STATISTICAL ASPECTS OF THE CHEBYSHEV CENTER ESTIMATE

Yonina C. Eldar

Technion–Israel Institute of Technology, Haifa, Israel
Email: yonina@ee.technion.ac.il.

ABSTRACT

We treat the linear regression problem of estimating $\boldsymbol{\theta}$ from noisy observations where the norm of $\boldsymbol{\theta}$ is bounded. Instead of using the constrained least-squares approach which minimizes the data error over bounded norm vectors, we explore the use of the Chebyshev-center estimate (CC), that is aimed at minimizing the worst-case squared-error over all bounded-norm vectors $\boldsymbol{\theta}$ and bounded noise. We derive an explicit expression for the CC estimate and explore some of its statistical properties. In particular, we show that it can be viewed as a generalized Bayesian estimate where both the parameter vector and the noise have hierarchial Gaussian priors.

***Index Terms*—** Regression, minimax, constrained least-squares.

1. INTRODUCTION

Over the past 50 years, a multitude of estimators have been proposed for the linear regression model $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{e}$, with the aim of improving the performance of the conventional least-squares (LS) method. Stein [1] was the first to show that when $\mathbf{A} = \mathbf{I}$ and the noise $\mathbf{e}$ is Gaussian, the LS can be improved in terms of mean-squared error (MSE) for all parameter vectors $\boldsymbol{\theta}$. Since then, a broad variety of approaches have been developed for this problem.

Due to the fact that the LS estimate is minimax, no alternative strategy can uniformly improve its MSE by a large amount, for all $\boldsymbol{\theta}$. One way to significantly improve LS is to incorporate constraints on $\boldsymbol{\theta}$. A popular restriction is that the norm of $\boldsymbol{\theta}$ is bounded [3]. From a Bayesian perspective, this can be interpreted as prior information on the variance of $\boldsymbol{\theta}$. The typical estimation strategy in this context is the constrained LS (CLS) technique in which the data error $\|\mathbf{y} - \mathbf{A}\boldsymbol{\theta}\|$ is minimized subject to the prior constraints on $\boldsymbol{\theta}$. However, this method does not deal directly with the estimation error $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|$. Consequently, the resulting estimate $\hat{\boldsymbol{\theta}}$ may be far from $\boldsymbol{\theta}$.

In some scenarios, the distribution of the noise $\mathbf{e}$ may not be known, or the noise may not be random. A common estimation technique in these settings is the bounded error approach, also referred to as set-membership estimation [4]. This strategy is designed to deal with bounded noise, and prior information on $\boldsymbol{\theta}$.

In this paper, we adopt the bounded error methodology and assume that the noise is norm-bounded $\|\mathbf{e}\|^2 \leq \rho$. The estimator we develop can also be used when $\mathbf{e}$ is random, for example by choosing ρ proportional to its variance. In Section 4 we explore several choices of ρ for the case in which $\mathbf{e}$ is Gaussian. We further suppose that $\|\boldsymbol{\theta}\|^2 \leq \eta$. In recent work, we proposed a minimax estimation strategy for this problem, which is aimed at designing an estimate $\hat{\boldsymbol{\theta}}$ that minimizes the worst-case estimation error $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2$ over all feasible $\boldsymbol{\theta}$ [5,6]. As we show in Section 2, the proposed 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 (CC) approach.

Finding the CC of a set is typically an intractable problem. In our previous work [5,6] we considered this approach for the general linear model $\mathbf{y} = \mathbf{A}\boldsymbol{\theta} + \mathbf{e}$ where $\boldsymbol{\theta}$ lies in an intersection of ellipsoids. To solve the problem we suggested an approximation based on Lagrange duality and semidefinite relaxation. We then showed through numerical simulations that this technique outperforms LS and Tikhonov with respect to the estimation error.

In this paper we focus on the white-noise model $\mathbf{y} = \boldsymbol{\theta} + \mathbf{e}$ and consider some statistical aspects of the CC strategy. We begin, in Section 2, by deriving an exact closed form solution for the CC estimate under the constraints $\|\boldsymbol{\theta}\|^2 \leq \eta$ and $\|\mathbf{e}\|^2 \leq \rho$. As we show, the CC shrinks the observations $\mathbf{y}$ towards the center of the prior constrain set. This is in contrast with the CLS technique which is either equal to LS, or lies on the boundary of the set. It is well known that moving away from the boundary can improve the MSE [7].

In Section 3 we develop a Bayesian interpretation of the CC method. Specifically we show that it can be viewed as a minimum MSE (MMSE) estimate assuming that both $\boldsymbol{\theta}$ and $\mathbf{e}$ are zero-mean random vectors with hierarchial Gaussian priors, whose covariance matrices are also random. In addition, we exploit the fact that the CC estimate depends only on the difference in the norm bounds $\rho - \eta$ and not on the values themselves. We model this in a Bayesian setting by presuming a strong prior only on the difference of the variances of $\boldsymbol{\theta}$ and $\mathbf{e}$ while a harmonic prior is assumed on the inverse sum of the variances. This interpretation sheds further insight into the properties of the CC strategy and further motivates its use.

2. THE CHEBYSHEV CENTER

2.1. Problem Formulation

We treat the problem of estimating a deterministic vector $\boldsymbol{\theta} \in \mathcal{R}^n$ from observations $\mathbf{y} \in \mathcal{R}^n$ which are related through

$$\mathbf{y} = \boldsymbol{\theta} + \mathbf{e}. \tag{1}$$

Here $\mathbf{e}$ is a perturbation vector with bounded norm $\|\mathbf{e}\|^2 \leq \rho$, and $\|\boldsymbol{\theta}\|^2 \leq \eta$. Combining the restrictions on $\boldsymbol{\theta}$ and $\mathbf{e}$, the feasible parameter set, which is the set of all possible values of $\boldsymbol{\theta}$, is given by

$$\mathcal{Q} = \{\boldsymbol{\theta} : \|\mathbf{y} - \boldsymbol{\theta}\|^2 \leq \rho, \|\boldsymbol{\theta}\|^2 \leq \eta\}. \tag{2}$$

Given that $\boldsymbol{\theta} \in \mathcal{Q}$, a popular estimation strategy is the CLS method in which the data error is minimized over $\mathcal{Q}$:

$$\min_{\|\boldsymbol{\theta}\|^2 \leq \eta} \|\mathbf{y} - \boldsymbol{\theta}\|^2. \tag{3}$$

This work was supported in part by the NEWCOM++ Network of Excellence funded through the EC 7th Framework Programme.

The resulting estimate is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{CLS}} = \begin{cases} \mathbf{y}, & \|\mathbf{y}\|^2 \le \eta; \\ \sqrt{\frac{\eta}{\mathbf{y}^T\mathbf{y}}}\mathbf{y}, & \|\mathbf{y}\|^2 \ge \eta. \end{cases} \tag{4}$$

The fact that $\hat{\boldsymbol{\theta}}_{\mathrm{CLS}}$ minimizes the data error does not mean it leads to a small estimation error $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|$. In fact, when $\|\mathbf{y}\|^2 \ge \eta$, $\hat{\boldsymbol{\theta}}_{\mathrm{CLS}}$ lies on the boundary of $\mathcal{Q}$. It is well known that the MSE of such an estimate can be improved by moving away from the boundary [7]. Consequently, we would like an estimate inside $\mathcal{Q}$.

Since we are not assuming a prior distribution on $\boldsymbol{\theta}$, the MSE of any estimate $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, defined by $E\{\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2\}$, will depend in general on $\boldsymbol{\theta}$ and therefore cannot be minimized. Furthermore, directly computing the MSE of a nonlinear estimate $\hat{\boldsymbol{\theta}}$ is typically difficult. Thus, instead, we suggest seeking an estimate that minimizes the worst-case squared error over all feasible vectors. This is equivalent to finding the CC of $\mathcal{Q}$:

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

In developing the CC we explicitly assume that $\mathcal{Q}$ is non empty.

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

$$\min_{\hat{\boldsymbol{\theta}}, r} \{r : \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \le r \text{ for all } \boldsymbol{\theta} \in \mathcal{Q}\}. \tag{6}$$

The set of all values of $\boldsymbol{\theta}$ satisfying $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|^2 \le r$ defines a ball with radius $\sqrt{r}$ and center $\hat{\boldsymbol{\theta}}$. Thus, the constraint in (6) is equivalent to the requirement that the ball defined by r and $\hat{\boldsymbol{\theta}}$ encloses the set $\mathcal{Q}$. Since the minimization is over the squared-radius r, it follows that the CC is the center of the minimum radius ball enclosing $\mathcal{Q}$. This is illustrated in Fig. 1, taken from [5], with the filled area being the intersection of two ellipsoids. The dotted circle is the minimum inscribing circle of the intersection. Evidently, in contrast with the CLS method, the CC will lie in the center of the set.

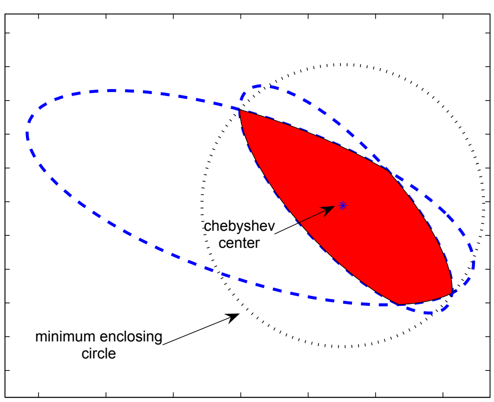

Fig. 1. The CC of the intersection of two ellipsoids.

2.2. Derivation of the CC Estimate

When $n = 1$ the set $\mathcal{Q}$ defined by (2) is just an interval in $\mathcal{R}$. Clearly the CC of the interval is its mid point. Thus, in the remainder of the paper we assume that $n \ge 2$. In this case, [8, Thm. 5.2] implies that the CC estimate for our model is the solution of

$$\begin{array}{ll} \max_{\boldsymbol{\theta}, t, \lambda_i} & t + \|\boldsymbol{\theta}\|^2 \\ \text{s.t.} & \begin{pmatrix} (\lambda_1 + \lambda_2 - 1)\mathbf{I} & \boldsymbol{\theta} - \lambda_1 \mathbf{y} \\ \boldsymbol{\theta}^T - \lambda_1 \mathbf{y}^T & t - \lambda_1(\rho - \mathbf{y}^T\mathbf{y}) - \lambda_2 \eta \end{pmatrix} \succeq 0, \\ & \lambda_1 \ge 0,\ \lambda_2 \ge 0, \end{array} \tag{7}$$

where the notation $\mathbf{A} \succeq 0$ means that $\mathbf{A}$ is nonnegative definite. Using convex analysis tools, it can be shown that the dual to (7) is

$$\begin{array}{ll} \max_{\boldsymbol{\theta}, t} & t \\ \text{s.t.} & \|\mathbf{y} - \boldsymbol{\theta}\|^2 + t \le \rho \\ & \|\boldsymbol{\theta}\|^2 + t \le \eta. \end{array} \tag{8}$$

Therefore, the CC tries to move the solution towards the center of the constraints: each constraint is satisfied with a margin of t and the aim is to maximize this gap. Clearly, this will move the solution away from the boundary of the set (unless the optimal t is $t = 0$).

Interestingly, the CLS can be written in a similar form to (8). Namely, it can be determined as the solution of

$$\begin{array}{ll} \max_{\boldsymbol{\theta}, t} & t \\ \text{s.t.} & \|\mathbf{y} - \boldsymbol{\theta}\|^2 + t \le \rho \\ & \|\boldsymbol{\theta}\|^2 \le \eta. \end{array} \tag{9}$$

Comparing (8) and (9) highlights the fundamental difference between the two strategies: In the CLS the margin is only on the data error constraint while in the CC formulation both restrictions are satisfied with a gap.

An important observation from (10), which will be instrumental in developing the Bayesian interpretation of the CC approach, is that the estimate depends only on the difference $\rho - \eta$ and not on ρ and η separately. This can be seen by substituting $t' = t - \rho$ into (8).

Problem (8) is a simple convex optimization problem and therefore can be solved using duality theory which leads to the solution

$$\hat{\boldsymbol{\theta}}_{\mathrm{CC}} = \frac{1}{2}\left(1 - \frac{\gamma}{\mathbf{y}^T\mathbf{y}}\right)_{[0,2]} \mathbf{y}. \tag{10}$$

Here $\gamma = \rho - \eta$, and we used the notation

$$x_{[a,b]} = \begin{cases} x, & a \le x \le b \\ a, & x \le a \\ b, & x \ge b. \end{cases} \tag{11}$$

It is interesting to note that $\hat{\boldsymbol{\theta}}_{\mathrm{CC}}$ has a similar form to the James-Stein (JS) estimate [2]. An important difference is the factor of $1/2$ that appears in (10). A Bayesian interpretation of $\hat{\boldsymbol{\theta}}_{\mathrm{CC}}$, which will also explain this scaling factor, is developed in the next section.

3. BAYESIAN INTERPRETATION

Throughout this section we consider the unrestricted CC estimate

$$\hat{\boldsymbol{\theta}}_{\mathrm{UCC}} = \frac{1}{2}\left(1 - \frac{\rho - \eta}{\mathbf{y}^T\mathbf{y}}\right)\mathbf{y}. \tag{12}$$

3.1. Empirical Bayes Interpretation

We first show that $\hat{\boldsymbol{\theta}}_{\mathrm{UCC}}$ of (12) can be viewed as an empirical Bayes estimate, in the limit of large n [9].

Suppose that $\boldsymbol{\theta}$ is a Gaussian vector consisting of independent identically distributed (iid) elements with variance τ, and that $\mathbf{e}$ is comprised of iid Gaussian variables with variance σ. If τ and σ are known, then the MMSE estimate of $\boldsymbol{\theta}$ from $\mathbf{y}$ is

$$\hat{\boldsymbol{\theta}} = \frac{\tau}{\sigma+\tau}\mathbf{y}. \quad (13)$$

Empirical Bayes methods are based on using (13) in conjunction with estimates for τ and σ.

There are two key properties of the CC approach which are instrumental in developing a Bayesian interpretation. The first, is that in deriving the CC estimate, both constraints $\|\mathbf{y}-\boldsymbol{\theta}\|^2 \le \rho$ and $\|\boldsymbol{\theta}\|^2 \le \eta$ are treated equally. This is in contrast with the CLS approach in which the first constraint is minimized subject to the second. In a Bayesian context, this implies that the covariance of $\boldsymbol{\theta}$ and $\mathbf{e}$ should be treated equally, namely both should be considered unknown, and not only that of $\boldsymbol{\theta}$ which is the typical approach in empirical Bayes methods. The second important feature is that only the difference $\rho-\eta$ plays a role. To incorporate this we express the estimate of (13) in a way that explicitly depends on the difference. Finally, in the CC context, we have prior knowledge on the norms of $\boldsymbol{\theta}$ and $\mathbf{e}$. Since the CC estimate depends only on the difference $\rho-\eta$, we assume that $\sigma-\tau$ is given. Using the fact that for large n, $\|\boldsymbol{\theta}\|^2 \to n\sigma$ and $\|\mathbf{e}\|^2 \to n\tau$, we choose

$$\sigma-\tau = \frac{\rho-\eta}{n}. \quad (14)$$

We can express the MMSE estimate of (13) as:

$$\hat{\boldsymbol{\theta}} = \frac{\tau}{\sigma+\tau}\mathbf{y} = \frac{1}{2}\left(1-\frac{\sigma-\tau}{\sigma+\tau}\right)\mathbf{y}. \quad (15)$$

Substituting (14), and using the fact that in the limit of large n,

$$\mathbf{y}^T\mathbf{y} \to n(\sigma+\tau), \quad (16)$$

results in the unrestricted CC estimate $\hat{\boldsymbol{\theta}}_{\mathrm{UCC}}$ of (12).

3.2. Generalized Bayesian Interpretation

We now show that the CC can be viewed as an exact MMSE estimate for a certain choice of priors on $\boldsymbol{\theta}$ and $\mathbf{e}$. Specifically we assume that $\boldsymbol{\theta}$ and $\mathbf{e}$ are zero-mean Gaussian vectors with covariance matrices $\tau\mathbf{I}$ and $\sigma\mathbf{I}$ respectively, where τ and σ are themselves random variables. This is in contrast to several generalized Bayesian estimates proposed in the literature in which the covariance of $\mathbf{e}$ is fixed, and a hierarchial prior is assumed only on $\boldsymbol{\theta}$ [10–12]. To build a prior on τ and σ we consider the norm constraints as well as the fact that the CC depends only on the difference $\rho-\eta$. Defining

$$r = \sigma-\tau, \quad s = \frac{1}{\sigma+\tau}, \quad (17)$$

we want to choose r as a function of $\rho-\eta$. On the other hand, the prior on s should not be too informative, as the sum $\rho+\eta$ plays no role in the Chebyshev estimate.

To incorporate the constraints $\|\boldsymbol{\theta}\|^2 \le \eta$ and $\|\mathbf{e}\|^2 \le \rho$ we assume that $E\{\tau\} = \eta/n$, and $E\{\sigma\} = \rho/n$. It then follows that $E\{r\} = (\rho-\eta)/n$. We do not make any further restrictions on the probability density function (pdf) of r. To ensure that s does not grow to fast we assume that s has the generalized pdf $p_s(s) = 1/s$. We further assume that r and s are independent. We now show that under this model the MMSE estimate of $\boldsymbol{\theta}$ from $\mathbf{y}$ is $\hat{\boldsymbol{\theta}}_{\mathrm{UCC}}$ of (12).

Given τ and σ, the MMSE estimate of $\boldsymbol{\theta}$ follows from (15) as $\hat{\boldsymbol{\theta}} = (1/2)(1-rs)\mathbf{y}$. Taking the expectation over τ and σ we have

$$\hat{\boldsymbol{\theta}} = \frac{1}{2}\left(1-E\left\{rs|\mathbf{y}\right\}\right)\mathbf{y}. \quad (18)$$

Now,

$$E\left\{rs|\mathbf{y}\right\} = \frac{\int\int rsp(\mathbf{y}|r,s)p_r(r)p_s(s)drds}{\int\int p(\mathbf{y}|r,s)p_r(r)p_s(s)drds} \quad (19)$$

where $p_r(r)$ is the pdf of r. Given r, s, the vector $\mathbf{y}$ is Gaussian with zero mean and covariance $(1/s)\mathbf{I}$. Therefore,

$$p(\mathbf{y}|r,s) \propto \exp(-\|\mathbf{y}\|^2 s/2)s^{n/2}. \quad (20)$$

Substituting (20) into (19) we have

$$E\left\{rs|\mathbf{y}\right\} = \frac{\int \exp(-\|\mathbf{y}\|^2 s/2)s^{n/2+1}p_s(s)ds \int rp_r(r)dr}{\int \exp(-\|\mathbf{y}\|^2 s/2)s^{n/2}p_s(s)ds}, \quad (21)$$

where we used the fact that $\int p_r(r)dr = 1$. Since $E\{r\} = (\rho-\eta)/n$ and $p_s(s) = 1/s$, (21) becomes

$$E\left\{rs|\mathbf{y}\right\} = \frac{(\rho-\eta)\int \exp(-\|\mathbf{y}\|^2 s/2)s^{n/2}ds}{n\int \exp(-\|\mathbf{y}\|^2 s/2)s^{n/2-1}ds}. \quad (22)$$

Applying integration by parts,

$$\int \exp(-\|\mathbf{y}\|^2 s/2)s^{n/2}ds = \frac{n}{\|\mathbf{y}\|^2}\int \exp(-\|\mathbf{y}\|^2 s/2)s^{n/2-1}ds, \quad (23)$$

so that (22) becomes

$$E\left\{rs|\mathbf{y}\right\} = \frac{\rho-\eta}{\|\mathbf{y}\|^2}. \quad (24)$$

Plugging (24) into (18) results in the estimate $\hat{\boldsymbol{\theta}}_{\mathrm{UCC}}$.

4. CHOOSING THE NOISE BOUND ρ

We now discuss two methods for choosing ρ assuming that $\mathbf{e}$ is a Gaussian random vector with iid random variables of variance σ^2.

In our first approach we use the fact that the random variable $z = \|\mathbf{y}-\boldsymbol{\theta}\|^2/\sigma^2$ has a chi-squared distribution with n degrees of freedom, denoted χ_2^n. Therefore, $p_t = P(z \le t) = \chi_n^2(t)$. We suggest choosing t such that p_t is close to 1. Denoting $\alpha = 1-p_t$, our strategy is to select $\rho = \sigma^2 t$ with $t = (\chi_n^2)^{1-\alpha}$, where $(\chi_n^2)^{1-\alpha}$ is the $1-\alpha$ percentile of the chi-squared distribution.

The second technique we suggest is based on minimizing the SURE estimate of the MSE subject to the constraint that $\mathcal{Q}$ is non empty. Let $\hat{\boldsymbol{\theta}} = \mathbf{y} + h(\mathbf{y})$ be an arbitrary estimate of $\boldsymbol{\theta}$ that is weakly differentiable in $\mathbf{y}$ and such that $E\left\{|h_i(\mathbf{y})|\right\}$ is bounded where $h_i(\mathbf{y})$ is the ith component of $h(\mathbf{y})$. Then an unbiased estimate of the MSE of $\hat{\boldsymbol{\theta}}$ is given from the SURE principle as [12]

$$S(\hat{\boldsymbol{\theta}}) = n\sigma^2 + \|h(\mathbf{y})\|^2 + 2\sigma^2\sum_{i=1}^{n}\frac{dh(\mathbf{y})}{d\mathbf{y}_i}. \quad (25)$$

The CC estimate (10) depends on a single parameter γ. Our approach is to choose the γ that minimizes the SURE estimate.

Theorem 4.1. *Let $\hat{\boldsymbol{\theta}}$ be the CC estimate that minimizes the SURE criterion and let $a = \mathbf{y}^T\mathbf{y}$. Then, for $\eta < 2\sigma^2$,*

$$\hat{\boldsymbol{\theta}} = \begin{cases} 0, & 0 \le a < a_0^2; \\ \sqrt{\frac{\eta}{a}}\mathbf{y}, & a \ge a_0^2. \end{cases} \quad (26)$$

For $\eta \ge 2\sigma^2$:

$$\hat{\boldsymbol{\theta}} = \begin{cases} 0, & 0 \le a < a_1; \\ \left(1-\frac{\sigma^2(n-2)}{\mathbf{y}^T\mathbf{y}}\right)\mathbf{y}, & a_1 \le a \le a_2; \\ \sqrt{\frac{\eta}{a}}\mathbf{y}, & a \ge a_2, \end{cases} \quad (27)$$

where a_0, a_1, a_2 are defined by

$$
\begin{aligned}
a_0 &= \frac{1}{4\sqrt{\eta}}\left(\eta + 2\sigma^2 + \sqrt{(\eta+2\sigma^2)^2 + 16\eta\sigma^2(n-2)}\right)\\
a_1 &= \sigma^2(n-1+\sqrt{2n-3})\\
a_2 &= \frac{1}{2}(\eta + 2\sigma^2(n-2) + \sqrt{\eta^2 + 4\eta\sigma^2(n-2)}).
\end{aligned}
$$

For small values of η, the resulting CC estimate can be viewed as a pre-test CLS method. Indeed, for values of a above a given threshold, the estimate coincides with the CLS solution of (3). For small values the estimate is 0. When η is increased, for a certain range of a values the estimate coincides with the JS technique.

5. SIMULATIONS RESULTS

We now demonstrate the behavior of our methods via several simulations. In the examples, we compare our estimators with the CLS estimate (3) and a projected JS approach which is a positive-part JS estimator truncated if its norm is too large:

$$
\hat{\boldsymbol{\theta}}_{\mathrm{PJS}} = \begin{cases} \left[1 - \frac{\sigma^2(n-2)}{\mathbf{y}^T\mathbf{y}}\right]_+ \mathbf{y}, & a \le a_{\mathrm{TH}};\\ \sqrt{\frac{\eta}{\mathbf{y}^T\mathbf{y}}}\mathbf{y}, & a \ge a_{\mathrm{TH}}. \end{cases} \tag{28}
$$

Here $[x]_+ = x$ if $x \geq 0$ and 0 otherwise, and a_{TH} is the unique value of a for which the norm of $\hat{\boldsymbol{\theta}}_{\mathrm{PJS}}$ as defined by the first row of (28) is equal to η. For each value of η we generate $5,000$ observation vectors $\mathbf{y}$, where the elements of $\boldsymbol{\theta}$ are chosen as iid Gaussian random variables. The entire vector is then scaled to norm 1 and multiplied by a uniform random variable on $[0, \sqrt{\eta}]$.

In Fig. 2 we plot the MSE for $\sigma^2 = 3$, $n = 7$, and $\alpha = 0.05$ in the chi-squared estimate. In this regime, the chi-squared Chebyshev approach seem to offer substantial MSE improvements over the other methods. Similar behavior was exhibited for $\alpha = 0.1$ and other choices of n. If the true norm of $\boldsymbol{\theta}$ is larger than η, then the performance of the chi-squared estimate deteriorates considerably. This can be seen in Fig. 3 where we repeat the simulations with $n = 3$ and the squared-norm of $\boldsymbol{\theta}$ distributed uniformly on $[0, 2\eta]$.

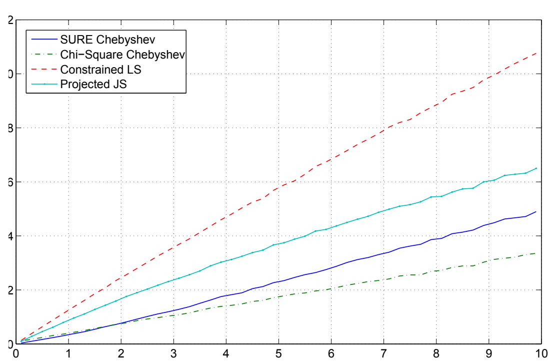

Fig. 2. MSE as a function of η for $n = 7$, $\sigma = 3$ and $\alpha = 0.05$.

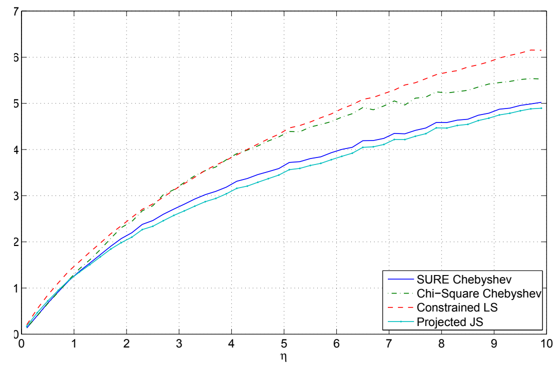

Fig. 3. MSE as a function of η for $n = 3$, $\sigma = 3$ and $\alpha = 0.05$ where the true value of η can be twice as large as the assumed value.

6. REFERENCES

[1] C. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," in *Proc. Third Berkeley Symp. Math. Statist. Prob.* 1956, vol. 1, pp. 197–206, University of California Press, Berkeley.

[2] W. James and C. Stein, "Estimation of quadratic loss," in *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* 1961, vol. 1, pp. 361–379, University of California Press, Berkeley.

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

[4] 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.

[5] 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.

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

[7] A. Charras and C. van Eeden, "Bayes and admissibility properties of estimators in truncated parameter spaces," *The Canadian Journal of Statistics*, vol. 19, no. 2, June 1991.

[8] A. Beck, "Convexity properties associated with nonconvex quadratic matrix functions and applications to quadratic programming," preprint.

[9] B. Efron and C. Morris, "Steins estimation rule and its competitors: an empirical Bayes approach," *J. Amer. Statist. Assoc.*, vol. 68, pp. 117–130, 1973.

[10] J. O. Berger, "A robust generalized Bayes estimator and confidence region for a multivariate normal mean," *Ann. Statist.*, vol. 8, no. 4, pp. 716–761, 1980.

[11] W. E. Strawderman, "Proper Bayes minimax estimators of multivariate normal mean," *Ann. Math. Statist.*, vol. 42, pp. 385–388, 1971.

[12] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Ann. Stat.*, vol. 9, no. 6, pp. 1135–1151, Nov. 1981.