mathbf{Y} = \mathbf{B}$ for some $\mathbf{Y} \in \mathbb{F}^{p \times k}$.

THEOREM 3.1. *Let* Ω *be the set given in* (10) *with* $\mathbb{F} = \mathbb{C}$. *Suppose that there exists* $\tilde{\mathbf{z}} \in \mathbb{F}^n$ *such that* $f_1(\tilde{\mathbf{z}}) < 0$ *and* $f_2(\tilde{\mathbf{z}}) < 0$ *and that* (9) *is satisfied. Then the solution to* (11) *is*

$$\hat{\mathbf{z}} = -\left(\alpha_1 \mathbf{A}_1 + \alpha_2 \mathbf{A}_2\right)^{-1} \left(\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2\right), \tag{12}$$

where (α_1, α_2) is an optimal solution of the following convex optimization problem in two variables:

$$
\begin{array}{ll}
\min_{\alpha_1,\alpha_2} & \left\{-c_1\alpha_1 - c_2\alpha_2 + (\alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2)^*(\alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2)^{-1}(\alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2)\right\} \\
\textit{s.t.} & \alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2 \succeq \mathbf{I}, \alpha_1 \geq 0, \alpha_2 \geq 0.
\end{array} \tag{13}
$$

Proof. Problem (11) can be rewritten as

$$\min_{\hat{\mathbf{z}}\in\mathbb{F}^n} \left\{ \|\hat{\mathbf{z}}\|^2 + \max_{\mathbf{z}\in\Omega} \left\{ \|\mathbf{z}\|^2 - 2\mathbf{z}^*\hat{\mathbf{z}} \right\} \right\}.$$

By using the strong duality result of Theorem 2.1 (note that all the conditions are satisfied), we conclude that the value of the inner maximization

$$\max_{\mathbf{z}\in\Omega}\{\|\mathbf{z}\|^2 - 2\mathbf{z}^*\hat{\mathbf{z}}\}$$

is equal to the value of the dual minimization problem (see section 2):

$$
\begin{array}{ll}
\min_{\alpha_1,\alpha_2,\lambda} & \lambda \\
\text{s.t.} & \begin{pmatrix} -\mathbf{I} & \hat{\mathbf{z}} \\ \hat{\mathbf{z}}^* & \lambda \end{pmatrix} + \alpha_1 \begin{pmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^* & c_1 \end{pmatrix} + \alpha_2 \begin{pmatrix} \mathbf{A}_2 & \mathbf{b}_2 \\ \mathbf{b}_2^* & c_2 \end{pmatrix} \succeq \mathbf{0}, \\
& \alpha_1 \geq 0, \alpha_2 \geq 0.
\end{array}
$$

Therefore, we can write (11) as

$$
\begin{array}{ll}
\min_{\alpha_1,\alpha_2,\hat{\mathbf{z}},\lambda} & \left\{\lambda + \|\hat{\mathbf{z}}\|^2\right\} \\
\text{s.t.} & \begin{pmatrix} -\mathbf{I} + \alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2 & \hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 \\ (\hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2)^* & \lambda + \alpha_1 c_1 + \alpha_2 c_2 \end{pmatrix} \succeq \mathbf{0}, \\
& \alpha_1 \geq 0, \alpha_2 \geq 0.
\end{array} \tag{14}
$$

Using Lemma 3.1 and Remark 3.1, problem (14) is equivalent to

$$
\begin{array}{ll}
\min_{\alpha_1,\alpha_2,\hat{\mathbf{z}},\lambda} & \left\{\lambda + \|\hat{\mathbf{z}}\|^2\right\} \\
\text{s.t.} & \mathcal{B}_\alpha \succeq \mathbf{0}, \\
& \hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 \in \mathcal{R}(\mathcal{B}_\alpha), \\
& \lambda + \alpha_1 c_1 + \alpha_2 c_2 \geq (\hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2)\mathcal{B}_\alpha^\dagger(\hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2), \\
& \alpha_1 \geq 0, \alpha_2 \geq 0,
\end{array} \tag{15}
$$

where we defined

$$\mathcal{B}_\alpha \equiv -\mathbf{I} + \alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2.$$

Noting that at the optimum we will have equality in the third constraint of (15), our problem reduces to

$$
\begin{array}{ll}
\min_{\alpha_1,\alpha_2,\hat{\mathbf{z}}} & \left\{-\alpha_1 c_1 - \alpha_2 c_2 + (\hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2)\mathcal{B}_\alpha^\dagger(\hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2) + \|\hat{\mathbf{z}}\|^2\right\} \\
\text{s.t.} & \mathcal{B}_\alpha \succeq \mathbf{0}, \\
& \hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 \in \mathcal{R}(\mathcal{B}_\alpha), \\
& \alpha_1 \geq 0, \alpha_2 \geq 0.
\end{array} \tag{16}
$$

The constraint $\hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 \in \mathcal{R}(\mathcal{B}_\alpha)$ is satisfied if and only if there exists $\mathbf{w} \in \mathbb{F}^n$ such that $\hat{\mathbf{z}} + \alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2 = \mathcal{B}_\alpha\mathbf{w}$. Using this observation combined with the identity $\mathcal{B}_\alpha\mathcal{B}_\alpha^\dagger\mathcal{B}_\alpha = \mathcal{B}_\alpha$, (16) becomes

$$
\begin{array}{ll}
\min_{\alpha_1,\alpha_2,\mathbf{w}} & \left\{-\alpha_1 c_1 - \alpha_2 c_2 + \mathbf{w}^*\mathcal{B}_\alpha\mathbf{w} + \|-\alpha_1\mathbf{b}_1 - \alpha_2\mathbf{b}_2 + \mathcal{B}_\alpha\mathbf{w}\|^2\right\} \\
\text{s.t.} & \mathcal{B}_\alpha \succeq \mathbf{0}, \\
& \alpha_1 \geq 0, \alpha_2 \geq 0.
\end{array} \tag{17}
$$

Fixing (α_1, α_2) and minimizing with respect to $\mathbf{w}$, we obtain that an optimal $\mathbf{w}$ is any vector satisfying

$$\mathcal{B}_\alpha(\mathbf{I} + \mathcal{B}_\alpha)\mathbf{w} = \mathcal{B}_\alpha(\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2).$$

Choosing $\mathbf{w} = (\mathbf{I} + \mathcal{B}_\alpha)^{-1}(\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2)$ together with the identity

$$\mathcal{B}_\alpha(\mathbf{I} + \mathcal{B}_\alpha)^{-1} = \mathbf{I} - (\mathbf{I} + \mathcal{B}_\alpha)^{-1}$$

leads to the following form of (11):

$$\begin{array}{ll} \min_{\alpha_1,\alpha_2} & \left\{-\alpha_1 c_1 - \alpha_2 c_2 + (\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2)^*(\alpha_1 \mathbf{A}_1 + \alpha_2 \mathbf{A}_2)^{-1}(\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2)\right\} \\ \text{s.t.} & \alpha_1 \mathbf{A}_1 + \alpha_2 \mathbf{A}_2 \succeq \mathbf{I}, \\ & \alpha_1 \geq 0, \alpha_2 \geq 0. \end{array} \tag{18}$$

Since the objective in (18) is convex and the constraints are convex conic constraints, the problem (18) is convex. Finally,

$$\begin{aligned} \hat{\mathbf{z}} &= -\alpha_1 \mathbf{b}_1 - \alpha_2 \mathbf{b}_2 + \mathcal{B}_\alpha \mathbf{w} \\ &= (-\mathbf{I} + \mathcal{B}_\alpha(\mathbf{I} + \mathcal{B}_\alpha)^{-1})(\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2) \\ &= -(\mathbf{I} + \mathcal{B}_\alpha)^{-1}(\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2), \end{aligned}$$

completing the proof. ◻

An immediate consequence of Theorem 3.1 is that at the expense of adding an additional variable, we can recast the problem of finding the Chebyshev center of Ω as an SDP with three variables.

COROLLARY 3.2. *Consider the setting of Theorem* 3.1. *Then the solution to* (11) *is given by*

$$\hat{\mathbf{z}} = -\left(\alpha_1 \mathbf{A}_1 + \alpha_2 \mathbf{A}_2\right)^{-1}\left(\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2\right), \tag{19}$$

where (α_1, α_2) *is an optimal solution of the SDP:*

$$\begin{array}{ll} \min_{\alpha_1,\alpha_2,t} & \{-\alpha_1 c_1 - \alpha_2 c_2 + t\} \\ \text{s.t.} & \alpha_1 \mathbf{A}_1 + \alpha_2 \mathbf{A}_2 \succeq \mathbf{I}, \\ & \begin{pmatrix} \alpha_1 \mathbf{A}_1 + \alpha_2 \mathbf{A}_2 & \alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2 \\ (\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2)^* & t \end{pmatrix} \succeq \mathbf{0}, \\ & \alpha_1 \geq 0, \alpha_2 \geq 0. \end{array} \tag{20}$$

Proof. The proof follows from rewriting (13) as

$$\begin{array}{ll} \min_{\alpha_1,\alpha_2,t} & \{-\alpha_1 c_1 - \alpha_2 c_2 + t\} \\ \text{s.t.} & \alpha_1 \mathbf{A}_1 + \alpha_2 \mathbf{A}_2 \succeq \mathbf{I}, \\ & (\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2)^*(\alpha_1 \mathbf{A}_1 + \alpha_2 \mathbf{A}_2)^{-1}(\alpha_1 \mathbf{b}_1 + \alpha_2 \mathbf{b}_2) \leq t, \\ & \alpha_1 \geq 0, \alpha_2 \geq 0, \end{array}$$

and invoking Lemma 3.1. ◻

Since problem (20) is an SDP, it can be solved efficiently via interior point methods [23]. Alternatively, we may solve the convex optimization problem (13) using the ellipsoid method [3], which is attractive given the small number of variables (two). In section 5 we compare these approaches.

The Chebyshev center of Ω can be calculated using Theorem 3.1 only when $\mathbb{F} = \mathbb{C}$. In the real case ($\mathbb{F} = \mathbb{R}$), strong duality is not guaranteed, and therefore the vector $\hat{\mathbf{z}}$

defined by (12), with (α_1, α_2) being the optimal solution of (13) (or of (20)), is not necessarily the exact Chebyshev center. In fact, the weak duality theorem implies that the resulting ball will enclose the set but will not necessarily be the smallest one possible. In Figure 2, four examples of intersections of ellipsoids in the real domain are given. The vector $\hat{\mathbf{z}}$ was calculated by solving the SDP problem (20) with the software package SeDuMi [27]. The radius of each ball is the square root of the corresponding optimal value of problem (20). In the two upper examples it seems that strong duality holds while in the two lower examples it is evident that the circle defined by Theorem 3.1 (or Corollary 3.2) is not minimal.

An important property of an *optimal* solution $(\bar{\alpha}_1, \bar{\alpha}_2)$ of problem (13) is that the matrix $\bar{\alpha}_1\mathbf{A}_1 + \bar{\alpha}_2\mathbf{A}_2 - \mathbf{I}$ is *not* positive definite, i.e., the minimum eigenvalue of $\bar{\alpha}_1\mathbf{A}_1 + \bar{\alpha_2}\mathbf{A}_2 - \mathbf{I}$ is zero. This is proved in Theorem 3.3 below. This result is valid both in the complex and real domains. In section 4.2 we use this result in order to further reduce (13) to a *single-variable* convex optimization problem when $\mathbf{L} = \mathbf{I}$.

THEOREM 3.3. *Suppose that there exists* $\tilde{\mathbf{z}} \in \mathbb{F}^n$ *such that* $f_1(\tilde{\mathbf{z}}) < 0$ *and* $f_2(\tilde{\mathbf{z}}) < 0$ *and that* (9) *is satisfied. Let* $(\bar{\alpha}_1, \bar{\alpha}_2)$ *be an optimal solution of* (13). *Then* $\lambda_{\min}(\bar{\alpha}_1\mathbf{A}_1 + \bar{\alpha}_2\mathbf{A}_2 - \mathbf{I}) = 0$.

Proof. Denote the objective function in (13) by

$$h(\boldsymbol{\alpha}) = -c_1\alpha_1 - c_2\alpha_2 + (\alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2)^*(\alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2)^{-1}(\alpha_1\mathbf{b}_1 + \alpha_2\mathbf{b}_2).$$

Then the following hold:

(i) $h(\boldsymbol{\alpha})$ is homogeneous, i.e., $h(\lambda\boldsymbol{\alpha}) = \lambda h(\boldsymbol{\alpha})$ for every $\lambda \neq 0$ and feasible $\boldsymbol{\alpha}$.

(ii) $h(\bar{\boldsymbol{\alpha}}) > 0$.

The first property is obvious by a simple substitution. To prove the second property note that by the weak duality theorem, $h(\bar{\boldsymbol{\alpha}})$ is greater than or equal to the value of the min-max problem (11). Let $\hat{\mathbf{z}}$ be the optimal solution of (11). Then $h(\bar{\boldsymbol{\alpha}}) \geq \max_{\mathbf{z}\in\Omega} \|\mathbf{z} - \hat{\mathbf{z}}\|^2$ and $\max_{\mathbf{z}\in\Omega} \|\mathbf{z} - \hat{\mathbf{z}}\|^2$ must be positive since, by our assumptions, Ω has a nonempty interior.

Suppose that $\bar{\alpha}_1\mathbf{A}_1 + \bar{\alpha}_2\mathbf{A}_2 \succ \mathbf{I}$. Then there exists $0 < \lambda < 1$ such that $\lambda\bar{\alpha}_1\mathbf{A}_1 + \lambda\bar{\alpha}_2\mathbf{A}_2 \succ \mathbf{I}$ so that $(\lambda\bar{\alpha}_1, \lambda\bar{\alpha}_2)$ is a feasible point of (13). However, from properties (i) and (ii),

$$h(\lambda\bar{\boldsymbol{\alpha}}) = \lambda h(\bar{\boldsymbol{\alpha}}) < h(\bar{\boldsymbol{\alpha}}),$$

contradicting the optimality of $\bar{\boldsymbol{\alpha}}$. ∎

4. The RCC estimator.

4.1. The RCC: Definition and form.

We now return to the problem of finding the Chebyshev center of FPS (5), which is the solution of the min-max problem (6). The set FPS can be represented as an intersection of two ellipsoids:

$$\text{FPS} = \{\mathbf{z} \in \mathbb{F}^n : \mathbf{z}^*\mathbf{L}^*\mathbf{L}\mathbf{z} \leq \eta, \quad \mathbf{z}^*\mathbf{A}^*\mathbf{A}\mathbf{z} - 2\Re(\mathbf{b}^*\mathbf{A}\mathbf{z}) + \|\mathbf{b}\|^2 \leq \rho\}.$$

We assume that condition (9) is satisfied, which means that

$$\exists\gamma_1 \geq 0, \gamma_2 \geq 0, \quad \gamma_1\mathbf{L}^*\mathbf{L} + \gamma_2\mathbf{A}^*\mathbf{A} \succ \mathbf{0}. \tag{21}$$

By Theorem 3.1, if $\mathbb{F} = \mathbb{C}$ and there exists $\tilde{\mathbf{z}}$ such that $\|\mathbf{A}\tilde{\mathbf{z}} - \mathbf{b}\|^2 < \rho$, $\|\mathbf{L}\tilde{\mathbf{z}}\|^2 < \eta$, then the Chebyshev center of FPS has the form

$$\hat{\mathbf{z}} = \alpha_2(\alpha_1\mathbf{L}^*\mathbf{L} + \alpha_2\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{b},$$

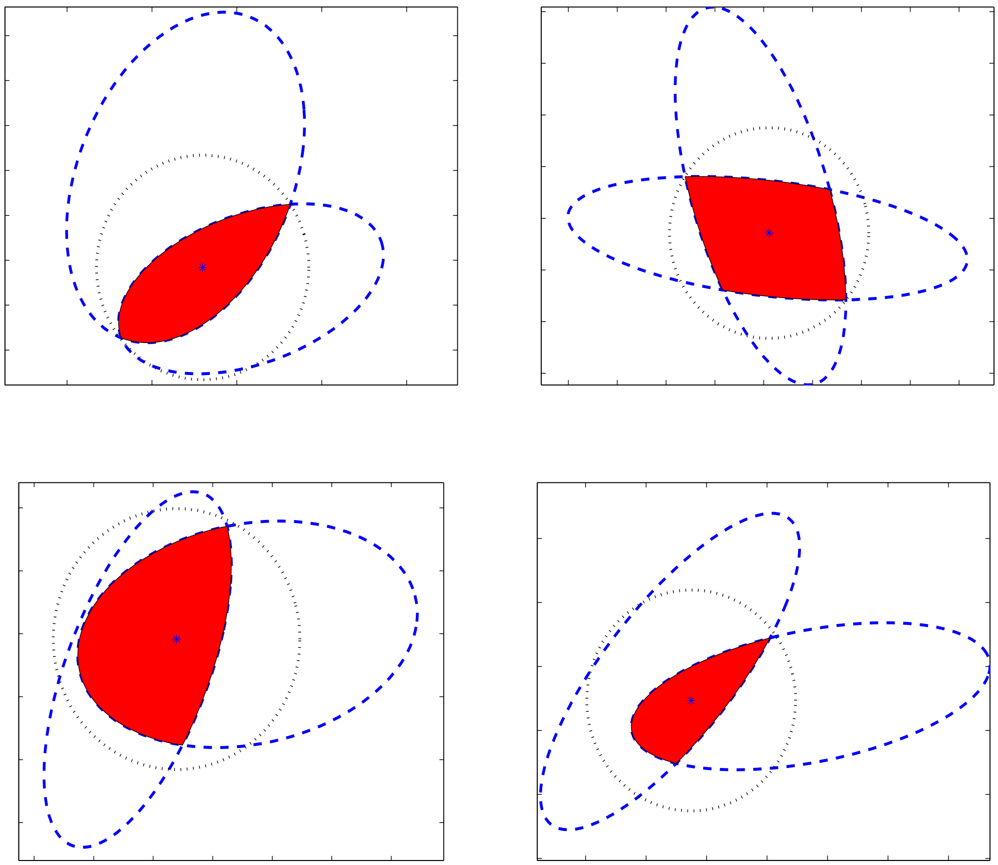

FIG. 2. *Four examples of intersection of ellipsoids (dashed lines). The filled area is the intersection of the ellipsoids. The center of the dotted circle is given by* (12) *with* (α_1, α_2) *being an optimal solution of* (13) *and the radius being the square root of the corresponding optimal value.*

where (α_1, α_2) is an optimal solution of the problem

$$\begin{array}{lll} & \min_{\alpha_1,\alpha_2} & \left\{\alpha_1\eta + \alpha_2(\rho - \|\mathbf{b}\|^2) + \alpha_2^2\mathbf{b}^*\mathbf{A}(\alpha_1\mathbf{L}^*\mathbf{L} + \alpha_2\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{b}\right\} \\ (22) & \text{s.t.} & \alpha_1\mathbf{L}^*\mathbf{L} + \alpha_2\mathbf{A}^*\mathbf{A} \succeq \mathbf{I}, \\ & & \alpha_1, \alpha_2 \geq 0. \end{array}$$

We now define $\hat{\mathbf{z}}$ for both the real and complex domains, and for the case when the conditions stated above are not necessarily satisfied.

DEFINITION 4.1. *The relaxed Chebyshev center (RCC) estimator is the vector*

$$\hat{\mathbf{z}}_{\text{RCC}} = \alpha_2(\alpha_1\mathbf{L}^*\mathbf{L} + \alpha_2\mathbf{A}^*\mathbf{A})^{-1}\mathbf{A}^*\mathbf{b},$$

where (α_1, α_2) *is an optimal solution of the convex optimization problem* (22).

If the optimal α_2 is positive, then the RCC estimator can be written as

$$\hat{\mathbf{z}}_{\text{RCC}} = (\mathbf{A}^*\mathbf{A} + (\alpha_1/\alpha_2)\mathbf{L}^*\mathbf{L})^{-1}\mathbf{A}^*\mathbf{b}. \tag{23}$$

Therefore, the RCC estimator is essentially a Tikhonov regularization with a special choice of λ that also takes into account the bounded noise constraint. This is in contrast to the choice of the regularization parameter in the RLS estimator that exploits only the norm constraint $\|\mathbf{Lz}\|^2 \le \eta$.

In section 5 we demonstrate that although the RCC estimator is only an approximation of the Chebyshev center in the real domain, it can still significantly outperform the LS and RLS methods with respect to the estimation error. This is the case even when the bound on the noise is loose; thus, with almost the same information as used by the RLS approach, we can significantly reduce the estimation error by using our proposed strategy. The two key ingredients that lead to the improved performance are treating the estimation error directly and the added constraint on the noise.

4.2. The case $\mathbf{L} = \mathbf{I}$. We now show that in the interesting special case $\mathbf{L} = \mathbf{I}$, the task of calculating the RCC estimator reduces to a single-variable convex minimization problem. To this end we rely on Theorem 3.3.

THEOREM 4.1. *Let $\mathbf{L} = \mathbf{I}$ and denote $\delta = \lambda_{\min}(\mathbf{A}^*\mathbf{A})$. Then the RCC estimator is given by*

$$\hat{\mathbf{z}}_{\mathrm{RCC}} = \begin{cases} (\mathbf{A}^*\mathbf{A} + \lambda\mathbf{I})^{-1}\mathbf{A}^*\mathbf{b}, & 0 \le \lambda < \infty, \\ \mathbf{0}, & \lambda = \infty, \end{cases}$$

where λ is determined as follows[3]*:*

(i) *If $\delta > 0$, then $\lambda = 1/\mu - \delta$, where μ is the solution of the convex minimization problem*

$$\min_{0 \le \mu \le 1/\delta} \left\{ (1 - \delta\mu)\eta + \mu(\rho - \|\mathbf{b}\|^2) + \mu^2 \mathbf{b}^*\mathbf{A}(\mu(\mathbf{A}^*\mathbf{A} - \delta\mathbf{I}) + \mathbf{I})^{-1}\mathbf{A}^*\mathbf{b} \right\}. \tag{24}$$

(ii) *If $\delta = 0$, then $\lambda = 1/\xi$, where ξ is the solution of the convex minimization problem*

$$\min_{\xi \ge 0} \left\{ \xi(\rho - \|\mathbf{b}\|^2) + \xi^2 \mathbf{b}^*\mathbf{A}(\xi\mathbf{A}^*\mathbf{A} + \mathbf{I})^{-1}\mathbf{A}^*\mathbf{b} \right\}. \tag{25}$$

Proof. Substituting $\mathbf{L} = \mathbf{I}$ into (22) we find

$$\begin{array}{ll} \min_{\alpha_1,\alpha_2} & \left\{ \alpha_1\eta + \alpha_2(\rho - \|\mathbf{b}\|^2) + \alpha_2^2 \mathbf{b}^*\mathbf{A}(\alpha_2\mathbf{A}^*\mathbf{A} + \alpha_1\mathbf{I})^{-1}\mathbf{A}^*\mathbf{b} \right\} \\ \text{s.t.} & \alpha_2\mathbf{A}^*\mathbf{A} + \alpha_1\mathbf{I} \succeq \mathbf{I}, \\ & \alpha_1, \alpha_2 \ge 0. \end{array} \tag{26}$$

The LMI constraint can be written equivalently as

$$\alpha_2\lambda_{\min}(\mathbf{A}^*\mathbf{A}) + \alpha_1 = \alpha_2\delta + \alpha_1 \ge 1. \tag{27}$$

From Theorem 3.3, we conclude that (27) must be satisfied with equality. Therefore,

$$\alpha_1 = 1 - \delta\alpha_2. \tag{28}$$

Substituting (28) into (26) we obtain that in the case $\delta > 0$, (26) becomes

$$\begin{array}{ll} \min_{\alpha_2} & \left\{ (1 - \delta\alpha_2)\eta + \alpha_2(\rho - \|\mathbf{b}\|^2) + \alpha_2^2 \mathbf{b}^*\mathbf{A}(\alpha_2(\mathbf{A}^*\mathbf{A} - \delta\mathbf{I}) + \mathbf{I})^{-1}\mathbf{A}^*\mathbf{b} \right\} \\ \text{s.t.} & 0 \le \alpha_2 \le 1/\delta, \end{array}$$

[3]We use the standard terminology $\frac{a}{0} = \infty$ whenever $a > 0$.

which is the same as (24) after μ is replaced by α_2. The result for the case $\delta = 0$ is similarly derived. □

To solve the single-variable convex problems (24) and (25), we can use any solver of one-dimensional convex minimization problems—for instance, a simple bisection algorithm on the derivative of the function. Denoting by $q(\mu)$ and $q'(\mu)$ the objective in (24) and its derivative, respectively, we have

$$\begin{aligned} q'(\mu) = &-\delta\eta + \rho - \|\mathbf{b}\|^2 + 2\mu\mathbf{b}^*\mathbf{A}(\mu(\mathbf{A}^*\mathbf{A} - \delta\mathbf{I}) + \mathbf{I})^{-1}\mathbf{A}^*\mathbf{b} \\ &-\mu^2\mathbf{b}^*\mathbf{A}(\mu(\mathbf{A}^*\mathbf{A} - \delta\mathbf{I}) + \mathbf{I})^{-1}(\mathbf{A}^*\mathbf{A} - \delta\mathbf{I})(\mu(\mathbf{A}^*\mathbf{A} - \delta\mathbf{I}) + \mathbf{I})^{-1}\mathbf{A}^*\mathbf{b}. \end{aligned}$$

Since $\mu(\mathbf{A}^*\mathbf{A} - \delta\mathbf{I}) + \mathbf{I}$ is a positive definite matrix for every choice of $\mu \geq 0$, we can calculate the derivative using a single Cholesky factorization in the following manner.

Calculation of $q'(\mu)$.

1. Calculate a Cholesky factorization $\mathbf{D}^*\mathbf{D} = \mu(\mathbf{A}^*\mathbf{A} - \delta\mathbf{I}) + \mathbf{I}$.
2. Solve the system $\mathbf{D}^*\mathbf{y} = \mathbf{A}^*\mathbf{b}$.
3. Solve the system $\mathbf{D}\mathbf{x} = \mathbf{y}$.
4. The derivative is given by $q'(\mu) = -\delta\eta+\rho-\|\mathbf{b}\|^2+2\mu\mathbf{b}^*\mathbf{A}\mathbf{x}-\mu^2\mathbf{x}^*(\mathbf{A}^*\mathbf{A}-\delta\mathbf{I})\mathbf{x}$.

Note that the Cholesky factorization is the most expensive component in the calculation of $q'(\mu)$ (the calculation of $\mathbf{A}^*\mathbf{A}$ is done in a preprocess). The other operations—solution of triangular systems and matrix/vector multiplications—are significantly cheaper. An alternative approach for computing the derivative is using the singular value decomposition of $\mathbf{A}$. This approach is viable for small-size problems but is not applicable for medium- and large-scale problems in which the Cholesky or the sparse Cholesky factorization can be employed. The complete description of the algorithm for calculating the RCC estimator when $\mathbf{L} = \mathbf{I}$ and $\delta > 0$ is as follows.

Algorithm RCC-S.

Input: $\mathbf{A} \in \mathbb{F}^{m\times n}$, the model matrix; $\mathbf{b} \in \mathbb{F}^m$, the (noisy) right-hand side vector; η, an upper bound on $\|\mathbf{z}\|^2$; and ρ, an upper bound on the squared-norm of the noise $\|\mathbf{A}\mathbf{z} - \mathbf{b}\|^2$.

Output: The RCC estimator $\hat{\mathbf{z}}_{\mathrm{RCC}}$, which is the solution to problem (22) with $\mathbf{L} = \mathbf{I}$.

1. If $q'(0) \geq 0$ then $h = 0$, and go to step 5.
2. If $q'(1/\delta) \leq 0$ then $h = 1/\delta$, and go to step 5.
3. Set $lb = 0, ub = 1/\delta$.
4. Repeat the following steps until $|ub - lb| < \eta$:
   (a) Set $h = \frac{lb+ub}{2}$.
   (b) Calculate $d = q'(h)$.
   (c) If $d \geq 0$ then $ub = h$; else $lb = h$.
5. Set $\hat{\mathbf{z}}_{\mathrm{RCC}} = (\mathbf{A}^*\mathbf{A} + (1 - \delta h)/h\mathbf{I})^{-1}\mathbf{A}^*\mathbf{b}$.

A similar algorithm can be defined for the case $\delta = 0$.

5. Numerical examples. We now present some examples comparing the RCC estimator with the LS and RLS methods, given by (2) and (3), respectively. The comparison was employed on two sets of problems: randomly chosen problems and the discretized inverse heat equation from “Regularization tools” [15]. All experiments were performed in MATLAB.

We note that in the simulations we assume knowledge of a very loose bound ρ on the noise so that essentially our method exploits almost the same knowledge as the RLS approach.

5.1. Random test problems. We chose a problem with dimensions $m = 10, n = 7$. Each component of $\mathbf{A}$ was randomly generated from a uniform distribution (i.e., $\mathbf{A} = \text{rand}(m, n)$). The "true" vector $\mathbf{z}_\text{T}$ is the vector of all ones. In the constraints, $\mathbf{L} = \mathbf{I}$ and $\eta = 2\|\mathbf{z}_\text{T}\|^2$. The observed vector $\mathbf{b}$ was generated by

$$\mathbf{b} = \mathbf{A}\mathbf{z}_\text{T} + \sigma\mathbf{w},$$

where σ takes the values $0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$ and each component of $\mathbf{w}$ was randomly generated from a standard normal distribution. The upper bound on the squared norm of the noise was chosen as $10\|\mathbf{w}\|^2$ (i.e., 10 times the true squared norm).

Table 1 describes the average of the data error $\|\mathbf{A}\hat{\mathbf{z}} - \mathbf{b}\|^2$ (here $\hat{\mathbf{z}}$ is $\hat{\mathbf{z}}_\text{LS}, \hat{\mathbf{z}}_\text{RLS}$, or $\hat{\mathbf{z}}_\text{RCC}$) and the squared error residual $\|\mathbf{z}_\text{T} - \hat{\mathbf{z}}\|^2$ over 100 realizations of $\mathbf{w}$. The best results in each half row are marked in boldface. The RLS approach (3) was implemented using the function lsqi from [15] and the RCC estimator was generated by the RCC-S algorithm of section 4.2.

TABLE 1
Comparison of the LS, RLS, and RCC estimators with respect to estimation error and data error.

σ	Squared estimation error			Squared data error		
	LS	RLS	RCC	LS	RLS	RCC
0.01	**1.9e-3**	**1.9e-3**	**1.9e-3**	**3.0e-4**	**3.0e-4**	**3.0e-4**
0.1	2.1e-1	2.1e-1	**1.8e-1**	**3.0e-2**	**3.0e-2**	3.1e-2
0.2	6.6e-1	6.6e-1	**3.6e-1**	**1.2e-1**	**1.2e-1**	1.5e-1
0.3	1.8e+0	1.8e+0	**2.0e-1**	**2.6e-1**	**2.6e-1**	1.2e+0
0.4	3.1e+0	2.9e+0	**2.5e-1**	**5.3e-1**	**5.3e-1**	2.7e+0
0.5	4.3e+0	3.9e+0	**3.3e-1**	**7.2e-1**	7.3e-1	4.8e+0
0.6	6.5e+0	5.0e+0	**4.7e-1**	**1.1e+0**	1.2e+0	7.3e+0
0.7	9.6e+0	5.4e+0	**5.5e-1**	**1.5e+0**	1.7e+0	1.0e+1
0.8	1.4e+1	6.6e+0	**6.8e-1**	**2.1e+0**	2.4e+0	1.4e+1
0.9	1.5e+1	6.7e+0	**9.0e-1**	**2.4e+0**	2.9e+0	1.8e+1
1.0	1.8e+1	6.8e+0	**9.9e-1**	**2.8e+0**	3.6e+0	2.2e+1

It is evident that the LS and RLS estimators are significantly and consistently worse than the RCC method with respect to the estimation error. This is despite the fact that the bound on the data error was chosen to be very large—much larger than the true bound. Thus, this approach does not require much prior information. On the other hand, the LS and RLS estimators result in a smaller data error than the RCC approach. This is not surprising since, as was already mentioned, the RCC method is designed to minimize a measure of estimation error while the LS and RLS strategies are aimed at minimizing the data error, which is less relevant in an *estimation* context.

Note that the RCC method was implemented in the case $\mathbb{F} = \mathbb{R}$. Recall that in the real domain the RCC estimator is only an approximation of the Chebyshev center of the FPS. We also implemented a set of random examples over $\mathbb{F} = \mathbb{C}$. The results were essentially the same as those reported in Table 1. Therefore, it seems that at least from an empirical point of view the RCC strategy is a "good enough" approximation of the Chebyshev center.

As we already pointed out, both the RCC and RLS strategies are Tikhonov estimators with different regularization parameters. In all the simulations in this section we observed that the regularization parameter of the RCC method,[4] denoted by λ_RCC,

[4] α_2 was always nonzero in our experiments.

is consistently greater than or equal to the parameter λ_{RLS} of the RLS approach. Furthermore, the RCC estimator was always feasible so that $\|\mathbf{L}\hat{\mathbf{z}}_{\mathrm{RCC}}\|^2 \leq \eta$. The latter observation explains why $\lambda_{\mathrm{RCC}} \geq \lambda_{\mathrm{RLS}}$. To see this, we define $\varphi(\lambda) \equiv \|\mathbf{L}\mathbf{z}_\lambda\|^2$, where $\mathbf{z}_\lambda$ is given by (4). It is straightforward to show that the function φ is strictly decreasing under our assumption (21). Now, if $\|\mathbf{L}\hat{\mathbf{z}}_{\mathrm{LS}}\|^2 \leq \eta$, then $\lambda_{\mathrm{RLS}} = 0$, which immediately implies $\lambda_{\mathrm{RCC}} \geq \lambda_{\mathrm{RLS}}$. Otherwise, when $\|\mathbf{L}\hat{\mathbf{z}}_{\mathrm{LS}}\|^2 > \eta$, λ_{RLS} satisfies $\varphi(\lambda_{\mathrm{RLS}}) = \eta$. On the other hand, $\varphi(\lambda_{\mathrm{RCC}}) = \|\mathbf{L}\hat{\mathbf{z}}_{\mathrm{RCC}}\|^2 \leq \eta$, and by the fact that φ is decreasing, $\lambda_{\mathrm{RCC}} \geq \lambda_{\mathrm{RLS}}$.

5.2. Inverse heat equation. We now treat the problem of estimating the function $f(t)$ that solves the heat equation

$$\int_0^1 k(s-t)f(t) = g(s),$$

with $k(t) = \frac{t^{-3/2}}{2\sqrt{\pi}} \exp\left(-\frac{1}{4t}\right)$. By means of a simple collocation and midpoint rule with n points, the problem reduces to an $n \times n$ linear system $\mathbf{A}\mathbf{z}_{\mathrm{T}} = \mathbf{b}_{\mathrm{T}}$. This system and its solution $\mathbf{z}_{\mathrm{T}}$ are implemented in the function heat(n,1) from [15]. We note that this example is ill-conditioned. We compare the RCC estimator to the RLS method (the results for the LS approach are not given because it produces extremely poor results).

The perturbed right-hand side is chosen as

$$\mathbf{b} = \mathbf{b}_{\mathrm{T}} + 10^{-4}\mathbf{w}, \tag{29}$$

where each component of $\mathbf{w}$ is generated from a standard normal distribution. The matrix $\mathbf{L}$ approximates the first-derivative operator implemented in the function get_l(n,1) from [15]. The upper bound η was chosen to be $2\|\mathbf{L}\mathbf{z}_{\mathrm{T}}\|^2$. In Figure 3 three possible values of ρ were employed: $\rho = \|\mathbf{w}\|^2$ (exact squared norm), $\rho = 2\|\mathbf{w}\|^2$, and $\rho = 10\|\mathbf{w}\|^2$. The results of the RCC estimator in these three cases are very similar and are much closer to the true vector $\mathbf{z}_{\mathrm{T}}$ than the RLS solution $\hat{\mathbf{z}}_{\mathrm{RLS}}$. Therefore, it seems that at least in this example, the performance of the RCC method is quite robust with respect to the choice of ρ. The fourth plot in Figure 3 describes the three vectors $\mathbf{A}\mathbf{z}_{\mathrm{T}}, \mathbf{A}\hat{\mathbf{z}}_{\mathrm{RLS}}$, and $\mathbf{A}\hat{\mathbf{z}}_{\mathrm{RCC}}$ (for $\rho = 10\|\mathbf{w}\|^2$). It can be readily seen that the three vectors are almost identical, implying that the data errors of the RLS and RCC approaches are both negligible.

5.3. Ellipsoid versus interior-point methods. In the case when $\mathbf{L} \neq \mathbf{I}$ (as in the inverse heat equation problem), we are required to solve the convex optimization problem (22) with two variables, or the SDP

$$\begin{array}{ll} \min_{\alpha_1,\alpha_2,t} & \left\{\alpha_1\eta + \alpha_2(\rho - \|\mathbf{b}\|^2) + t\right\} \\ \text{s.t.} & \alpha_1\mathbf{L}^*\mathbf{L} + \alpha_2\mathbf{A}^*\mathbf{A} \succeq \mathbf{I}, \\ & \begin{pmatrix} \alpha_1\mathbf{L}^*\mathbf{L} + \alpha_2\mathbf{A}\mathbf{A} & -\alpha_2\mathbf{A}^*\mathbf{b} \\ -\alpha_2\mathbf{A}\mathbf{b}^* & t \end{pmatrix} \succeq \mathbf{0}, \alpha_1 \geq 0, \alpha_2 \geq 0. \end{array} \tag{30}$$

Now, consider an SDP of the general form

$$\min\left\{\mathbf{c}^T\mathbf{x} : \sum_{i=1}^m x_i\mathbf{B}_i \succeq \mathbf{E}\right\},$$

where $\mathbf{c} \in \mathbb{R}^m$ and $\mathbf{E}, \mathbf{B}_i, i = 1, \ldots, m$, are $n \times n$ Hermitian matrices. In order to solve the general form SDP we can use a primal-dual interior-point method, which

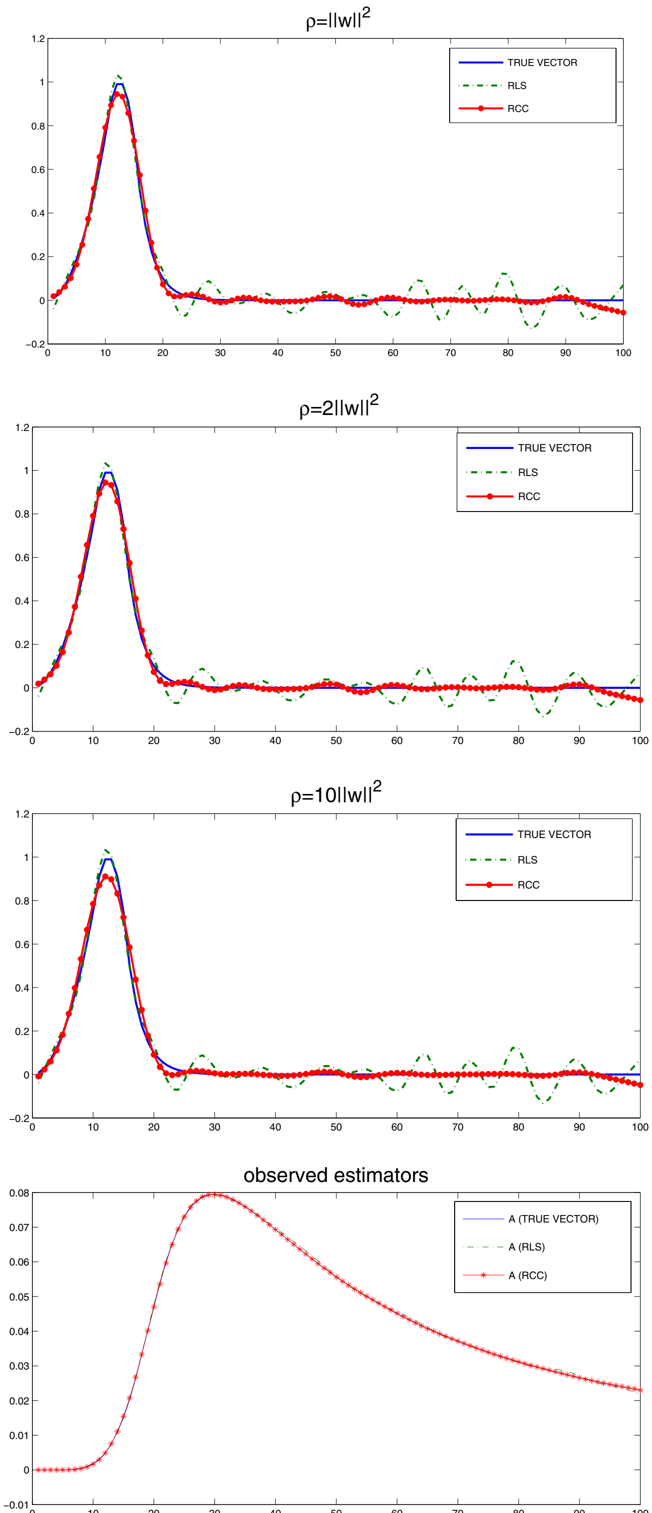

FIG. 3. *Results for the inverse heat problem of the RCC and RLS estimators.*

requires $O(n^{3.5}m^{1.5}+n^{2.5}m^2+n^{0.5}m^{2.5})$ operations per accuracy digit. For the specific problem (30) we have $m = 3$, and the amount of operations is therefore $O(n^{3.5})$.

Another alternative is to use the ellipsoid method [3] directly on the problem (22). This algorithm requires $O(n^3)$ operations per accuracy digit since it requires at most two Cholesky factorizations at each iteration. Therefore, it is cheaper than the SDP approach by a factor of order $\sqrt{n}$. To compare the performance of the two algorithms, we implemented the ellipsoid method (see the appendix for full details) and compared it to the interior-point method implemented in SeDuMi [27] on the inverse heat equation problem with various values of n. The CPU time in seconds of the ellipsoid and interior-point algorithms averaged over 10 realizations of the noise $\mathbf{w}$ is given in Table 2 below (σ was fixed to be 1e-4). For $n = 1000$ SeDuMi failed due to memory difficulties. Table 2 demonstrates the efficiency of the ellipsoid method.

TABLE 2
CPU time in seconds on a Pentium 4, 1.8Ghz.

n	Ellipsoid	SeDuMi
10	1.4e-1	5.5e-1
20	1.6e-1	9.3e-1
50	2.9e-1	3.5e+0
100	8.5e-1	1.8e+1
200	6.0e+0	1.0e+2
500	3.8e+1	8.3e+2
1000	2.4e+2	–

6. Extensions to other estimation problems. The RCC estimator was constructed to handle the situation in which only the right-hand side of the linear system $\mathbf{Ax} \approx \mathbf{b}$ is contaminated by noise. The same methodology can be applied to deal with other sources of noise. In this section, we briefly outline the resulting estimators in two scenarios: (i) both $\mathbf{A}$ and $\mathbf{b}$ are uncertain, and (ii) $\mathbf{A}$ and $\mathbf{b}$ are uncertain and regularization is required. In the first scenario, the proposed estimator has a similar structure to the well-known TLS method [13, 17], and in the second scenario, the estimator has a form similar to that of the RTLS solution [12]. Thus, these popular methods of handling uncertainties in the basic regression model (1) can be shown to be special cases of our general results.

The derivation of the estimators is very similar to that described in section 3; therefore, we present the main results without proof.

6.1. Uncertainty in both A and b. Suppose that both $\mathbf{A}$ and $\mathbf{b}$ are uncertain and are given by $\mathbf{A}+\boldsymbol{\Delta}, \mathbf{b}+\mathbf{w}$ with $\boldsymbol{\Delta}, \mathbf{w}$ being unknown but bounded perturbations. This setting is assumed in the robust LS approach [11]. We assume that the bound constraint is given by[5] $\|(\boldsymbol{\Delta}, \mathbf{w})\|_{\mathrm{F}}^2 \le \rho$. The corresponding FPS is

$$\mathrm{FPS}_1 = \{\mathbf{z} \in \mathbb{F}^n : \exists \boldsymbol{\Delta} \in \mathbb{F}^{m\times n}, \mathbf{w} \in \mathbb{F}^m : (\mathbf{A}+\boldsymbol{\Delta})\mathbf{z} = \mathbf{b}+\mathbf{w}, \|(\boldsymbol{\Delta}, \mathbf{w})\|_{\mathrm{F}}^2 \le \rho\}.$$

To apply our results, we first note that FPS_1 can be written as the single quadratic constraint

$$\mathrm{FPS}_1 = \{\mathbf{z} \in \mathbb{F}^n : \mathbf{z}^*(\mathbf{A}^*\mathbf{A} - \rho\mathbf{I})\mathbf{z} - 2\mathbf{z}^*\mathbf{A}^*\mathbf{b} + \|\mathbf{b}\|^2 - \rho \le 0\}. \tag{31}$$

This follows from writing FPS_1 as

$$\mathrm{FPS}_1 = \{\mathbf{z} \in \mathbb{F}^n : \mathbf{Az} - \mathbf{b} = \mathbf{E}\tilde{\mathbf{z}} \text{ for some } \|\mathbf{E}\|_{\mathrm{F}}^2 \le \rho\}$$

[5] For a matrix $\mathbf{B}$, $\|\mathbf{B}\|_{\mathrm{F}}$ denotes the Frobenius norm of $\mathbf{B}$.

and applying the following simple lemma.

LEMMA 6.1. *Let $\mathbf{x} \in \mathbb{F}^n$ and $\mathbf{y} \in \mathbb{F}^m$, and let η be a positive scalar. Then the following two statements are equivalent:*

(i) *There exists $\Delta \in \mathbb{F}^{m\times n}$ such that $\Delta\mathbf{x} = \mathbf{y}$ and $\|\Delta\|_F \le \eta$.*

(ii) *The inequality $\|\mathbf{y}\| \le \eta\|\mathbf{x}\|$ holds.*

Under the assumption that $\rho < \lambda_{\min}(\mathbf{A}^*\mathbf{A})$, it can be shown, using the same line of analysis of section 3, that the Chebyshev center of FPS_1 is given by

$$\hat{\mathbf{z}} = (\mathbf{A}^*\mathbf{A} - \rho\mathbf{I})^{-1}\mathbf{A}^*\mathbf{b}, \tag{32}$$

which is a *deregularization* of the LS solution. Since FPS_1 consists of a single quadratic constraint, this result is valid both in the real and in the complex domains. We note that when $\rho > \lambda_{\min}(\mathbf{A}^*\mathbf{A})$, FPS_1 is unbounded, and as a result the value of the inner maximization problem in (6) is always ∞, which implies that the Chebyshev center in this case is meaningless.

If we choose

$$\rho = \lambda_{\min}\begin{pmatrix} \mathbf{A}^*\mathbf{A} & \mathbf{A}^*\mathbf{b} \\ \mathbf{b}^*\mathbf{A} & \|\mathbf{b}\|^2 \end{pmatrix},$$

then the estimator (32) coincides with the TLS estimator [17, Theorem 2.7].

6.2. Uncertainty in both A and b with regularization. Suppose now we add regularization to the previous scenario; i.e., we consider the feasible set

$$\text{FPS}_2 = \{\mathbf{z} \in \mathbb{F}^n : \|\mathbf{L}\mathbf{z}\|^2 \le \eta, \exists\boldsymbol{\Delta} \in \mathbb{F}^{m\times n}, \mathbf{w} \in \mathbb{F}^m : (\mathbf{A}+\boldsymbol{\Delta})\mathbf{z} = \mathbf{b}+\mathbf{w}, \|(\boldsymbol{\Delta},\mathbf{w})\|_{\text{F}}^2 \le \rho\},$$

which can also be written as

$$\text{FPS}_2 = \{\mathbf{z} \in \mathbb{F}^n : \|\mathbf{L}\mathbf{z}\|^2 \le \eta, \mathbf{z}^*(\mathbf{A}^*\mathbf{A} - \rho\mathbf{I})\mathbf{z} - 2\mathbf{z}^*\mathbf{A}^*\mathbf{b} + \rho - \|\mathbf{b}\|^2 \le 0\}.$$

In the case $\mathbb{F} = \mathbb{C}$, the Chebyshev center of FPS_2 is given by

$$\hat{\mathbf{z}} = \alpha_2(\alpha_1\mathbf{L}^*\mathbf{L} + \alpha_2(\mathbf{A}^*\mathbf{A} - \rho\mathbf{I}))^{-1}\mathbf{A}^*\mathbf{b}, \tag{33}$$

where (α_1, α_2) is an optimal solution of the convex optimization problem

$$\begin{array}{ll} \min_{\alpha_1,\alpha_2} & \{\alpha_1\eta + \alpha_2(\rho - \|\mathbf{b}\|^2) + \alpha_2^2\mathbf{b}^*\mathbf{A}(\alpha_1\mathbf{L}^*\mathbf{L} + \alpha_2(\mathbf{A}^*\mathbf{A} - \rho\mathbf{I}))^{-1}\mathbf{A}^*\mathbf{b}\} \\ \text{s.t.} & \alpha_1\mathbf{L}^*\mathbf{L} + \alpha_2\mathbf{A}^*\mathbf{A} \succeq \mathbf{I}, \\ & \alpha_1, \alpha_2 \ge 0. \end{array}$$

If $\alpha_2 \ne 0$ then $\hat{\mathbf{z}}$ of (33) can be written as

$$\hat{\mathbf{z}} = (\mathbf{A}^*\mathbf{A} - \rho\mathbf{I} + \alpha_1/\alpha_2\mathbf{L}^*\mathbf{L})^{-1}\mathbf{A}^*\mathbf{b}.$$

This estimator has the same structure as the RTLS method, which solves the equation

$$(\mathbf{A}^*\mathbf{b} - \lambda\mathbf{I} + \mu\mathbf{L}^*\mathbf{L})\mathbf{x}_{\text{RTLS}} = \mathbf{A}^*\mathbf{b}$$

for some choice of parameters λ, μ [12].

7. Conclusion. In this paper we discussed a Chebyshev center regularization method that is based on an estimation error criterion. In contrast to previous regularization strategies that invoke a data error–based criterion, here we focus on the estimation error and try to minimize it in some sense. Since the estimation error depends on the unknown vector, we choose as our estimate the Chebyshev center of an FPS, which consists of a constraint both on the data error and on the weighted norm of the true parameter. Although the resulting problem is nonconvex, by exploiting recent duality results, we show that in the complex domain it can be formulated as a solution to a convex optimization problem in two unknowns, and in the real case the same approach can be used to get a "pretty good" approximation of the true Chebyshev center. From a numerical standpoint, we provide two solution methods and compare their performance. The first is based on an SDP and the second on an ellipsoid algorithm. The latter turns out to be more efficient as the problem size grows. Finally, we show that the popular TLS and RTLS methods can also be formulated within our framework.

Appendix. The ellipsoid method for problem (13).

In this appendix we describe in detail the ellipsoid method as applied to the convex optimization problem (13).

The two basic ingredients in the ellipsoid method are a separation oracle and a first-order oracle (see, e.g., [3]). The main linear algebra procedure we use in both oracles is the Cholesky factorization. We assume that the input to the Cholesky procedure is a symmetric matrix $\mathbf{B}$, and its output consists of three arguments **flag,** $\mathbf{D}$, and $\mathbf{x}$. If $\mathbf{flag = 1}$ then $\mathbf{B}$ is positive definite, $\mathbf{B} = \mathbf{D}^*\mathbf{D}$ with $\mathbf{D}$ being a lower triangular matrix, and $\mathbf{x}$ is NULL. If $\mathbf{flag = 0}$ then $\mathbf{B}$ is not positive definite, $\mathbf{x}$ is a vector satisfying $\mathbf{x}^*\mathbf{B}\mathbf{x} \le 0$, and $\mathbf{D}$ is NULL.

The input to the separation oracle is a vector $\boldsymbol{\alpha} \in \mathbb{R}^2$. The output is either a statement that the vector is feasible (up to some tolerance) or a hyperplane separating the vector from the feasible set. ϵ is a tolerance parameter chosen as 10^{-6} in our implementation.

Algorithm SEP-ORA.
Input: $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T \in \mathbb{R}^2$.
Output: **flag** equals one if $\boldsymbol{\alpha}$ is feasible (up to some tolerance) and zero otherwise. $\mathbf{d} \in \mathbb{R}^2$ is a separating hyperplane.

1. If $\alpha_1 \le -\epsilon$ then **flag**=0, $\mathbf{d} = (-1, 0)^T$, STOP.
2. If $\alpha_2 \le -\epsilon$ then **flag**=0, $\mathbf{d} = (0, -1)^T$, STOP.
3. Set $\mathbf{M} = \alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2 - \mathbf{I} + \epsilon\mathbf{I}$.
4. Invoke the Cholesky factorization procedure with input $\mathbf{M}$ and obtain an output $\{\mathbf{flag, D, x}\}$.
   (a) If $\mathbf{flag = 1}$ then STOP.
   (b) If $\mathbf{flag = 0}$ then $(d_1, d_2) = (\mathbf{x}^*\mathbf{A}_1\mathbf{x}, \mathbf{x}^*\mathbf{A}_2\mathbf{x})$, STOP.

The first-order oracle is invoked in the case when the current vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$ is feasible. Its main computational effort is the Cholesky factorization of the matrix $\alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2$, which by feasibility of $\boldsymbol{\alpha}$, must be positive definite.

Algorithm FO-ORA.
Input: $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T \in \mathbb{R}^2$, an η-feasible solution of (13).
Output: $\mathbf{f}$, the gradient of the objective function of (13) at $\boldsymbol{\alpha}$.

1. Set $\mathbf{M} = \alpha_1\mathbf{A}_1 + \alpha_2\mathbf{A}_2$.
2. Invoke the Cholesky factorization procedure with input $\mathbf{M}$ and obtain an output $\{\mathbf{flag}, \mathbf{D}, \mathbf{x}\}$.

3. Solve the following linear systems in $\mathbf{x}_1, \mathbf{x}_2$: $\mathbf{D}^*\mathbf{x}_1 = \mathbf{b}_1, \mathbf{D}^*\mathbf{x}_2 = \mathbf{b}_2$.
4. Solve the following linear systems in $\mathbf{v}_1, \mathbf{v}_2$: $\mathbf{D}\mathbf{v}_1 = \mathbf{x}_1, \mathbf{D}\mathbf{v}_2 = \mathbf{x}_2$.
5. Set

$$f_1 = -c_1 + 2\alpha_1 \mathbf{b}_1^*\mathbf{v}_1 - \alpha_1^2 \mathbf{v}_1^*\mathbf{A}_1\mathbf{v}_1 + 2\alpha_2 \mathbf{b}_1^*\mathbf{v}_2 - 2\alpha_1\alpha_2 \mathbf{v}_1^*\mathbf{A}_1\mathbf{v}_2 - \alpha_2^2 \mathbf{v}_2^*\mathbf{A}_1\mathbf{v}_2,$$
$$f_2 = -c_2 + 2\alpha_2 \mathbf{b}_2^*\mathbf{v}_2 - \alpha_2^2 \mathbf{v}_2^*\mathbf{A}_2\mathbf{v}_2 + 2\alpha_1 \mathbf{b}_1^*\mathbf{v}_2 - 2\alpha_1\alpha_2 \mathbf{v}_1^*\mathbf{A}_2\mathbf{v}_2 - \alpha_1^2 \mathbf{v}_1^*\mathbf{A}_2\mathbf{v}_1.$$

We are now ready to describe the implementation of the ellipsoid method on the convex optimization problem (13).

Algorithm Ellipsoid.

Input: The optimization problem (13).

Output: $\boldsymbol{\alpha} \in \mathbb{R}^2$, a solution to problem (13) (up to some tolerance).

1. Set $R = 10^8, \mathbf{B} = R\mathbf{I}_2, \boldsymbol{\alpha} = (0,0)^T, v = \pi 10^{16}$.
2. Repeat the following steps until $v < \epsilon$.
   (a) Invoke the separation oracle SEP-ORA with input $\boldsymbol{\alpha}$ and obtain an output $\{\mathbf{flag}, \mathbf{d}\}$. If $\mathbf{flag = 0}$ then go to step (c).
   (b) Invoke the first-order oracle FO-ORA with input $\boldsymbol{\alpha}$ and obtain an output $\mathbf{d}$.
   (c) $\mathbf{p} = \frac{\mathbf{B}^T\mathbf{d}}{\sqrt{\mathbf{d}^T\mathbf{B}\mathbf{B}^T\mathbf{d}}}$.
   (d) $\boldsymbol{\alpha} = \boldsymbol{\alpha} - \frac{1}{3}\mathbf{B}\mathbf{p}$.
   (e) $\mathbf{B} = \frac{2}{\sqrt{3}}\mathbf{B} + (\frac{2}{3} - \frac{2}{\sqrt{3}})\mathbf{B}\mathbf{p}\mathbf{p}^T$.
   (f) $v = \pi \det(\mathbf{B})$.

REFERENCES

[1] A. Beck and Y. C. Eldar, *Strong duality in nonconvex quadratic optimization with two quadratic constraints*, SIAM J. Optim., 17 (2006), pp. 844–860.
[2] Z. Ben-Haim and Y. C. Eldar, *Maximum set estimators with bounded estimation error*, IEEE Trans. Signal Process., 53 (2005), pp. 3172–3182.
[3] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization*, MPS SIAM Ser. Optim. 2, SIAM, Philadelphia, 2001.
[4] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont, MA, 1999.
[5] A. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996.
[6] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
[7] Y. C. Eldar, *Minimum variance in biased estimation: Bounds and asymptotically optimal estimators*, IEEE Trans. Signal Process., 52 (2004), pp. 1915–1930.
[8] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, *Linear minimax regret estimation of deterministic parameters with bounded data uncertainties*, IEEE Trans. Signal Process., 52 (2004), pp. 2177–2188.
[9] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, *Robust mean-squared error estimation in the presence of model uncertainties*, IEEE Trans. Signal Process., 53 (2005), pp. 168–181.
[10] C. Fortin and H. Wolkowicz, *The trust region subproblem and semidefinite programming*, Optim. Methods Softw., 19 (2004), pp. 41–67.
[11] L. El Ghaoui and H. Lebret, *Robust solutions to least-squares problems with uncertain data*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 1035–1064.
[12] G. H. Golub, P. C. Hansen, and D. P. O'Leary, *Tikhonov regularization and total least squares*, SIAM J. Matrix Anal. Appl., 21 (1999), pp. 185–194.
[13] 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.
[14] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., Johns Hopkins University Press, Baltimore, MD, 1996.
[15] P. C. Hansen, *Regularization tools, a MATLAB package for analysis of discrete regularization problems*, Numer. Algorithms, 6 (1994), pp. 1–35.
[16] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed problems: Numerical Aspects of Linear Inversion*, SIAM Monogr. Math. Model. Comput. 4, SIAM, Philadelphia, 1997.

[17] S. Van Huffel and J. Vandewalle, *The Total Least Squares Problem: Computational Aspects and Analysis*, Frontiers Appl. Math. 9, SIAM, Philadelphia, 1991.

[18] T. Kailath, *Lectures on Linear Least-Squares Estimation*, Springer-Verlag, Wein, New York, 1976.

[19] M. Milanese and G. Belforte, *Estimation theory and uncertainty intervals evaluation in the presence of unknown but bounded errors: Linear families of models and estimators*, IEEE Trans. Automat. Control, 27 (1982), pp. 408–414.

[20] M. Milanese and R. Tempo, *Optimal algorithms theory for robust estimation and prediction*, IEEE Trans. Automat. Control, 30 (1985), pp. 730–738.

[21] M. Milanese and A. Vicino, *Optimal estimation theory for dynamic systems with set membership uncertainty: An overview*, Automatica J. IFAC, 27 (1991), pp. 997–1009.

[22] J. J. Moré, *Generalizations of the trust region subproblem*, Optim. Methods Softw., 2 (1993), pp. 189–209.

[23] Y. Nesterov and A. Nemirovskii, *Interior Point Polynomial Algorithms in Convex Programming*, Studies in Applied Mathematics 13, SIAM, Philadelphia, 1994.

[24] J. P. Norton, *Identification and application of bounded parameter models*, Automatica J. IFAC, 23 (1987), pp. 497–507.

[25] M. S. Pinsker, *Optimal filtering of square-integrable signals in Gaussian noise*, Prob. Inf. Transm., 16 (1980), pp. 120–133.

[26] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[27] J. F. Sturm, *Using SeDuMi* 1.02*, a MATLAB toolbox for optimization over symmetric cones*, Optim. Methods Softw., 11/12 (1999), pp. 625–653.

[28] A. N. Tikhonov, *Solution of incorrectly formulated problems and the regularization method*, Soviet. Math. Dokl., 5 (1963), pp. 1035–1038.

[29] J. F. Traub, G. Wasikowski, and H. Wozinakowski, *Information-based Complexity*, Academic Press, New York, 1988.

[30] S. Xu, R. M. Freund, and J. Sun, *Solution methodologies for the smallest enclosing circle problem*, Comput. Optim. Appl., 25 (2003), pp. 283–292.