BOUNDED ERROR ESTIMATION: A CHEBYSHEV CENTER APPROACH

Yonina C. Eldar, Amir Beck

Technion–Israel Institute of Technology
Haifa, 32000, Israel
{yonina@ee, becka@ie}.technion.ac.il

Marc Teboulle

Tel Aviv University
Tel Aviv 69978, Israel
{teboulle@post.tau.ac.il}

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 then demonstrate through simulations that the RCC can significantly improve the estimation error over the conventional constrained least-squares method.

Index Terms— Estimation, regression, minimax.

1. INTRODUCTION

A broad range of estimation problems 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

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

where $\mathbf{A}$ is a known matrix and $\mathbf{w}$ is an unknown perturbation.

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 $\mathbf{y}$. Although this approach is deterministic in nature, 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 $\hat{\mathbf{x}} - \mathbf{x}$. Thus, many attempts have been made to develop estimators that may be biased but closer to $\mathbf{x}$ in some statistical sense [1–4].

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 [1]. 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 $\mathcal{C}$ [5]. 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, problems resulting from quantization). A common estimation technique in these settings is the bounded error approach, or set-membership estimation [6, 7]. This strategy is designed to deal with bounded noise, and prior information of the form $\mathbf{x} \in \mathcal{C}$ for some set $\mathcal{C}$.

Here 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 our objective is to choose $\hat{\mathbf{x}}$ to be close to $\mathbf{x}$ in the squared error sense, 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 2, 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 [8]. In Section 4 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 [9], and when the set is finite [10]. Recently, we considered this approach for $\mathcal{C}$ given by an ellipsoid [11]. 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 [12]. 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.

In this paper 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. Furthermore, our development of the RCC estimate in this paper is different than that presented for the single ellipsoid case in [11]. Here we use the fact that the RCC can be cast as a solution to a convex-concave saddle point program.

Omitted proofs, and further details on the RCC and its relation to the CLS can be found in [13].

The paper is organized as follows. In Section 2 we discuss the geometrical properties of the Chebyshev center. We then develop in Section 3 the RCC using a simpler method than that in [11]. In Section 4 we demonstrate via examples that the RCC can dramatically reduce the estimation error with respect to the CLS method.

2. 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 identity matrix is denoted by $\mathbf{I}$, and $\mathbf{A}^T$ is the transpose of $\mathbf{A}$. Given two matrices $\mathbf{A}$

978-1-4244-1714-8/07/$25.00

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 (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:

$$\mathcal{C} = \{\mathbf{x} : f_i(\mathbf{x}) \stackrel{\triangle}{=} \mathbf{x}^T\mathbf{Q}_i\mathbf{x} + 2\mathbf{g}_i^T\mathbf{x} + d_i \leq 0, 1 \leq i \leq k\}, \quad (2)$$

where $\mathbf{Q}_i \succeq \mathbf{0}, \mathbf{g}_i \in \mathcal{R}^m$ 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

$$\mathcal{Q} = \{\mathbf{x} : \mathbf{x} \in \mathcal{C}, \|\mathbf{y} - \mathbf{A}\mathbf{x}\|^2 \leq \rho\}. \quad (3)$$

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 $\mathbf{x} \in \mathcal{C}$, 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

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

Note that the fact that $\hat{\mathbf{x}}_{\mathrm{CLS}}$ 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 4 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 $\mathcal{Q}$:

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

The minimax problem of (5) can be written equivalently as

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

For a given r, the set of vectors $\mathbf{x}$ 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 $\mathcal{Q}$. Therefore, 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 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 the 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 [12], 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 2 ellipsoids. The same approach was then used over the reals to develop an approximation of the Chebyshev center. Here we extend these ideas to a more general quadratic constraint set. The importance of this extension is that in many practical applications there are more than 2 constraints. For example, interval restrictions are popular in image processing in which the components of $\mathbf{x}$ represent individual

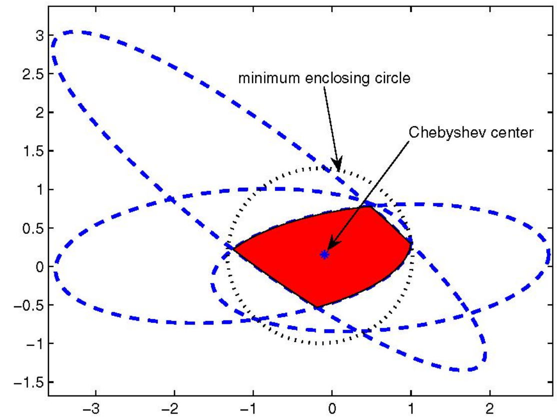

Fig. 1. The Chebyshev center of the intersection of three ellipsoids.

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 $(x_i - \ell_i)(x_i - u_i) \leq 0$. Another popular constraint is $\|\mathbf{L}\mathbf{x}\| \leq \eta$ for some $\eta > 0$ where $\mathbf{L}$ is the discretization of a differential operator that accounts for smoothness properties of $\mathbf{x}$ [14].

3. 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):

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

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

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

where

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

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

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

The RCC is the solution to the resulting minimax problem:

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

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 [15], resulting in the equivalent problem

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

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

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

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).

In Theorem 3.1 below we present an explicit representation of the RCC. The proof can be found in [13].

Theorem 3.1. *The RCC estimate, $\hat{\mathbf{x}}_{\mathrm{RCC}}$ which is the solution to (15), is given by*

$$\hat{\mathbf{x}}_{\mathrm{RCC}} = -\left(\sum_{i=0}^k \alpha_i \mathbf{Q}_i\right)^{-1}\left(\sum_{i=0}^k \alpha_i \mathbf{g}_i\right), \tag{16}$$

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

$$\begin{array}{ll} \min_{\alpha_i} & \left\{(\sum_{i=0}^k \alpha_i\mathbf{g}_i)^T(\sum_{i=0}^k \alpha_i\mathbf{Q}_i)^{-1}(\sum_{i=0}^k \alpha_i\mathbf{g}_i) \right. \\ & \qquad\qquad \left. - \sum_{i=0}^k d_i\alpha_i\right\} \\ \textit{s.t.} & \sum_{i=0}^k \alpha_i\mathbf{Q}_i \succeq \mathbf{I}, \\ & \alpha_i \geq 0, \quad 0 \leq i \leq k. \end{array} \tag{17}$$

For $k = 1$, the expression for the RCC reduces to the one obtained in [11]. We note that our derivation in Theorem 3.1 for an arbitrary k is significantly simpler than the derivation in [11] for the special case $k = 1$. The main difference is that here we replace the inner maximization with its semidefinite relaxation, while in [11], 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 [16].

4. EXAMPLES

To illustrate the effectiveness of the RCC approach in comparison with the CLS method, we consider two examples from the "Regularization Tools" [17].

4.1. Heat Equation

The first example is a discretization of the heat integral equation implemented in the function heat(90,1). In this case, $\mathbf{A}\mathbf{x} = \mathbf{g}$, 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. 4.1 (True Signal) and resides in the set

$$\mathcal{C} = \{\mathbf{x} \in \mathcal{R}^{90} : \mathbf{x} \geq \mathbf{0}, \mathbf{x}^T\mathbf{e} \leq \eta\}, \tag{18}$$

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. 4.1 (Observation).

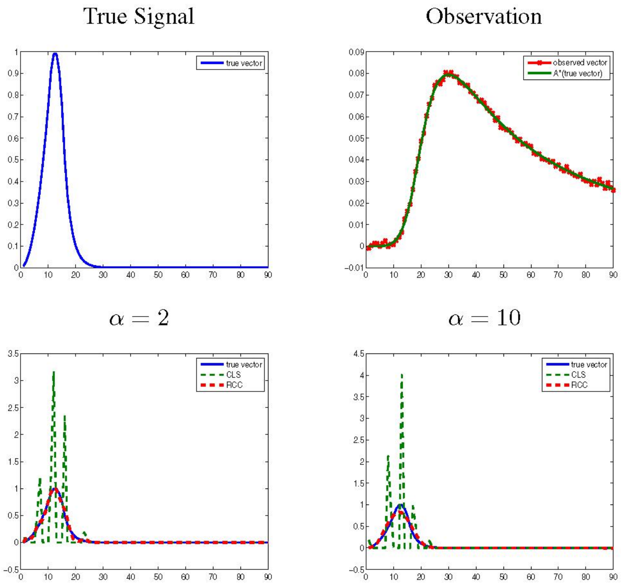

Fig. 2. Comparison between the RCC and CLS estimates.

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

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

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 $\mathcal{C}$:

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

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. 4.1. 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.

4.2. 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 [18]; it is implemented in the function 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. 3. The image is scaled so that each pixel is bounded below and above by 0 and 4 respectively.

The observed vector was generated by $\mathbf{A}\mathbf{x}+\mathbf{w}$ 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. 3 (Observation). To estimate $\mathbf{x}$ we considered several approaches:

- **LS.** The LS estimator given by $\hat{\mathbf{x}}_{\text{LS}} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}$. As can be seen in Fig. 3, the resulting image is very poor.
- **RLS.** The regularized LS solution which is the CLS corresponding to the norm constraint $\|\mathbf{x}\|^2 \leq \rho$. In our experiments ρ was chosen to be $1.1\|\mathbf{x}\|^2$. As can be seen from Fig. 3, 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

  $$\min\{\|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 : x_i(x_i - 4) \leq 0, 1 \leq i \leq 256\}.$$

  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

  $$\mathcal{Q} = \{\mathbf{x} : \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 \leq \rho, x_i(x_i - 4) \leq 0, 1 \leq i \leq 256\}.$$

  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 $\|\hat{\mathbf{x}}_{\text{RCC}} - \mathbf{x}\|^2$ was 33% smaller than the squared error of the CLS image $\|\hat{\mathbf{x}}_{\text{CLS}} - \mathbf{x}\|^2$.

5. REFERENCES

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

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

[3] Y. C. Eldar, A. Ben-Tal, and A. Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Trans. Signal Processing*, vol. 53, pp. 168–181, Jan. 2005.

[4] Z. Ben-Haim and Y. C. Eldar, "Blind minimax estimation," to appear in *IEEE Trans. Info. Theory.*

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

[6] 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.

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

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

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

[10] 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.

[11] 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.

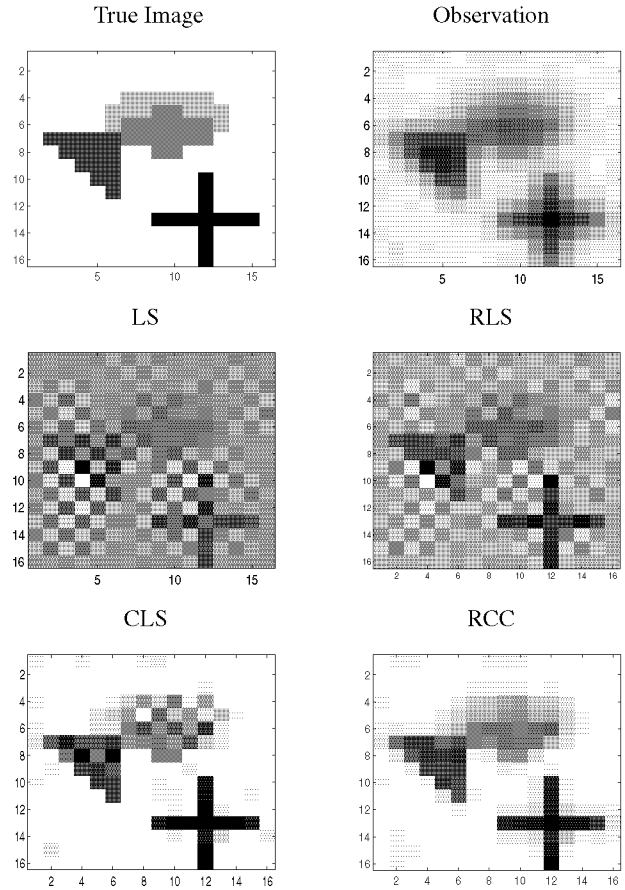

Fig. 3. Comparison between different estimators.

[12] 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.

[13] Y. C. Eldar, A. Beck, and M. Teboulle, "A minimax Chebyshev estimator for bounded error estimation," to appear in *IEEE Trans. Signal Processing*.

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

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

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

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

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