A Minimax Chebyshev Estimator for Bounded Error Estimation

Yonina C. Eldar, *Senior Member, IEEE*, Amir Beck, and Marc Teboulle

Abstract—We develop a nonlinear minimax estimator for the classical linear regression model assuming that the true parameter vector lies in an intersection of ellipsoids. We seek an estimate that minimizes the worst-case estimation error over the given parameter set. Since this problem is intractable, we approximate it using semidefinite relaxation, and refer to the resulting estimate as the relaxed Chebyshev center (RCC). We show that the RCC is unique and feasible, meaning it is consistent with the prior information. We then prove that the constrained least-squares (CLS) estimate for this problem can also be obtained as a relaxation of the Chebyshev center, that is looser than the RCC. Finally, we demonstrate through simulations that the RCC can significantly improve the estimation error over the CLS method.

Index Terms—Bounded error estimation, Chebyshev center, constrained least-squares, semidefinite programming, semidefinite relaxation.

I. INTRODUCTION

MANY estimation problems in a broad range of applications can be written in the form of a linear regression model. In this class of problems, the goal is to construct an estimate $\hat{\mathbf{x}}$ of a deterministic parameter vector $\mathbf{x}$ from noisy observations $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$, where $\mathbf{A}$ is a known model matrix and $\mathbf{w}$ is an unknown perturbation vector.

The celebrated least-squares (LS) method minimizes the data error $\|\hat{\mathbf{y}} - \mathbf{y}\|^2$ between the estimated data $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}}$ and $\mathbf{y}$. This approach is deterministic in nature, as no statistical information is assumed on $\mathbf{x}$ or $\mathbf{w}$. Nonetheless, if the covariance of $\mathbf{w}$ is known, then it can be incorporated as a weighting matrix, such that the resulting weighted LS estimate minimizes the variance among all unbiased methods. However, this does not necessarily lead to a small estimation error $\hat{\mathbf{x}} - \mathbf{x}$. In particular, when $\mathbf{A}$ is ill-conditioned, for example, when the system is obtained via discretization of ill-posed problems such as integral equations of the first kind [1] or in certain image processing problems [2],

Y. C. Eldar is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: yonina@ee.technion.ac.il).

A. Beck is with the Department of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa 32000, Israel (e-mail: becka@ie.technion.ac.il).

M. Teboulle is with the School of Mathematical Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel (e-mail: teboulle@post.tau.ac.il).

the LS will give poor results with respect to the estimation error. Thus, many attempts have been made to develop estimators that may be biased but closer to $\mathbf{x}$ in some statistical sense [3]–[10]. By now it is well established that even though unbiasedness may be appealing intuitively, it does not necessarily lead to a small estimation error $\hat{\mathbf{x}} - \mathbf{x}$ [11].

A popular strategy for improving the estimation error of LS is to incorporate prior information on $\mathbf{x}$. For example, the Tikhonov estimator minimizes the data error subject to a weighted norm constraint on $\mathbf{x}$ [5]. In practical applications, more general restrictions on $\mathbf{x}$ can be given, such as interval constraints on the individual components of $\mathbf{x}$. These type of bounds rise naturally e.g., in image processing where the pixel values are limited. To deal with more general type of restrictions, the constrained LS estimator (CLS) has been proposed, which minimizes the data error subject to the constraint that $\mathbf{x}$ lies in a convex set $\mathcal{C}$ [12]. However, this method does not deal directly with the estimation error.

In some scenarios, the distribution of the noise may not be known, or the noise may not be random (for example, in problems resulting from quantization). A common estimation technique in these settings is the bounded error approach, also referred to as set-membership estimation (see e.g., [13] and the survey papers [14], [15]). This strategy is designed to deal with bounded noise, and prior information of the form $\mathbf{x} \in \mathcal{C}$ for some set $\mathcal{C}$.

In this paper, we adopt the bounded error methodology and assume that the noise is norm-bounded $\|\mathbf{w}\|^2 \leq \rho$. The estimator we develop can also be used when $\mathbf{w}$ is random by choosing ρ proportional to its variance. We further suppose that $\mathbf{x} \in \mathcal{C}$ where we focus on sets $\mathcal{C}$ that are given by an intersection of ellipsoids. This form of $\mathcal{C}$ is quite general and includes a large variety of structures, among them are weighted norm constraints, and interval restrictions. Since we do not assume a statistical model, our objective is to choose $\hat{\mathbf{x}}$ to be close to $\mathbf{x}$ in the squared error sense. Thus, instead of minimizing the data error, we suggest minimizing the worst-case estimation error $\|\hat{\mathbf{x}} - \mathbf{x}\|^2$ over all feasible solutions. As we show in Section II, the proposed minimax estimator has a nice geometric interpretation in terms of the center of the minimum radius ball enclosing the feasible set. Therefore, this methodology is also referred to as the Chebyshev center approach [16], [17]. In Section VII, we demonstrate that this strategy can indeed reduce the estimation error dramatically with respect to the CLS method.

Finding the Chebyshev center of a set is a difficult and typically intractable problem. Two exceptions are when the set is polyhedral and the estimation error is measured by the l_∞ norm [18], and when the set is finite [19]. Recently, we considered this approach for $\mathcal{C}$ given by an ellipsoid [20]. When the problem is

defined over the complex domain we showed that the Chebyshev center can be computed exactly by relying on strong duality results [21]. In the real domain, we suggested an approximation based on Lagrange duality and semidefinite relaxation, referred to as the relaxed Chebyshev center (RCC). We then showed through numerical simulations that the RCC estimate outperforms other estimates such as least squares and Tikhonov with respect to the estimation error.

Here, we generalize the RCC estimator to the intersection of several ellipsoids in order to extend its applicability to a larger set of signal processing problems. In addition, we further explore the properties of the RCC estimate, and compare it to the CLS approach, thus providing new results even for the case of a single ellipsoid treated in [20]. In particular, we show that both the exact Chebyshev center and the RCC are unique and feasible, meaning they satisfy the prior constraints and are therefore consistent with the given prior knowledge. Our development of the RCC estimate in this paper is different than that presented for the single ellipsoid case in [20]. Here, we use the fact that the RCC can be cast as a solution to a convex-concave saddle point program. This method of proof leads to an alternative representation of the RCC which has several advantages. First, it is simpler to solve than the minimization problem obtained in [20]. Second, the same method can be used to show that the CLS estimator can also be viewed as a relaxation of the Chebyshev center, however this relaxation is looser than that of the RCC. This motivates the observation that in many different examples the RCC leads to smaller estimation error. Third, this approach immediately reveals the feasibility of the RCC solution. Finally, as we discuss next, this strategy can be used to explore methods for representing linear constraints.

In many applications there are interval restrictions on $\mathbf{x}$ of the form $\ell_i \leq x_i \leq u_i$. These constraints can be written as two linear inequalities, or a single quadratic constraint $(x_i - \ell_i)(x_i - u_i) \leq 0$. The true Chebyshev center is the same regardless of the specific characterization chosen, as it is defined by the constraint set $\mathcal{C}$ which is the same in both cases. However, as we show, the RCC depends on the specific representation of the set. An important question therefore in the context of the RCC is how to best represent linear constraints in order to obtain a smaller estimation error. Here we show that the quadratic characterization leads to a tighter approximation of the true minimax value and is therefore preferable in this sense.

The paper is organized as follows. In Section II, we discuss the geometrical properties of the Chebyshev center and its application to estimation problems. We then develop in Section III the RCC using a simpler method than that in [20]. The feasibility of the Chebyshev center and the RCC estimate are discussed in Section IV. In Section V, we show that the CLS estimate can also be viewed as a relaxation of the Chebyshev center, which is looser than the RCC. The representation of linear constraints is discussed in Section VI. In Section VII, we demonstrate via examples that the RCC strategy can dramatically reduce the estimation error with respect to the CLS method.

II. THE CHEBYSHEV CENTER

We denote vectors by boldface lowercase letters, e.g., $\mathbf{y}$, and matrices by boldface uppercase letters, e.g., $\mathbf{A}$. The ith compo-

nent of a vector $\mathbf{y}$ is written as y_i, and $\hat{(\cdot)}$ is an estimated vector. The identity matrix is denoted by $\mathbf{I}$, and $\mathbf{A}^T$ is the transpose 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).

We treat the problem of estimating a deterministic parameter vector $\mathbf{x} \in \mathcal{R}^m$ from observations $\mathbf{y} \in \mathcal{R}^n$ which are related through the linear model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}. \tag{1}$$

Here, $\mathbf{A}$ is a known $n \times m$ model matrix, $\mathbf{w}$ is a perturbation vector with bounded norm $\|\mathbf{w}\|^2 \leq \rho$, and $\mathbf{x}$ lies in the set $\mathcal{C}$ defined by the intersection of k ellipsoids:

$$\mathcal{C} = \left\{ \mathbf{x} : f_i(\mathbf{x}) \triangleq \mathbf{x}^T \mathbf{Q}_i \mathbf{x} + 2\mathbf{g}_i^T \mathbf{x} + d_i \leq 0, 1 \leq i \leq k \right\} \tag{2}$$

where $\mathbf{Q}_i \succeq 0$, $\mathbf{g}_i \in \mathcal{R}^m$ and $d_i \in \mathcal{R}$. To simplify notation, we present the results for the real case, however all the derivations hold true for complex-valued data as well. Combining the restrictions on $\mathbf{x}$ and $\mathbf{w}$, the feasible parameter set, which is the set of all possible values of $\mathbf{x}$, is given by

$$\mathcal{Q} = \left\{ \mathbf{x} : \mathbf{x} \in \mathcal{C}, \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \leq \rho \right\}. \tag{3}$$

In order to obtain strictly feasible optimization problems, we assume throughout that there is at least one point in the interior of $\mathcal{Q}$. In addition, we require that $\mathcal{Q}$ is compact. To this end, it is sufficient to assume that $\mathbf{A}^T \mathbf{A}$ is invertible.

Given the prior knowledge $x \in \mathcal{C}$, a popular estimation strategy is the CLS approach, in which the estimate is chosen to minimize the data error over the set $\mathcal{C}$. Thus, the CLS estimate, denoted $\hat{\mathbf{x}}_{\mathrm{CLS}}$, is the solution to

$$\min_{\mathbf{x} \in \mathcal{C}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2. \tag{4}$$

Clearly $\hat{\mathbf{x}}_{\mathrm{CLS}}$ is feasible, namely it satisfies the constraints defining our prior knowledge. However, the fact that it minimizes the data error over $\mathcal{C}$ does not mean that it leads to a small estimation error $\|\hat{\mathbf{x}} - \mathbf{x}\|$. In fact, the simulations in Section VII demonstrate that the resulting error can be quite large.

To design an estimator with small estimation error, we suggest minimizing the worst-case error over all feasible vectors. This is equivalent to finding the Chebyshev center of $\mathcal{Q}$:

$$\min_{\hat{\mathbf{x}}} \max_{\mathbf{x} \in \mathcal{Q}} \|\hat{\mathbf{x}} - \mathbf{x}\|^2. \tag{5}$$

To develop a geometrical interpretation of $\hat{\mathbf{x}}$, note that (5) can be written equivalently as

$$\min_{\hat{\mathbf{x}}, r} \left\{ r : \|\hat{\mathbf{x}} - \mathbf{x}\|^2 \leq r \text{ for all } \mathbf{x} \in \mathcal{Q} \right\}. \tag{6}$$

For a given r, the set of all values of $\mathbf{x}$ satisfying $\|\hat{\mathbf{x}} - \mathbf{x}\|^2 \leq r$ defines a ball with radius $\sqrt{r}$ and center $\hat{\mathbf{x}}$. Thus, the constraint in (6) is equivalent to the requirement that the ball defined by r and $\hat{\mathbf{x}}$ encloses the set $\mathcal{Q}$. Since the minimization is over the squared-radius r, it follows that the Chebyshev center is the center of the minimum radius ball enclosing $\mathcal{Q}$ and the squared radius of the ball is the optimal minimax value of (5). This is illustrated in Fig. 1 with the filled area being the intersection
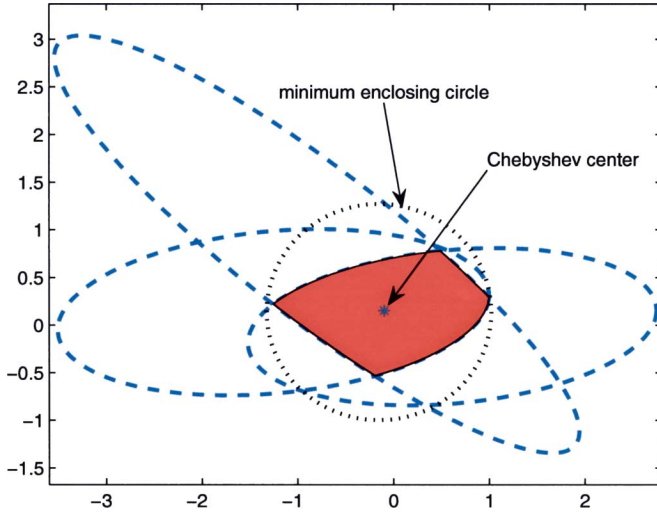
Fig. 1. Chebyshev center of the intersection of three ellipsoids.

of three ellipsoids. The dotted circle is the minimum inscribing circle of the intersection of the ellipsoids.

Computing the Chebyshev center (5) is a hard optimization problem. To better understand its intrinsic difficulty, note that the inner maximization is a non-convex quadratic optimization problem. Relying on strong duality results derived in the context of quadratic optimization [21], it was recently shown that despite the non-convexity of the problem, it can be solved efficiently over the complex domain when $\mathcal{Q}$ is the intersection of two ellipsoids. The same approach was then used over the real domain to develop an approximation of the Chebyshev center. Here, we extend these ideas to a more general quadratic constraint set. The importance of this generalization is that in many practical applications there are more than two constraints. For example, interval restrictions are popular in image processing in which the components of $\mathbf{x}$ represent individual pixel values which are limited to a fixed interval (e.g., [0,255]). A bound of the form $\ell_i \leq x_i \leq u_i$ can be represented by the ellipsoid $(x_i - \ell_i)(x_i - u_i) \leq 0$, or by the intersection of the two degenerate ellipsoids $x_i - u_i \leq 0$ and $\ell_i - x_i \leq 0$. These types of sets and the best way to represent them are discussed in Section VI. Another popular constraint is $\|\mathbf{L}\mathbf{x}\| \leq \eta$ for some $\eta > 0$, where $\mathbf{L}$ is the discretization of a first- or second-order differential operator that accounts for smoothness properties of $\mathbf{x}$ [1], [22].

III. THE RELAXED CHEBYSHEV CENTER

The RCC estimator, denoted $\hat{\mathbf{x}}_{\mathrm{RCC}}$, is obtained by replacing the non-convex inner maximization in (5) by its semidefinite relaxation, and then solving the resulting convex-concave minimax problem.

To develop $\hat{\mathbf{x}}_{\mathrm{RCC}}$, consider the inner maximization in (5):

$$\max_{\mathbf{x}} \left\{ \|\hat{\mathbf{x}} - \mathbf{x}\|^2 : f_i(\mathbf{x}) \leq 0, 0 \leq i \leq k \right\} \quad (7)$$

where $f_i(\mathbf{x}), 1 \leq i \leq k$ are defined by (2), and $f_0(\mathbf{x})$ is defined similarly with $\mathbf{Q}_0 = \mathbf{A}^T \mathbf{A}$, $\mathbf{g}_0 = -\mathbf{A}^T \mathbf{y}$, $d_0 = \|\mathbf{y}\|^2 - \rho$ so that $f_0(\mathbf{x}) = \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 - \rho$. Thus, the set $\mathcal{Q}$ can be written as

$$\mathcal{Q} = \{\mathbf{x} : f_i(\mathbf{x}) \leq 0, 0 \leq i \leq k\}. \quad (8)$$

Denoting $\Delta = \mathbf{x}\mathbf{x}^T$, (7) can be written equivalently as

$$\max_{(\Delta, \mathbf{x}) \in \mathcal{G}} \left\{ \|\hat{\mathbf{x}}\|^2 - 2\hat{\mathbf{x}}^T \mathbf{x} + \mathrm{Tr}(\Delta) \right\} \quad (9)$$

where

$$\mathcal{G} = \left\{ (\Delta, \mathbf{x}) : f_i(\Delta, \mathbf{x}) \leq 0, 0 \leq i \leq k, \Delta = \mathbf{x}\mathbf{x}^T \right\} \quad (10)$$

and we defined

$$f_i(\Delta, \mathbf{x}) = \mathrm{Tr}(\mathbf{Q}_i \Delta) + 2\mathbf{g}_i^T \mathbf{x} + d_i, \quad 0 \leq i \leq k. \quad (11)$$

The objective in (9) is concave (linear) in $(\Delta, \mathbf{x})$, but the set $\mathcal{G}$ is not convex. To obtain a relaxation of (9) we may replace $\mathcal{G}$ by the convex set

$$\mathcal{T} = \left\{ (\Delta, \mathbf{x}) : f_i(\Delta, \mathbf{x}) \leq 0, 0 \leq i \leq k, \Delta \succeq \mathbf{x}\mathbf{x}^T \right\}. \quad (12)$$

The RCC is the solution to the resulting minimax problem:

$$\min_{\hat{\mathbf{x}}} \max_{(\Delta, \mathbf{x}) \in \mathcal{T}} \left\{ \|\hat{\mathbf{x}}\|^2 - 2\hat{\mathbf{x}}^T \mathbf{x} + \mathrm{Tr}(\Delta) \right\}. \quad (13)$$

The objective in (13) is concave (linear) in Δ and $\mathbf{x}$ and convex in $\hat{\mathbf{x}}$. Furthermore, the set $\mathcal{T}$ is bounded. Therefore, we can replace the order of the minimization and maximization [23], resulting in the equivalent problem

$$\max_{(\Delta, \mathbf{x}) \in \mathcal{T}} \min_{\hat{\mathbf{x}}} \left\{ \|\hat{\mathbf{x}}\|^2 - 2\hat{\mathbf{x}}^T \mathbf{x} + \mathrm{Tr}(\Delta) \right\}. \quad (14)$$

The inner minimization is a simple quadratic problem, whose optimal value is $\hat{\mathbf{x}} = \mathbf{x}$. Thus, (14) reduces to

$$\max_{(\Delta, \mathbf{x}) \in \mathcal{T}} \left\{ -\|\mathbf{x}\|^2 + \mathrm{Tr}(\Delta) \right\} \quad (15)$$

which is a convex optimization problem with a concave objective and linear matrix inequality constraints. The RCC estimate is the $\mathbf{x}$-part of the solution to (15).

The RCC is not generally equal to the Chebyshev center of $\mathcal{Q}$. An exception is when $k = 1$ with the problem defined over the complex domain [20]. Since clearly $\mathcal{G} \subseteq \mathcal{T}$, we have that

$$\min_{\hat{\mathbf{x}}} \max_{\mathbf{x} \in \mathcal{Q}} \|\hat{\mathbf{x}} - \mathbf{x}\|^2$$
$$= \min_{\hat{\mathbf{x}}} \max_{(\Delta, \mathbf{x}) \in \mathcal{G}} \left\{ \|\hat{\mathbf{x}}\|^2 - 2\hat{\mathbf{x}}^T \mathbf{x} + \mathrm{Tr}(\Delta) \right\}$$
$$\leq \min_{\hat{\mathbf{x}}} \max_{(\Delta, \mathbf{x}) \in \mathcal{T}} \left\{ \|\hat{\mathbf{x}}\|^2 - 2\hat{\mathbf{x}}^T \mathbf{x} + \mathrm{Tr}(\Delta) \right\}. \quad (16)$$

Therefore, the RCC provides an upper bound on the optimal minimax value.

In Theorem III.1 below, we present an explicit representation of the RCC.

Theorem III.1: The RCC estimate, $\hat{\mathbf{x}}_{\mathrm{RCC}}$ which is the solution to (15), is given by

$$\hat{\mathbf{x}}_{\mathrm{RCC}} = -\left(\sum_{i=0}^{k} \alpha_i \mathbf{Q}_i \right)^{-1} \left(\sum_{i=0}^{k} \alpha_i \mathbf{g}_i \right) \quad (17)$$

where $(\alpha_1, \ldots, \alpha_k)$ is an optimal solution of the following convex optimization problem in $k + 1$ variables:

$$\min_{\alpha_i} \left\{ \left(\sum_{i=0}^{k} \alpha_i \mathbf{g}_i \right)^T \left(\sum_{i=0}^{k} \alpha_i \mathbf{Q}_i \right)^{-1} \left(\sum_{i=0}^{k} \alpha_i \mathbf{g}_i \right) - \sum_{i=0}^{k} d_i \alpha_i \right\}$$

$$\text{s.t.} \quad \sum_{i=0}^{k} \alpha_i \mathbf{Q}_i \succeq \mathbf{I},$$

$$\alpha_i \geq 0, \quad 0 \leq i \leq k. \tag{18}$$

Before proving the theorem, note that (18) can be cast as a semidefinite program (SDP) [24]:

$$\min_{\alpha_i} \left\{ t - \sum_{i=0}^{k} d_i \alpha_i \right\}$$

$$\text{s.t.} \quad \begin{pmatrix} \sum_{i=0}^{k} \alpha_i \mathbf{Q}_i & \sum_{i=0}^{k} \alpha_i \mathbf{g}_i \\ \sum_{i=0}^{k} \alpha_i \mathbf{g}_i^T & t \end{pmatrix} \succeq 0,$$

$$\sum_{i=0}^{k} \alpha_i \mathbf{Q}_i \succeq \mathbf{I},$$

$$\alpha_i \geq 0, \quad 0 \leq i \leq k. \tag{19}$$

This SDP can be solved by one of the many available SDP solvers such as SeDuMi [25] and SDPT3 [26].

Proof: To prove the theorem we show that (18) is the dual of (15). Since (15) is convex and strictly feasible (because we assumed that there is a point in the interior of $\mathcal{Q}$), its optimal value is equal to that of its dual problem. To compute the dual, we first form the Lagrangian:

$$\mathcal{L} = -\|\mathbf{x}\|^2 + \text{Tr}(\Delta) + \text{Tr}\left(\Pi(\Delta - \mathbf{x}\mathbf{x}^T) \right)$$
$$- \sum_{i=0}^{k} \alpha_i \left(\text{Tr}(\mathbf{Q}_i \Delta) + 2\mathbf{g}_i^T \mathbf{x} + d_i \right) \tag{20}$$

where $\alpha_i \geq 0$ and $\Pi \succeq 0$ are the dual variables. Differentiating $\mathcal{L}$ with respect to $\mathbf{x}$ and equating to 0,

$$\mathbf{x} = -(\mathbf{I} + \Pi)^{-1} \sum_{i=0}^{k} \alpha_i \mathbf{g}_i \tag{21}$$

where we used the fact that since $\Pi \succeq 0$, $\mathbf{I} + \Pi$ is invertible. The derivative with respect to Δ yields

$$\mathbf{I} + \Pi = \sum_{i=0}^{k} \alpha_i \mathbf{Q}_i. \tag{22}$$

Combining (21) and (22) yields (17).

Next we note that since $\Pi \succeq 0$, we must have from (22) that $\sum_{i=0}^{k} \alpha_i \mathbf{Q}_i \succeq \mathbf{I}$. Finally, substituting (21) and (22) into (20), we obtain the dual problem (18). ∎

For $k = 1$, the expression for the RCC reduces to the one obtained in [20]. We note that our derivation in Theorem III.1 for an arbitrary k is significantly simpler than the derivation in [20] for the special case $k = 1$. The main difference is that here we replace the inner maximization with its semidefinite

relaxation, while in [20], this maximization was replaced by its Lagrangian dual. These derivations are equivalent since the dual problem of the inner maximization problem is also the dual of the (convex) semidefinite relaxation [24]. Besides its simplicity, this approach can be used to show that the CLS may also be interpreted as a relaxation of the Chebyshev center. This is the topic of Section V.

IV. FEASIBILITY OF THE EXACT CHEBYSHEV CENTER AND THE RCC

The Chebyshev center and RCC approaches are designed to estimate $\mathbf{x}$ when we have the prior knowledge $\mathbf{x} \in \mathcal{Q}$ where $\mathcal{Q}$ is defined by (8) [or (3)]. Therefore, a desirable property of these estimates is that they should also reside in the set $\mathcal{Q}$, meaning they should be consistent with the prior information.

In Section IV-A, we use the definition of the exact Chebyshev center to establish its feasibility. We then prove in Section IV-B the feasibility of the RCC by using the representation (15).

A. Feasibility of the Chebyshev Center

Proposition IV.1: The solution $\hat{\mathbf{x}}$ of (5) is unique and feasible.
Proof: Problem (5) can be written as

$$\min_{\hat{\mathbf{x}}} \left\{ \|\hat{\mathbf{x}}\|^2 + \varphi(\hat{\mathbf{x}}) \right\} \tag{23}$$

where $\varphi(\hat{\mathbf{x}}) \equiv \max_{\mathbf{x} \in \mathcal{Q}} \{ -2\mathbf{x}^T \hat{\mathbf{x}} + \|\mathbf{x}\|^2 \}$. Since φ is a maximum of convex functions of $\hat{\mathbf{x}}$, it is convex. As a result, $\|\hat{\mathbf{x}}\|^2 + \varphi(\hat{\mathbf{x}})$ is strictly convex which readily implies [27] that (23) has a unique solution.

In order to prove the feasibility of the Chebyshev center, we will assume by contradiction that the optimal solution $\hat{\mathbf{x}}$ of (5) is infeasible, i.e., $\hat{\mathbf{x}} \notin \mathcal{Q}$. Let $\mathbf{y} = P_{\mathcal{Q}}(\hat{\mathbf{x}})$, where $P_{\mathcal{Q}}(\cdot)$ denotes the orthogonal projection operator onto the convex set $\mathcal{Q}$. By the projection theorem on convex sets [27], we have

$$(\hat{\mathbf{x}} - \mathbf{y})^T (\mathbf{x} - \mathbf{y}) \leq 0 \quad \text{for every } \mathbf{x} \in \mathcal{Q}. \tag{24}$$

Therefore, for every $\mathbf{x} \in \mathcal{Q}$,

$$\|\mathbf{y} - \mathbf{x}\|^2 < \|\mathbf{y} - \mathbf{x}\|^2 + \|\hat{\mathbf{x}} - \mathbf{y}\|^2$$
$$= \|\mathbf{y} - \mathbf{x} + \hat{\mathbf{x}} - \mathbf{y}\|^2 + 2(\mathbf{x} - \mathbf{y})^T (\hat{\mathbf{x}} - \mathbf{y})$$
$$\leq \|\hat{\mathbf{x}} - \mathbf{x}\|^2$$

where the first inequality follows from the fact that $\mathbf{y} \neq \hat{\mathbf{x}}$ and the last inequality is due to (24). Thus

$$\|\mathbf{y} - \mathbf{x}\|^2 < \|\hat{\mathbf{x}} - \mathbf{x}\|^2 \quad \text{for every } \mathbf{x} \in \mathcal{Q} \tag{25}$$

which using the compactness of $\mathcal{Q}$ implies that

$$\max_{\mathbf{x} \in \mathcal{Q}} \|\mathbf{y} - \mathbf{x}\|^2 < \max_{\mathbf{x} \in \mathcal{Q}} \|\hat{\mathbf{x}} - \mathbf{x}\|^2$$

contradicting the optimality of $\hat{\mathbf{x}}$. ∎

B. Feasibility of the RCC

Proposition IV.2: The RCC estimate $\hat{\mathbf{x}}_{\text{RCC}}$ of (13) is unique and feasible.

Proof: The uniqueness follows from similar arguments as in the proof of Proposition IV.1. The feasibility results from the observation that $\hat{\mathbf{x}}_{\mathrm{RCC}}$ can be expressed as the solution to (15).

Specifically, let $(\widehat{\Delta}, \hat{\mathbf{x}}_{\mathrm{RCC}})$ be a solution to (15). Since $(\widehat{\Delta}, \hat{\mathbf{x}}_{\mathrm{RCC}}) \in \mathcal{T}$, where $\mathcal{T}$ is defined by (12), we have that

$$\mathrm{Tr}(\mathbf{Q}_i \widehat{\Delta}) + 2\mathbf{g}_i^T \hat{\mathbf{x}}_{\mathrm{RCC}} + d_i \leq 0, \quad 0 \leq i \leq k. \quad (26)$$

In addition, $\widehat{\Delta} \succeq \hat{\mathbf{x}}_{\mathrm{RCC}} \hat{\mathbf{x}}_{\mathrm{RCC}}^T$. Since $\mathbf{Q}_i \succeq 0$, this, in turn, implies that

$$\begin{aligned} f_i(\hat{\mathbf{x}}_{\mathrm{RCC}}) &= \hat{\mathbf{x}}_{\mathrm{RCC}}^T \mathbf{Q}_i \hat{\mathbf{x}}_{\mathrm{RCC}} + 2\mathbf{g}_i^T \hat{\mathbf{x}}_{\mathrm{RCC}} + d_i \\ &= \mathrm{Tr}\left(\mathbf{Q}_i \hat{\mathbf{x}}_{\mathrm{RCC}} \hat{\mathbf{x}}_{\mathrm{RCC}}^T\right) + 2\mathbf{g}_i^T \hat{\mathbf{x}}_{\mathrm{RCC}} + d_i \\ &\leq \mathrm{Tr}(\mathbf{Q}_i \widehat{\Delta}) + 2\mathbf{g}_i^T \hat{\mathbf{x}}_{\mathrm{RCC}} + d_i \leq 0 \end{aligned} \quad (27)$$

where we used the fact that if $\mathbf{A} \succeq \mathbf{B}$ and $\mathbf{Q} \succeq 0$, then $\mathrm{Tr}(\mathbf{AQ}) \geq \mathrm{Tr}(\mathbf{BQ})$. We conclude that $f_i(\hat{\mathbf{x}}_{\mathrm{RCC}}) \leq 0$ for $0 \leq i \leq k$, which implies that $\hat{\mathbf{x}}_{\mathrm{RCC}} \in \mathcal{Q}$ with $\mathcal{Q}$ defined by (8). ∎

V. RELATION WITH THE CLS

Using a similar derivation to that of the RCC estimator we now show that the CLS estimate can also be viewed as a relaxation of the Chebyshev center. However, this relaxation is looser than that of the RCC meaning that the resulting bound on the minimax value is larger.

The RCC estimate was based on the prior information $\mathbf{x} \in \mathcal{C}$ with $\mathcal{C}$ given by (2), and the bounded error constraint $\|\mathbf{y} - \mathbf{Ax}\|^2 \leq \rho$. The CLS estimate, which is the solution to (4), tries to minimize the noise (or the data error) over $\mathbf{x} \in \mathcal{C}$.

To establish the relation with the Chebyshev center, we first note that $\mathbf{x} \in \mathcal{Q}$ is equivalent to $\mathbf{x} \in \mathcal{C}$ and $\|\mathbf{y} - \mathbf{Ax}\|^2 \leq \rho$. Now, the RCC was obtained by replacing $\mathbf{x} \in \mathcal{Q}$ by the equivalent set $(\mathbf{x}, \Delta) \in \mathcal{G}$ with $\mathcal{G}$ given by (10), and then relaxing the non-convex constraint in $\mathcal{G}$ to obtain the convex set $\mathcal{T}$ of (12). As we now show, the CLS can be viewed as a relaxed Chebyshev center where $\mathcal{G}$ is replaced by a different convex set $\mathcal{V}$, that is larger than $\mathcal{T}$. To obtain $\mathcal{V}$, note that $\mathcal{G}$ can be written as

$$\mathcal{G} = \big\{(\Delta, \mathbf{x}) : \mathbf{x} \in \mathcal{C}, \mathrm{Tr}(\mathbf{A}^T \mathbf{A} \Delta) - 2\mathbf{y}^T \mathbf{A}^T \mathbf{x} \\ + \|\mathbf{y}\|^2 - \rho \leq 0, \Delta = \mathbf{x}\mathbf{x}^T\big\}. \quad (28)$$

In this representation, we substituted $\Delta = \mathbf{x}\mathbf{x}^T$ only in the noise constraint, but not in the constraints $\mathbf{x} \in \mathcal{C}$. We now relax the non-convex equality and obtain the relaxed convex set

$$\mathcal{V} = \big\{(\Delta, \mathbf{x}) : \mathbf{x} \in \mathcal{C}, \mathrm{Tr}(\mathbf{A}^T \mathbf{A} \Delta) - 2\mathbf{y}^T \mathbf{A}^T \mathbf{x} \\ + \|\mathbf{y}\|^2 - \rho \leq 0, \Delta \succeq \mathbf{x}\mathbf{x}^T\big\}. \quad (29)$$

Evidently, the relaxation only effected the noise constraint and not those in $\mathcal{C}$. This results in a set $\mathcal{V}$ that includes the set $\mathcal{T}$ of (12). To see this, note that since $\mathbf{Q}_i \succeq 0$, for any $\Delta \succeq \mathbf{x}\mathbf{x}^T$, we have that

$$f_i(\mathbf{x}) \leq f_i(\Delta, \mathbf{x}) \quad (30)$$

where $f_i(\mathbf{x})$ and $f_i(\Delta, \mathbf{x})$ are defined by (2) and (11), respectively. Now, if $(\Delta, \mathbf{x}) \in \mathcal{T}$, then $\Delta \succeq \mathbf{x}\mathbf{x}^T$ and $f_i(\Delta, \mathbf{x}) \leq 0$ which implies from (46) that $f_i(\mathbf{x}) \leq 0$ so that $\mathbf{x} \in \mathcal{C}$.

Theorem V.1 establishes that $\hat{\mathbf{x}}_{\mathrm{CLS}}$ is the solution to the resulting minimax problem:

$$\min_{\hat{\mathbf{x}}} \max_{(\Delta, \mathbf{x}) \in \mathcal{V}} \left\{\|\hat{\mathbf{x}}\|^2 - 2\hat{\mathbf{x}}^T \mathbf{x} + \mathrm{Tr}(\Delta)\right\}. \quad (31)$$

Theorem V.1: The CLS estimate of (4) is the same as the relaxed Chebyshev center (31).

Proof: To prove the theorem we first follow similar steps as in the proof of the RCC, and replace the order of the minimization and the maximization. Noting that as before the optimal choice of $\hat{\mathbf{x}}$ is $\hat{\mathbf{x}} = \mathbf{x}$ we arrive at the equivalent problem

$$\max_{(\Delta, \mathbf{x}) \in \mathcal{V}} \left\{-\|\mathbf{x}\|^2 + \mathrm{Tr}(\Delta)\right\}. \quad (32)$$

We now show that (32) and the CLS problem (4) have the same solution $\mathbf{x}$. To see this, let $(\mathbf{x}_1, \Delta)$ be an optimal solution of (32) and let $\mathbf{x}_2$ be the solution of (4). Suppose to the contrary that $\mathbf{x}_1 \neq \mathbf{x}_2$. Since the objective function in the CLS problem is strictly convex (because $\mathbf{A}^T \mathbf{A}$ is invertible), $\mathbf{x}_2$ is the unique optimal solution, which implies that

$$r \equiv \|\mathbf{y} - \mathbf{A}\mathbf{x}_1\|^2 - \|\mathbf{y} - \mathbf{A}\mathbf{x}_2\|^2 > 0. \quad (33)$$

Define

$$\Delta' = \Delta + \mathbf{x}_2 \mathbf{x}_2^T - \mathbf{x}_1 \mathbf{x}_1^T + \frac{r}{\mathrm{Tr}(\mathbf{A}^T \mathbf{A})} \mathbf{I}. \quad (34)$$

It is easy to see that $(\Delta', \mathbf{x}_2) \in \mathcal{V}$. Indeed, $\mathbf{x}_2 \in \mathcal{C}$, and since $\Delta \succeq \mathbf{x}_1 \mathbf{x}_1^T$ we have immediately that $\Delta' \succ \mathbf{x}_2 \mathbf{x}_2^T$. Furthermore

$$\begin{aligned} \mathrm{Tr}(\Delta' \mathbf{A}^T \mathbf{A}) &= \mathrm{Tr}(\Delta \mathbf{A}^T \mathbf{A}) + r + \mathbf{x}_2^T \mathbf{A}^T \mathbf{A} \mathbf{x}_2 - \mathbf{x}_1^T \mathbf{A}^T \mathbf{A} \mathbf{x}_1 \\ &\leq \rho - \|\mathbf{y}\|^2 + 2\mathbf{y}^T \mathbf{A}^T \mathbf{x}_1 + r \\ &\quad + \mathbf{x}_2^T \mathbf{A}^T \mathbf{A} \mathbf{x}_2 - \mathbf{x}_1^T \mathbf{A}^T \mathbf{A} \mathbf{x}_1 \\ &< \rho - \|\mathbf{y}\|^2 + 2\mathbf{y}^T \mathbf{A}^T \mathbf{x}_2 \end{aligned} \quad (35)$$

where the last inequality follows from (33). Denote the objective value of (32) by $P(\Delta, \mathbf{x})$. Then

$$P(\Delta', \mathbf{x}_2) = P(\Delta, \mathbf{x}_1) + r \frac{n}{\mathrm{Tr}(\mathbf{A}^T \mathbf{A})} > P(\Delta, \mathbf{x}_1) \quad (36)$$

which contradicts the optimality of $(\Delta, \mathbf{x}_1)$. Therefore, we must have $\mathbf{x}_2 = \mathbf{x}_1$. ∎

We conclude that the RCC and CLS estimates can both be viewed as relaxations of the Chebyshev center problem (5). Since $\mathcal{T} \subseteq \mathcal{V}$, the CLS estimate is the solution of a looser relaxation than the one associated with the RCC. As we show in Section VII, the estimation error corresponding to the CLS estimate is typically much larger than that resulting from the RCC approach.

VI. MODELLING OF LINEAR CONSTRAINTS

As we noted in Section II, there are many signal processing examples in which there are interval constraints on the elements

of $\mathbf{x}$. In this section we address the question of how to best represent such restrictions.

Specifically, suppose that one of the constraints defining the set $\mathcal{Q}$ is a double-sided linear inequality of the form

$$\ell \leq \mathbf{a}^T \mathbf{x} \leq u \tag{37}$$

where $\ell < u$ and $\mathbf{a} \in \mathcal{R}^m$ is a nonzero vector. The constraint (37) can also be written in quadratic form as

$$(\mathbf{a}^T \mathbf{x} - \ell)(\mathbf{a}^T \mathbf{x} - u) \leq 0. \tag{38}$$

An important question that arises is whether or not the RCC depends on the specific representation of the set $\mathcal{Q}$. Clearly the CLS and Chebyshev center estimates are independent of the representation of $\mathcal{Q}$, as they depend only the set $\mathcal{Q}$ itself. However, the RCC estimate is more involved as it is a result of a relaxation of $\mathcal{Q}$, so that different characterizations may lead to different relaxed sets. In this section we show that indeed the RCC depends on the specific form of $\mathcal{Q}$ chosen. Furthermore, we prove that the quadratic representation (38) is better than the linear characterization (37) in the sense that the resulting minimax value is smaller.

Suppose that we have the following two representations of the same set:

$$\mathcal{Q}_1 = \left\{\mathbf{x} : f_i(\mathbf{x}) \leq 0, \mathbf{a}^T \mathbf{x} - u \leq 0, -\mathbf{a}^T \mathbf{x} + \ell \leq 0\right\}$$
$$\mathcal{Q}_2 = \left\{\mathbf{x} : f_i(\mathbf{x}) \leq 0, (\mathbf{a}^T \mathbf{x} - \ell)(\mathbf{a}^T \mathbf{x} - u) \leq 0\right\} \tag{39}$$

where $0 \leq i \leq p$, and $f_i(\mathbf{x}) = \mathbf{x}^T \mathbf{Q}_i \mathbf{x} + 2\mathbf{g}_i^T \mathbf{x} + d_i$ with $\mathbf{Q}_i \succeq 0$, $\mathbf{g}_i \in \mathcal{R}^m$ and $d_i \in \mathcal{R}$. The RCC estimates corresponding to these sets are obtained by forming the relaxed sets $\mathcal{T}_1$ and $\mathcal{T}_2$, as defined by (12), namely:

$$\mathcal{T}_1 = \left\{(\Delta, \mathbf{x}) : \text{Tr}(\mathbf{Q}_i \Delta) + 2\mathbf{g}_i^T \mathbf{x} + d_i \leq 0, \mathbf{a}^T \mathbf{x} - u \leq 0 \right.$$
$$\left. -\mathbf{a}^T \mathbf{x} + \ell \leq 0, \Delta \succeq \mathbf{x}\mathbf{x}^T\right\} \tag{40}$$
$$\mathcal{T}_2 = \left\{(\Delta, \mathbf{x}) : \text{Tr}(\mathbf{Q}_i \Delta) + 2\mathbf{g}_i^T \mathbf{x} + d_i \leq 0, \text{Tr}(\mathbf{a}\mathbf{a}^T \Delta) \right.$$
$$\left. -(u+\ell)\mathbf{a}^T \mathbf{x} + u\ell \leq 0, \Delta \succeq \mathbf{x}\mathbf{x}^T\right\} \tag{41}$$

and then solving (13). Denote by $\mathcal{V}_1$ and $\mathcal{V}_2$ the objective values resulting from $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively. As in (16), both values are upper bounds on the exact squared radius of the minimum enclosing ball of $\mathcal{Q} = \mathcal{Q}_1 = \mathcal{Q}_2$, i.e.,

$$\mathcal{V}_1 \geq \min_{\hat{\mathbf{x}}} \max_{\mathbf{x} \in \mathcal{Q}} \|\hat{\mathbf{x}} - \mathbf{x}\|^2, \quad \mathcal{V}_2 \geq \min_{\hat{\mathbf{x}}} \max_{\mathbf{x} \in \mathcal{Q}} \|\hat{\mathbf{x}} - \mathbf{x}\|^2. \tag{42}$$

This follows from the fact that $\mathcal{Q} \subseteq \mathcal{T}_1$ and $\mathcal{Q} \subseteq \mathcal{T}_2$. In the next theorem we show that the "quadratic" representation $\mathcal{Q}_2$ provides a tighter upper bound than the one given by the "linear" representation $\mathcal{Q}_1$.

Theorem VI.1: Let $\mathcal{V}_1$ and $\mathcal{V}_2$ denote the optimal values of (13) when $\mathcal{T} = \mathcal{T}_1$ and $\mathcal{T} = \mathcal{T}_2$, respectively. Then $\mathcal{V}_2 \leq \mathcal{V}_1$.

Proof: To prove the result, we will show that

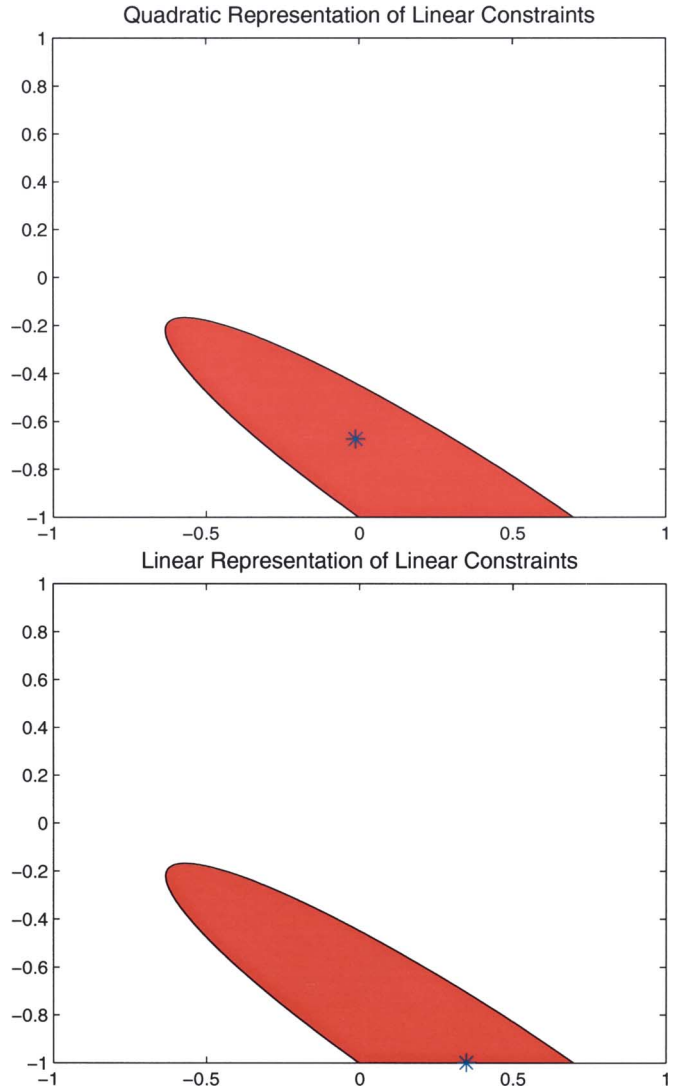$$\mathcal{T}_2 \subseteq \mathcal{T}_1 \tag{43}$$

Fig. 2. The RCC of the intersection of an ellipsoid with the box $[-1, 1] \times [-1, 1]$ using a quadratic representation (top) and a linear representation (bottom).

which implies

$$\mathcal{V}_1 = \min_{\hat{\mathbf{x}}} \max_{(\Delta, \mathbf{x}) \in \mathcal{T}_1} \left\{\|\hat{\mathbf{x}}\|^2 - 2\hat{\mathbf{x}}^T \mathbf{x} + \text{Tr}(\Delta)\right\}$$
$$\geq \min_{\hat{\mathbf{x}}} \max_{(\Delta, \mathbf{x}) \in \mathcal{T}_2} \left\{\|\hat{\mathbf{x}}\|^2 - 2\hat{\mathbf{x}}^T \mathbf{x} + \text{Tr}(\Delta)\right\} = \mathcal{V}_2. \tag{44}$$

To show (43), it is sufficient to prove that if $(\Delta, \mathbf{x})$ is in $\mathcal{T}_2$, then it is also in $\mathcal{T}_1$. Let $(\Delta, \mathbf{x}) \in \mathcal{T}_2$. Since the first $p + 1$ constraints defining $\mathcal{T}_2$ are the same as the first $p+1$ constraints defining $\mathcal{T}_1$, we only need to show that the last two constraints in $\mathcal{T}_1$ are satisfied, i.e.,

$$\mathbf{a}^T \mathbf{x} - u \leq 0, \quad -\mathbf{a}^T \mathbf{x} + \ell \leq 0. \tag{45}$$

Since $(\Delta, \mathbf{x}) \in \mathcal{T}_2$, we have

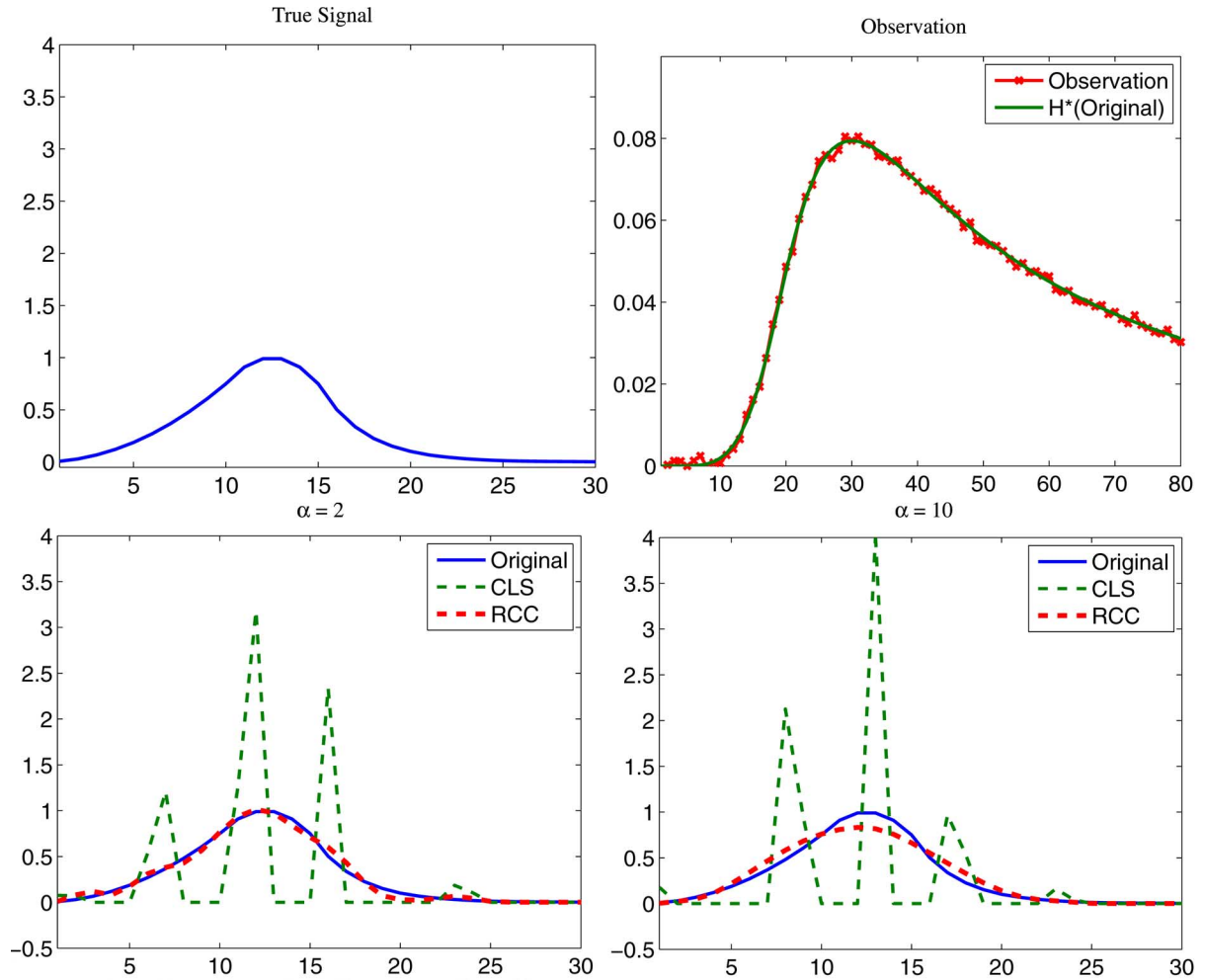$$\text{Tr}(\mathbf{a}\mathbf{a}^T \Delta) - (u + \ell)\mathbf{a}^T \mathbf{x} + u\ell \leq 0. \tag{46}$$

Fig. 3.   Comparison between the RCC and CLS estimates.

Combining (46) with the matrix inequality $\Delta \succeq \mathbf{x}\mathbf{x}^T$, it follows that

$$\mathbf{x}^T \mathbf{a}\mathbf{a}^T \mathbf{x} - (u + \ell)\mathbf{a}^T \mathbf{x} + u\ell \leq 0. \qquad (47)$$

But (47) is equivalent to $(\mathbf{a}^T \mathbf{x} - \ell)(\mathbf{a}^T \mathbf{x} - u) \leq 0$, which, in turn, is the same as (45). Therefore, $(\Delta, \mathbf{x}) \in \mathcal{T}_1$, proving the result.   ∎

A consequence of the theorem, is that in the presence of several double sided linear constraints, it is best to represent all of them as quadratic constraints.

The importance of the representation of linear constraints in the context of Lagrangian dual bounds was noted in[1] [28, Example 1] where the following two representations of the same optimization problem were considered:

$$\max \left\{ -x_1^2 + x_2^2 : x_2 = 1 \right\}$$
$$\max \left\{ -x_1^2 + x_2^2 : (x_2 - 1)^2 = 0 \right\}.$$

The Lagrangian dual bound of the first problem is ∞ (i.e., meaningless); the dual bound of the second problem is equal to 1

[1]We are indebted to H. Wolkowicz for refereeing us to this example.

which is the optimal value of the two (identical) problems. Theorem VI.1 provides another motivation to using the quadratic representation.

We have observed through numerical examples that the linear representation provides a bound on the squared radius of the minimum enclosing circle that is much worse than the quadratic formulation. A typical example can be seen in Fig. 2. The filled region describes the intersection of a randomly generated ellipsoid $\mathcal{E}$ with the box $[-1, 1] \times [-1, 1]$. The asterisk in the top picture is the RCC of the set $\mathcal{E} \cap ([-1, 1] \times [-1, 1])$ when the box constraints are modelled as $x_i^2 \leq 1$, $i = 1$, 2. The asterisk in the bottom picture is the RCC when the box constraints are represented as $-1 \leq x_i \leq 1$, $i = 1$, 2. Clearly, the RCC using the linear representation is far from the center of the filled region (actually, it is on the boundary of the area!). In contrast, the RCC corresponding to the quadratic representation seems like a good measure of the center of the set. The minimax value using the linear representation was approximately 37% higher than that resulting from the quadratic representation.

VII. EXAMPLES

To illustrate the effectiveness of the RCC approach in comparison with the CLS method, we consider two examples from the "Regularization Tools" [29].
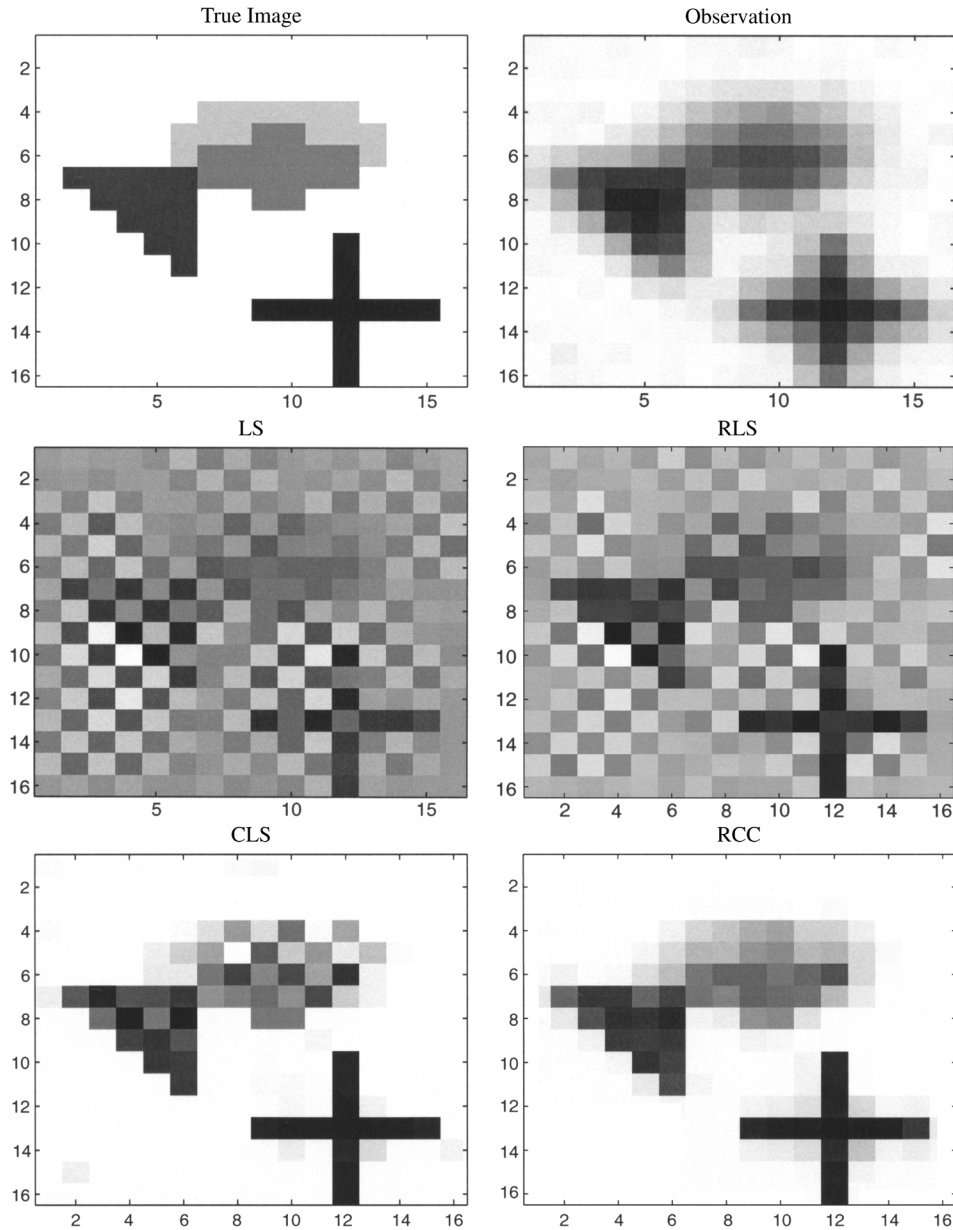
Fig. 4.   Comparison between different estimators.

A. Heat Equation

The first example is a discretization of the heat integral equation implemented in the function heat$(90, 1)$. In this case,

$$\mathbf{A}\mathbf{x} = \mathbf{g} \qquad (48)$$

where $\mathbf{A} \in \mathcal{R}^{90 \times 90}$ and $\mathbf{x}, \mathbf{g} \in \mathcal{R}^{90}$. The matrix $\mathbf{A}$ in this problem is extremely ill-conditioned. The true vector $\mathbf{x}$ is shown in Fig. 3 (True Signal) and resides in the set

$$\mathcal{C} = \{\mathbf{x} \in \mathcal{R}^{90} : \mathbf{x} \geq 0, \mathbf{x}^T \mathbf{e} \leq \eta\} \qquad (49)$$

where $\mathbf{e}$ is the vector of all ones. The observed vector is given by $\mathbf{y} = \mathbf{g} + \mathbf{w}$ where the elements of $\mathbf{w}$ are zero-mean, independent Gaussian random variables with standard deviation 0.001. Both $\mathbf{g}$ and $\mathbf{y}$ are shown in Fig. 3 (Observation).

To compute the RCC, we chose the set $\mathcal{Q}$ as

$$\mathcal{Q} = \left\{\|\mathbf{A}\mathbf{x} - \mathbf{b}\|^2 \leq \rho, \mathbf{x} \in \mathcal{C}\right\}$$

with $\rho = \alpha\|\mathbf{w}\|^2$ for some constant α, and $\eta = \alpha(\sum_{i=1}^{90} x_i)$. We then used the following quadratic representation of $\mathcal{C}$:

$$\left\{\mathbf{x} \in \mathcal{R}^{90} : x_i(x_i - \eta) \leq 0, (\mathbf{x}^T\mathbf{e})^2 \leq \eta^2, i = 1, \ldots, 90\right\}.$$

For comparison, we computed the CLS estimate which is the solution to $\min\{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 : \mathbf{x} \in \mathcal{C}\}$.

The results of the RCC and CLS estimates for $\alpha = 2$ and $\alpha = 10$ are shown at the bottom of Fig. 3. Evidently, the RCC approach leads to the best performance. The squared error of the RCC image $\|\hat{\mathbf{x}}_{\mathrm{RCC}} - \mathbf{x}\|^2$ was 196 and 55 times smaller than that of the CLS solution for $\alpha = 2$ and $\alpha = 10$ respectively.

The performance of both methods is better when $\alpha = 2$, as expected. However, it is interesting to note that even when

$\alpha = 10$, so that very loose prior information is used, the RCC results in very good performance.

B. Image Deblurring

As a second example, we consider a small image deblurring problem, again from the regularization tools.

In this problem the true value of $\mathbf{x}$ is of length 256 and is obtained by stacking the columns of the 16×16 image. The matrix $\mathbf{A}$ is of size 256×256 and represents an atmospheric turbulence blur originating from [30]; it is implemented in the function $\mathsf{blur}(16, 4, 0.8)$ (4 is the half bandwidth and 0.8 is the standard deviation associated with the corresponding point spread function). The image corresponding to $\mathbf{x}$ is shown at the top left corner of Fig. 4. The image is scaled so that each pixel is bounded below and above by 0 and 4, respectively.

The observed vector was generated by $\mathbf{Ax} + \mathbf{w}$ where each component of $\mathbf{w} \in \mathcal{R}^{256}$ was independently generated from a normal distribution with zero mean and standard deviation 0.05. The noisy image is shown in Fig. 4 (Observation). To estimate $\mathbf{x}$ we considered several approaches.

- **LS**. The LS estimator given by $\hat{\mathbf{x}}_{\mathrm{LS}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$. As can be seen in Fig. 4, the resulting image is very poor.
- **RLS**. The regularized LS solution which is the CLS corresponding to the norm constraint $\|\mathbf{x}\|^2 \le \rho$. In our experiments ρ was chosen to be $1.1\|\mathbf{x}\|^2$. As can be seen from Fig. 4, the RLS method also generates a poor image.
- **CLS**. Here we consider the CLS estimator when the bound on the pixels are taken into account. That is, the CLS image is the solution to the minimization problem

$$\min\left\{\|\mathbf{Ax} - \mathbf{b}\|^2 : x_i(x_i - 4) \le 0, 1 \le i \le 256\right\}.$$

Note that we use the quadratic representation of the constraints $0 \le x_i \le 4$. The image resulting from the CLS approach is much clearer than those resulting from the LS and RLS strategies.
- **RCC**. Finally, we compare the previous results with the RCC estimate corresponding to the set

$$\mathcal{Q} = \left\{\mathbf{x} : \|\mathbf{Ax} - \mathbf{b}\|^2 \le \rho, x_i(x_i - 4) \le 0, 1 \le i \le 256\right\}.$$

The upper bound on the squared norm of the noise vector was chosen to be $\rho = 1.3\|\mathbf{w}\|^2$. Evidently, the RCC results in the best image quality. The squared error of the RCC image $\|\hat{\mathbf{x}}_{\mathrm{RCC}} - \mathbf{x}\|^2$ was 33% smaller than the squared error of the CLS image $\|\hat{\mathbf{x}}_{\mathrm{CLS}} - \mathbf{x}\|^2$.

VIII. CONCLUSION

We introduced a new estimation strategy for estimating a deterministic parameter vector $\mathbf{x}$ in the linear regression model, when $\mathbf{x}$ is known to lie in a known parameter set $\mathcal{C}$. In our development, we considered the case in which $\mathcal{C}$ can be described by a set of quadratic inequality constraints. This includes bounded norm and interval (linear) restrictions. Instead of minimizing the data error, which results from a maximum likelihood approach, we suggested maximizing the worst-case estimation error over the given parameter set, which is equivalent to finding the Chebyshev center of the set. This allows control of the estimation error which is a direct measure of estimator quality. To

this end we add a constraint on the resulting data error which can be chosen to be proportional to the variance of the noise. Since the resulting problem is intractable, we suggested a relaxation based on semidefinite programming which can be solved efficiently. We then demonstrated via examples that the performance of the proposed RCC estimate can be much better than that of conventional least-squares based methods in terms of estimation error.

REFERENCES

[1] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion.* Philadelphia, PA: SIAM, 1998.

[2] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images—Matrices, Spectra and Filtering.* Philadelphia, PA: SIAM, 2006.

[3] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, Feb. 1970.

[4] D. W. Marquardt, "Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation," *Technometrics*, vol. 12, no. 3, pp. 592–612, Aug. 1970.

[5] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-Posed Problems.* Washington, DC: V.H. Winston, 1977.

[6] L. S. Mayer and T. A. Willke, "On biased estimation in linear models," *Technometrics*, vol. 15, pp. 497–508, Aug. 1973.

[7] Y. C. Eldar and A. V. Oppenheim, "Covariance shaping least-squares estimation," *IEEE Trans. Signal Process.*, vol. 51, no. 3, pp. 686–697, Mar. 2003.

[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.*, vol. 52, no. 8, pp. 2177–2188, Aug. 2004.

[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.*, vol. 53, no. 1, pp. 168–181, Jan. 2005.

[10] Z. Ben-Haim and Y. C. Eldar, "Blind minimax estimation," *IEEE Trans. Info. Theory*, vol. 53, no. 9, pp. 3145–3157, Sep. 2007.

[11] B. Efron, "Biased versus unbiased estimation," *Adv. Math.*, vol. 16, pp. 259–277, 1975.

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

[13] M. Milanese and G. Belforte, "Estimation theory and uncertainty intervals evaluation in presence of unknown but bounded errors: Linear families of models and estimators," *IEEE Trans. Autom. Control*, vol. 27, no. 2, pp. 408–414, 1982.

[14] M. Milanese and A. Vicino, "Optimal estimation theory for dynamic systems with set membership uncertainty: An overview," *Automatica J. IFAC*, vol. 27, no. 6, pp. 997–1009, 1991.

[15] J. P. Norton, "Identification and application of bounded parameter models," *Automatica*, vol. 23, pp. 497–507, 1987.

[16] D. Amir, "Chebyshev centers and uniform convexity," *Pacific J. Math.*, vol. 77, no. 1, pp. 1–6, 1978.

[17] J. F. Traub, G. Wasikowski, and H. Wozinakowski, *Information-Based Complexity.* New York: Academic, 1988.

[18] M. Milanese and R. Tempo, "Optimal algorithms theory for robust estimation and prediction," *IEEE Trans. Autom. Control*, vol. 30, no. 8, pp. 730–738, Aug. 1985.

[19] S. Xu, R. M. Freund, and J. Sun, "Solution methodologies for the smallest enclosing circle problem," *Comput. Optim. Appl.*, vol. 25, no. 1–3, pp. 283–292, 2003, A tribute to Elijah (Lucien) Polak.

[20] A. Beck and Y. C. Eldar, "Regularization in regression with bounded noise: A chebyshev center approach," *SIAM J. Matrix Anal. Appl.*, vol. 29, no. 2, pp. 606–625, 2007.

[21] A. Beck and Y. C. Eldar, "Strong duality in nonconvex quadratic optimization with two quadratic constraints," *SIAM J. Optim.*, vol. 17, no. 3, pp. 844–860, 2006.

[22] P. C. Hansen, "The use of the L-curve in the regularization of discrete ill-posed problems," *SIAM J. Sci. Stat. Comput.*, vol. 14, pp. 1487–1503, 1993.

[23] R. T. Rockafellar, *Convex Analysis.* Princeton, NJ: Princeton Univ. Press, 1970.

[24] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 40–95, Mar. 1996.

[25] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optim. Methods Softw.*, vol. 11–12, pp. 625–653, 1999.

[26] K. C. Toh, M. J. Todd, and R. H. Tütüncü, "SDPT3—A MATLAB software package for semidefinite programming, version 1.3," *Optim. Methods Softw.*, vol. 11/12, no. 1–4, pp. 545–581, 1999, Interior point methods.

[27] D. P. Bertsekas, *Nonlinear Programming*, 2nd ed. Belmont, MA: Athena Scientific, 1999.

[28] S. Poljak, F. Rendl, and H. Wolkowicz, "A recipe for semidefinite relaxation for (0,1)-quadratic programming," *J. Global Optim.*, vol. 7, no. 1, pp. 51–73, 1995.

[29] P. C. Hansen, "Regularization tools, a Matlab package for analysis of discrete regularization problems," *Numer. Algorithms*, vol. 6, pp. 1–35, 1994.

[30] M. Hanke and P. C. Hansen, "Regularization methods for large-scale problems," *Surveys Math. Indust.*, vol. 3, no. 4, pp. 253–315, 1993.

Yonina C. Eldar (S'98–M'02–SM'07) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, in 2001.

From January 2002 to July 2002, she was a Postdoctoral Fellow at the Digital Signal Processing Group at MIT. She is currently an Associate Professor in the Department of Electrical Engineering at the Technion—Israel Institute of Technology, Haifa, Israel. She is also a Research Affiliate with the Research Laboratory of Electronics at MIT. Her research interests are in the general areas of signal processing, statistical signal processing, and computational biology.

Dr. Eldar was in the program for outstanding students at TAU from 1992 to 1996. In 1998, she held the Rosenblith Fellowship for study in Electrical Engineering at MIT, and in 2000, she held an IBM Research Fellowship. From 2002 to 2005, she was a Horev Fellow of the Leaders in Science and Technology program at the Technion and an Alon Fellow. In 2004, she was awarded the Wolf Foundation Krill Prize for Excellence in Scientific Research, in 2005 the Andre and Bella Meyer Lectureship, and in 2007 the Henry Taub Prize for Excellence in Research. She is a member of the IEEE Signal Processing Theory and Methods Technical Committee and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the EURASIP Journal of Signal Processing, and she is on the Editorial Board of Foundations and Trends in Signal Processing.

Amir Beck was born in Israel in 1975. He received the B.Sc. degree in pure mathematics (*cum laude*) in 1991, the M.Sc. degree in operations research (*suma cum laude*), and the Ph.D. degree in operations research—all from Tel Aviv University (TAU), Tel Aviv, Israel.

From 2003 to 2005, he was a Postdoctoral Fellow at the Minerva Optimization Center, Technion, Haifa, Israel. He is currently a Senior Lecturer in the Department of Industrial Engineering at the Technion—Israel Institute of Technology, Haifa, Israel. His research interests are in continuous optimization, in particular, large-scale problems, conic and robust optimization as well as nonconvex quadratic optimization: theory, algorithms, and applications in signal processing and communication.

Marc Teboulle received the D.Sc. degree from the Technion—Israel Institute of Technology, Haifa, in 1985.

From 1999 to 2002, he served as Chairman of the Department of Statistics and Operations Research at the School of Mathematical Sciences of Tel-Aviv University, Tel-Aviv, Israel. He is currently a Professor at the School of Mathematical Sciences of Tel-Aviv University, Tel-Aviv, Israel. He has held a position of applied mathematician at the Israel Aircraft Industries, and academic appointments at Dalhousie University, Halifax, Canada, and the University of Maryland, Baltimore. He has also held visiting appointments at several institutions, including IMPA, Rio de Janeiro; the University of Montpellier, France; the University of Paris, France; the University of Lyon, France; the University of Michigan, Ann Arbor; and the University of Texas at Austin. His research interests are in the area of continuous optimization, including theory, algorithmic analysis, and its applications. He has published numerous papers and two books and has given invited lectures at many international conferences. His research has been supported by various funding agencies, including the National Science Foundation, the French-Israeli Ministry of Sciences, the Bi-National Israel–United States Science Foundation, and the Israel Science Foundation. He currently serves on the editorial board of Mathematics of Operations Research, and the European Series in Applied and Industrial Mathematics, Control, Optimisation and Calculus of Variations.