Computing linear mini-max estimators

by

K. Helmes[1] *and C. Srinivasan*[2]

Abstract:

Consider a vector of data $y = \theta + \varepsilon$, $y \in I\!R^n$, where $\varepsilon = (\varepsilon_i)_{1 \le i \le n}$ is an error vector which satisfies the "usual" conditions, and $\theta \in \Theta \subset I\!R^n$. The problem is to find/compute a vector m for which $\langle m, y \rangle$ minimizes the maximal risk among all linear estimators, i.e. m is a solution of $\min_{m \in I\!R^n} \max_{\theta \in \Theta} E\left[\left(\langle \ell, \theta\rangle - \langle m, \theta \rangle\right)^2\right]$, $\ell \in I\!R^n$ given. A solution of this mini-max problem is determined by the solution of the fractional optimization problem $\max_{\theta \in \Theta} \frac{\langle \ell, \theta \rangle^2}{1 + \| \theta \|^2}$. We present an efficient method to solve the fractional programming problem for the special case when Θ is a symmetric, bounded set described by linear inequalities, i.e. $\Theta = \{\theta \mid A\theta \le b\}$. We also report on studies where the new method has been compared with direct approaches to the non-linear optimization problem.

1 Introduction

The following estimation problem has been extensively studied in the literature, cf., for instance, [2] – [16], [18] – [20], and the references cited therein. Consider a vector of data

$$y = \theta + \varepsilon, \tag{1.1}$$

where $y \in I\!R^n$, θ is an unknown vector, but $\theta \in \Theta$, Θ a given subset of $I\!R^n$, and $\varepsilon = (\varepsilon_i)_{1 \le i \le n}$ is a sequence of uncorrelated random variables with mean value zero and variance equal to one. The problem is to find a (linear) estimator for the number $\langle \ell, \theta \rangle = \sum_{i=1}^n \ell_i \theta_i$, where $\ell \in I\!R^n$ is given; more precisely, the question is to specify a linear expression of the data y, $\langle m^*, y \rangle = \sum_{i=1}^n m_i^* y_i$, which minimizes the maximal risk of all such procedures, i.e. m^* is a solution of the mini-max problem

$$\min_{m \in I\!R^n} \max_{\theta \in \Theta} E\left[\left(\langle \ell, \theta \rangle - \langle m, y \rangle\right)^2\right] = \min_{m \in I\!R^n} \max_{\theta \in \Theta} \left(\langle \ell - m, \theta \rangle^2 + \|m\|^2\right). \tag{1.2}$$

The fundamental result concerning this particular problem is the following mini-max.

[1]Inst. for Operations Research, Humboldt University of Berlin; supported in part by NSF Grant DMS 94 04 990.

[2]Dept. of Statistics, University of Kentucky.

Here we shall only give a special version of a more general result; for generalizations, modifications and a proof of the (general) theorem see [11], [7], [7] and [16].

Theorem 1.1 *Let Θ be a convex compact set which is symmetric about the origin. Then*

$$\begin{aligned} \min_{m\in\mathbb{R}^n}\max_{\theta\in\Theta} E\left[\Big(\langle\ell,\theta\rangle-\langle m,y\rangle\Big)^2\right] &= \min_{m\in\mathbb{R}^n}\max_{\theta\in\Theta}\Big(\langle\ell-m,\theta\rangle^2+\|m\|^2\Big) \\ &= \max_{\theta\in\Theta}\frac{\langle\ell,\theta\rangle^2}{1+\|\theta\|^2}\,. \end{aligned} \tag{1.3}$$

The up-shot of this result is the fact that the mini-max problem has been reduced to a fairly simple fractional programming problem. While there is an extensive literature on theoretical analyses of the linear mini-max estimation problem the aspect of numerically computing the optimal estimator has been somewhat "neglected", cf., however, [20]. The algorithm proposed in [20] is based on a direct approach to the mini-max problem, viz. computing the left-hand-side of Eq. (1.3). Here we consider algorithms to evaluate the right-hand-side of Eq. (1.3).

A first idea is to simply use any non-linear optimizer to solve the fractional programming problem. This approach has the benefit of being applicable no matter what kind of functionals describe the set Θ. A second idea is to solve a sequence of related maximization problems (in general, each problem is a *degenerate* quadratic optimization problem). To this end, consider the functional, $\lambda\in\mathbb{R}$,

$$\Phi(\lambda)=\max_{\theta\in\Theta}\left\{\langle\ell,\theta\rangle^2-\lambda\Big(1+\|\theta\|^2\Big)\right\}. \tag{1.4}$$

If can be easily shown that if λ^* is a root of Φ, $\Phi(\lambda^*)=0$, then λ^* is the value of the fractional programming problem described above. For the special case when Θ is described by linear constraints one can do better. For this case we propose an algorithm based on a change of variables. A series of numerical tests have shown that this algorithm is usually "better"than any of the other ones.

The paper is organized as follows. In Section 2 we describe the change of variables for the special case where $\ell=e_n=(0,0,\ldots,0,1)$, and describe a numerically efficient procedure how the case of a general vector ℓ can be reduced to the special case $\ell=e_n$. In the final section of the manuscript we provide some test results.

2 The change of variable approach

From now on we shall assume the set Θ to be defined as the set all solution vectors for a linear system of inequalities, i.e.

$$\Theta = \left\{ \theta \;\middle|\; A\theta \leq b \right\}, \tag{2.1}$$

where A and b are such that Θ is a compact, symmetric and, trivially, convex set. Assuming Θ to be symmetric implies that for every row $A_{[i]}$ of A the vector $-A_{[i]}$ is also a row of A, and that the corresponding co-ordinates of b are the same.
While assuming the prior knowledge about Θ to be captured by a set of linear inequalities is certainly a loss of generality, it can be justified for mainly three reasons:
(1) the appearance of linear constraints in many applications, e.g. for integral estimation, cf. [17], (2) the fact that non-linear constraints could be approximated using linear constraints and, by the way, are approximated this way at each iteration by any non-linear programming solver, and (3) the algorithm described below.

To find the linear mini-max estimator for the special case under consideration we need to solve the following optimization problem (cf. Introduction):

$$\max_{A\theta \leq b} \left\{ \frac{\langle \ell, \theta \rangle^2}{1 + \|\theta\|^2} \right\}. \tag{2.2}$$

Below we shall use the following notation: If not otherwise stated any vector is written as a coloumn matrix. If A is a $p \times q$-matrix then $A^{[i]} \in I\!R^p$, $1 \leq i \leq q$, denotes the i-th coloumn of A, and A can be written as:

$$A = \left[A^{[i]}, \ldots, A^{[q]} \right].$$

Theorem 2.1 *Let $\Theta = \{A\theta \leq b\}$ be a compact, symmetric (and convex) set, and let $\ell \neq 0$; without loss of generality assume $\ell_n < 0$ and $||\ell|| = 1$. Put $a = \ell - \|\ell\| e_n$, $q = \frac{a}{\|a\|}$ and $H = id - 2qq^T$. Define $\tilde{A} = AH^T$. Then*

$$\max_{A\theta \leq b} \left\{ \frac{\langle \ell, \theta \rangle^2}{1 + \|\theta\|^2} \right\} = \left[1 + \min_{\mathcal{A}u \geq \beta} \left\{ \|u\|^2 \right\} \right]^{-1}, \tag{2.3}$$

where $\mathcal{A} = \left[-\tilde{A}^{[1]}, \ldots, -\tilde{A}^{[n-1]}, +b \right]$ and $\beta = +\tilde{A}^{[n]}$.

Proof. By construction $H\ell = e_n$ (recall $||\ell|| = 1$), and H is a symmetric and orthogonal matrix. When checking $H\ell = e_n$ note that $2\langle a, \ell \rangle = \|a\|^2$, and that the assumption $\ell_n < 0$ holds. Define

$$\nu := H\theta. \tag{2.4}$$

Then

$$\max_{A\theta\le b}\left\{\frac{\langle \ell,\theta\rangle^2}{1+\|\theta\|^2}\right\}=\max_{\tilde{A}\nu\le b}\left\{\frac{\langle e_n,\nu\rangle^2}{1+\|\nu\|^2}\right\} = \max_{\tilde{A}\nu\le b}\left\{\frac{\nu_n^2}{1+\|\nu\|^2}\right\}. \tag{2.5}$$

To exclude a trivial case, we may assume the maximum to be different from zero. Let us write any feasible ν as $\nu=(x,\xi)$, where $x\in I\!R^{n-1}$, $\xi\in I\!R$. Taking the reciprocal of $\nu_n^2/\left(1+\|\nu\|^2\right)$, we obtain the minimization problem:

$$\min_{\tilde{A}\nu\le b}\left\{\frac{1+\|\nu\|^2}{\nu_n^2}\right\} = \min_{\substack{\tilde{A}\nu\le b\\ \nu=(x,\xi)}}\left\{\frac{1}{\xi^2}+\sum_{i=1}^{n-1}\frac{\nu_i^2}{\xi^2}+1\right\}. \tag{2.6}$$

Now change variables as follows. Let

$$\zeta := 1/\xi,\ ,\xi\neq 0,\quad \text{and}\quad z=\frac{x}{\xi}\in I\!R^{n-1}. \tag{2.7}$$

Furthermore, for $\nu=(x,\xi)$, $\xi>0$ (without loss of generality), multiply the vector inequality $\tilde{A}\nu\le b$ by $1/\xi$ to obtain:

$$\frac{\nu_1}{\xi}\tilde{A}^{[1]}+\ldots+\frac{\nu_{n-1}}{\xi}\tilde{A}^{[n-1]}+\tilde{A}^{[n]} \ \le\ \frac{1}{\xi}b.$$

Rearranging this linear combination, and using the variables $u=(z,\zeta)$ we see the equivalence of the maximization problem and the minimum distance problem, $u\neq 0$,

$$\min_{\mathcal{A}u\ge\beta}\left\{\|u\|^2\right\}, \tag{2.8}$$

where $\mathcal{A}$ und β are described in Theorem 2.1. □

Thus we may use any quadratic optimization solver for linear constraints to find the linear mini-max estimator; for instance, Wolfe's modified simplex algorithm would do; for more up-to-date methods cf. [1] and [13]. Given a solution of the minimization problem, i.e.

$$w^*=\|u^*\|^2=\min_{\mathcal{A}u\ge\beta}\left\{\|u\|^2\right\},$$

the best linear mini-max estimator for the case of a quadratic loss function, and its mini-max risk is given as follows:

Theorem 2.2 *Let $u^*=(z^*,\zeta^*)$ be a solution of the minimum distance problem* (2.8)*, and let w^* denote the minimal distance. Put*

$$\xi^* = \frac{1}{\zeta^*},\ x^* = \frac{z^*}{\zeta^*},\ \nu^* = (x^*, \xi^*)\ \textit{and}\ \theta^* = H^T\nu^*. \tag{2.9}$$

The linear minimax estimator is defined by the linear form $\langle m^*, y\rangle$ *where*

$$m^* = \frac{\langle \ell, \theta^*\rangle}{1 + \|\theta^*\|^2}\theta^*, \tag{2.10}$$

and the mini-max risk equals $\frac{1}{1+w^*}$.

We are now in the position to describe the "change-of-variables algorithm“ for computing the linear mini-max estimator for the case of linear constraints and a quadratic loss function.

Algorithm. Given the mini-max problem described in the theorem stated in the Introduction, see Eq. (1.3), and assuming w.l.o.g. $||\ell|| = 1$.

Step 1: Define the Householder matrix H as given in Theorem 2.1. Put $\tilde{A} = AH^T$, and consider the transformed fractional programming problem described by Eq. (2.5).

Step 2: Solve the minimum distance problem (2.8); see Theorem 2.1 for the notation used.

Step 3: Use the transformation rules stated in Theorem 2.2 to go back from the solution of the minimum distance problem to a solution of the original problem.

3 Some test results[1]

Here we provide some of the numerical results which we used to compare the three different ideas (algorithms) for computing linear mini-max estimators; all of them are using exististing optimization packages, f.i. SAS, NAG, IMSL and GINO. Here we shall only report on two (unrelated) test cases, viz.

case I: The vector $\ell = e_n$, and the linear system $A\theta \leq b$ is given by, $2m$ is the number of inequalities,

[1]We thank Tom Powers who helped us with some test runs.

$$a_{ij} = \begin{cases} (-1)^{i+j}(i+j-1)\,, & 1 \le i \le 2m-1 \quad \text{and } \ i \ \text{ is odd} \\ -a_{ij}\,, & 2 \le i \le 2m \quad \text{and } \ i \ \text{ is even} \end{cases}$$

$$b_i = 10p\,, \qquad i = 2p-1 \quad \text{and} \quad i = 2p\,, \quad 1 \le p \le m.$$

case II: The set of inequalities $A\theta \le b$ are defined as linear approximations to an ellipsoid, and $\ell = (1, 1, \ldots, 1)$. For the case of an ellipsoid there is an analytical solution for the linear minimax estimation problem. So we considered case II, among other reasons, to be able to check our numerical computations.

The maximization and minimization problems (2.2) and (2.8) respectively, were solved using different size linear systems of case I. The results in Table 1 (test runs using GINO) show that the minimization problem almost always yielded an optimal solution in fewer iterations than the maximization problem. The two cases, 50×50 and 100×50, which have the "max" outperforming the "min" are due to the fact that the solution obtained to the "max problem" was clearly not optimal. The same qualitative behaviour has been observed when we employed SAS-NLP and the NAG-library. Table 2 provides similar results for case I with a slightly modified objective function, i.e. we assumed the variance of the noise to be $\frac{2}{3}$ instead of 1. For case II we have analyzed a large number of test-runs using SAS-NLP. We have considered systems with 10^5 inequalities approximating ellipsoids in $\mathbb{R}^4$ and higher. The observations we made were the same as in case I.

Table 1: Comparison of Max and Min problems for Case I

Typ	Size	Iterations
Max	20×15	22
Min	20×15	4
Max	20×50	21
Min	20×50	7
Max	50×50	4
Min	50×50	7
Max	100×50	6
Min	100×50	9
Max	50×100	18
Min	50×100	7
Max	10×10	11
Min	10×10	7

Table 2: Comparison of Max and Min problems for Case I with noise -variance $\frac{2}{3}$

Typ	Size	Iterations
Max	20×15	20
Min	20×15	8
Max	20×50	25
Min	20×50	6
Max	50×50	24
Min	50×50	6
Max	100×50	46
Min	100×50	7
Max	50×100	39
Min	50×100	6
Max	10×10	13
Min	10×10	7
Max	100×70	53
Min	100×70	6
Min	100×80	6

Finally, table 3 shows that, depending on the package used, e.g. GINO compared with NAG, the number of iterations for the minimization problem would even be smaller. The qualitative behaviour of number of iterations for the max problem and the min problem is, however, the same irrespectively of the package used.

Table 3: Number of iterations for min-problems

Size	Gino	Nag
Min 20×15	8	2
Min 20×50	6	3
Min 50×50	6	3
Min 100×50	7	3
Min 50×100	6	3
Min 10×10	7	2
Min 100×70	6	3
Min 100×80	6	3

References

[1] Bertsekas, D.P. (1995). Nonlinear Programming. Athena Scientific, Belmont.

[2] Drygas, H. (1991). On an extension of the Girko inequality in linear minimax estimation. Proceeding of the Probastat '91 Conference, Bratislava/CSFR.

[3] Drygas, H. (1993). Spectral methods in linear minimax estimation. Kasseler Mathematische Schriften 4/93. To be published in the Proceedings of the Oldenburg Minimax Workshop (3./4.8.1992).

[4] Drygas, H. and Läuter (1993). On the representation of the linear minimax estimator in the convex linear model. Kasseler Mathematische Schriften No. 7/93.

[5] Gaffke, N. and Heiligers (1989). Bayes, admissible, and minimax linear estimators in linear models with restricted parameter space. *Statistics* 20, 487–508.

[6] Gaffke, N. and Heiligers (1991). Note on a Paper by P. Alson. *Statistics* 22, 3–8.

[7] Helmes, K. and Srinivasan (1992). Linear minimax estimation. *Lecture Notes in Economics and Mathematics,* 389, 9–23.

[8] Helmes, K. and Christopeit (1996). Linear estimation with ellipsoidal constraints. To appear in *Acta Applicandae Mathematicae*, vol. 43, No. 1, 3–15.

[9] Hoffmann, K. (1979). Characterization of minimax linear estimators in linear regression. *Math. Operationsforsch. u. Statistik, Ser. Statist.*, 20, 19–26.

[10] Ibragimov, A.D. and Hasminski (1987). Estimation of linear functionals in Gaussian noise. *Theory Prob. Appl.,* vol. 32, No. 1, 30–39.

[11] Ibragimov, A.D. and Hasminski (1984). On non-parametric estimation of the value of a linear function in Gaussian white noise. *Theory of Probability and Its Applications,* Vol. XXIX, No. 1, 18–32.

[12] Läuter, H. (1975). A minimax linear estimator for linear parameters under restrictions in form of inequalities. *Math. Operationsforsch. u. Statist.,* 6, Heft 5, 689–695.

[13] Nash, S.G. and Sofer (1996). Linear and Nonlinear Programming. McGraw-Hill, New York.

[14] Pilz, J. (1986). Minimax linear regression estimation with symmetric parameter restrictions. *J. Statist. Plann. Inference* 13, 297–318.

[15] Pilz, J. (1991). Bayesian estimation and experimental design in linear regression models. 2nd edition, Wiley, Chichester, New York.

[16] Pinelis, I.F. (1988). On minimax risk. *Theory Prob. Appl.*, vol. 33, 104–109.

[17] Ritter, K. (1995). Asymptotic optimality of regular sequence design. To appear in *Annals of Statistics*.

[18] Stahlecker, P. and Drygas (1992). Representation theorems in linear minimax estimation. Report No. V-85-92, University of Oldenburg.

[19] Stahlecker, P., Jänner and Schmidt (1991). Linear-affine Minimax-Schätzer unter Ungleichungsrestriktionen. *Allg. Statist. Archiv,* 75, 245–264.

[20] Stahlecker, P. and Trenkler (1991). Linear and ellipsoidal restrictions in linear regression. *Statistics* 22, 163–176.