BIT
2002, Vol. 42, No. 4, pp. 856–866
0006-3835/02/4204-0856 $16.00

LINEAR HUBER M-ESTIMATOR UNDER ELLIPSOIDAL DATA UNCERTAINTY *

M. Ç. PINAR[1]

[1] *Department of Industrial Engineering, Bilkent University, 06533 Ankara, Turkey*
email: mustafap@bilkent.edu.tr

Abstract.

The purpose of this note is to present a robust counterpart of the Huber estimation problem in the sense of Ben-Tal and Nemirovski when the data elements are subject to ellipsoidal uncertainty. The robust counterparts are polynomially solvable second-order cone programs with the strong duality property. We illustrate the effectiveness of the robust counterpart approach on a numerical example.

AMS subject classification: 62J05, 90C25, 65D10, 65F20.

Key words: Data fitting, least squares problems, robustness, Huber's M-estimator, second-order cone programming.

1 Introduction and background.

An important problem of data analysis is to estimate a set of parameters in a linear model specified as an overdetermined set of equations $Ax \approx b$ where $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$. The most common way of determining the values of parameters is to minimize the residual $Ax - b$ in some norm where the norms of choice are the 1, ∞ and 2-norms. The latter case, by far the most popular, is known as the least-squares problem. When the problem data are known to be plagued with errors, or, simply cannot be measured accurately, several variants of the least squares criteria were proposed to compute solutions "immune" to such uncertainties, such as total least squares, ridge regression, regularized least squares, etc. Reference [7] gives a selected list of important contributions in this area. Recently, Chandrasekaran et al. [6] and El-Ghaoui and Lebret [9] initiated independently the study of a variant of the least squares problem where A and b were subject to unknown but bounded errors. In an important departure from previous approaches to uncertainty, they proposed minimizing the maximum error under such bounded errors, and derived a closed-form objective function for the problem. Following the publication of these papers, Watson [18] and Hindi and Boyd [11] gave extensions of this variant to the 1 and ∞-norm cases. More precisely, Watson [18] extends the bounded perturbation case to general p-norms, and studies solution algorithms while Hindi and Boyd [11] consider the 1, ∞, and 2-norm cases, for bounded, stochastic (2-norm only) and structured uncertainty cases.

In a parallel but independent line of work, Ben-Tal and Nemirovski [2] introduced a new concept of robustness for mathematical programming problems where data

*Received July 2001. Revised May 2002. Communicated by Kaj Madsen.

are subject to ellipsoidal uncertainty. They derive *robust counterparts* of linear, quadratic, second-order cone and semidefinite programming problems. They also give an example of an engineering design (array synthesis design) problem in the ∞-norm and the 2-norm and derive robust counterparts in [4]. The survey by Lobo et al. [14] reserves a paragraph to the robust least-squares problem where ellipsoidal uncertainty is also briefly considered.

Although the case of bounded uncertainty is exposed at length in the references cited above, the application of robust counterpart technique of Ben-Tal and Nemirovski remains scattered through sections of a few, more general research articles, and as an example in a book on convex optimization applications in engineering [4]. However, linear data fitting applications are so pervasive in applications that these ideas, in our view, deserve to be disseminated beyond the linear algebra and optimization community. Hence, our goal in this note is to compile important aspects of this new modeling effort in a clear, simple and easily accessible form, and add to the spectrum yet another criterion for data fitting, namely the Huber criterion.

Let us illustrate the idea of a robust counterpart using the 1-norm data fitting problem. Incidentally, the robust counterpart of this problem does not appear in any of the references cited above. The 1-norm data fitting problem which consists of finding a minimizer of the function $\|Ax - b\|_1$ can be expressed as the following problem:

$$\begin{array}{ll} \min & \sum_{i=1}^{m} t_i \\ \text{s.t.} & |a_i^T x - b_i| \le t_i, \quad \forall i = 1, \ldots, m. \end{array}$$

Assume that the rows of A are subject to independent errors, but known to lie in a given ellipsoid: $a_i \in \mathcal{E}_i$, where

$$\mathcal{E}_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \le 1\}, \tag{1.1}$$

with $P_i \in \mathbf{R}^{n \times n}$ a symmetric matrix. Ben-Tal and Nemirovski [3] show that such uncertainty sets are quite accurate representations of modeling situations where we have access to the mean value and standard deviation of data, and we act as an engineer who is willing to accept a certain deviation around the mean, measured by a constant times the square root of the variance-covariance matrix of the uncertain data vector. These considerations typically lead to ellipsoidal uncertainty sets of the type (1.1).

Now, the robust counterpart of the 1-norm data fitting problem in the sense of Ben-Tal–Nemirovski is the following problem:

$$\begin{array}{ll} \min & \sum_{i=1}^{m} t_i \\ \text{s.t.} & |a_i^T x - b_i| \le t_i, \quad a_i \in \mathcal{E}_i, \quad \forall i = 1, \ldots, m. \end{array}$$

The above problem can be rewritten

$$\begin{array}{lrcll} \min & \sum_{i=1}^{m} t_i & & & \\ \text{s.t.} & a_i^T x - b_i & \le & t_i, & \forall i = 1, \ldots, m, \\ & -a_i^T x + b_i & \le & t_i, & \forall i = 1, \ldots, m, \\ & a_i \in \mathcal{E}_i, & & & \forall i = 1, \ldots, m. \end{array}$$

In other words, we require that the constraints be satisfied for all realizations of the rows of A, and we want to pick the best solution among all feasible solutions that satisfy all possible realizations. Hence, although the above problem has infinitely many constraints, it can be cast into the following equivalent problem:

$$\begin{array}{lrcll} \min & \sum_{i=1}^{m} t_i & & & \\ \text{s.t.} & \max_{\|u\|_2 \le 1} \{\bar{a}_i^T x - b_i + u^T P_i x\} & \le & t_i, & \forall i = 1, \ldots, m, \\ & \max_{\|u\|_2 \le 1} \{-\bar{a}_i^T x + b_i - u^T P_i x\} & \le & t_i, & \forall i = 1, \ldots, m. \end{array}$$

Since $\max_{\|u\|_2 \le 1} u^T P_i x = \max_{\|u\|_2 \le 1} -u^T P_i x = \|P_i x\|_2$ we obtain the following robust counterpart program **L1R**:

$$\begin{array}{lrcll} \min & \sum_{i=1}^{m} t_i & & & \\ \text{s.t.} & |\bar{a}_i^T x - b_i| + \|P_i x\|_2 & \le & t_i, & \forall i = 1, \ldots, m, \end{array}$$

which is a particular instance of a convex, second-order cone program [4, 14], i.e., a problem of the following form:

$$\begin{array}{lrcll} \min & f^T x & & & \\ \text{s.t.} & \|A_i x + b_i\|_2 & \le & c_i^T x + d_i, & \forall i = 1, \ldots, N, \end{array}$$

for which polynomial interior point methods and efficient implementations exist, e.g., the software systems SOCP, SEDUMI, and LOQO [15, 17, 1]. The reader interested in second-order cone programming is directed to the excellent survey article [14]. The above instance of the second-order cone programming problem is equivalent to finding a minimizer of the 1-norm of a vector with ith component equal to $|\bar{a}_i^T x - b_i| + \|P_i x\|_2$. Using the same derivation technique, one can show that the 2-norm (least squares) and the Chebyshev norm (∞-norm) give rise respectively to robust counterpart problems where the residual vector is replaced by the vector whose ith component is $|\bar{a}_i^T x - b_i| + \|P_i x\|_2$.

Against this background, we add to the repertoire of robust counterparts the robust counterpart of the Huber estimation problem [12] under ellipsoidal uncertainty in the data. Although the Huber function is not a norm (i.e., does not satisfy all the axioms of a vector norm), we find a robust counterpart for it similar to those listed above. This development is given in the next section.

2 Huber estimator and its robust counterpart.

Huber estimation is concerned with identifying "outliers" among data points b_i and giving them less weight. The Huber estimator is essentially the least squares estimator, but uses the 1-norm for points that are considered outliers with respect to a certain threshold. Hence, the Huber criterion is less sensitive to the presence of outliers, and its usage would be appropriate when deviations from the normality assumption in the estimation errors are present. Boyd mentions the use of the Huber estimator in signal processing applications where the errors have exponentially distributed tails while following a Gaussian distribution otherwise [5]. The structural properties of this problem along with solution algorithms can be found in the extensive references of [13, 16].

More precisely, Huber's M-estimate is a minimizer $x^* \in \mathbf{R}^n$ of the function

$$F(x) = \sum_{i=1}^{m} \rho(r_i(x)) \tag{2.1}$$

where

$$\rho(t) = \begin{cases} \frac{1}{2\gamma} t^2, & \text{if } |t| \le \gamma, \\ |t| - \frac{1}{2}\gamma, & \text{if } |t| > \gamma, \end{cases} \tag{2.2}$$

with a tuning constant $\gamma > 0$. The residual $r_i(x)$ is defined as

$$r_i(x) = a_i^T x - b_i, \tag{2.3}$$

for all $i = 1, \dots, m$ with $r = Ax - b$.

To derive the robust counterpart problem, we pose the primal problem as a quadratic programming problem that we refer to as **HQP**:

$$\begin{array}{ll} \min & \dfrac{1}{2\gamma} \displaystyle\sum_{i=1}^{m} p_i^2 + \sum_{i=1}^{m} (q_i - \gamma/2) \\ \text{s.t.} & -p - q \le b - A^T x \le p + q, \\ & \quad 0 \le p \le \gamma e, \qquad q \ge 0, \end{array}$$

where e denotes a vector with all components unity.

PROPOSITION 2.1. *Any optimal solution to the quadratic program HQP is a minimizer of F, and conversely.*

PROOF. Let x be a minimizer of F and define $p_i = \min\{|a_i^T x - b_i|, \gamma\}$, and $q_i = |a_i^T x - b_i| - p_i$. This point is feasible for (HQP). Moreover,

$$\frac{1}{2\gamma} \sum_{i=1}^{m} p_i^2 + \sum_{i=1}^{m} (q_i - \gamma/2) = \sum_{i=1}^{m} \rho(a_i^T x - b_i).$$

Furthermore, let $\bar{x}, \bar{p}, \bar{q}$ be an optimal solution to (HQP). It is easy to see that $|a_i^T \bar{x} - b_i| = \bar{p}_i + \bar{q}_i$ for $i = 1, \dots, m$. Therefore,

$$\frac{1}{2\gamma} \sum_{i=1}^{m} \bar{p}_i^2 + \sum_{i=1}^{m} (\bar{q}_i - \gamma/2) = \sum_{i=1}^{m} \rho(a_i^T \bar{x} - b_i).$$

□

Now, consider the problem HQP where the rows a_i of A are confined to stay in ellipsoids as in the previous section, i.e., $a_i \in \mathcal{E}_i = \{\bar{a}_i + P_i u | \ \|u\|_2 \leq 1\}$ with $P_i \in \mathbf{R}^{n\times n}$ a symmetric matrix for all $i = 1,\dots,m$. We have immediately the following robust counterpart:

$$\begin{array}{ll} \min & \frac{1}{2\gamma}\sum_{i=1}^m p_i^2 + \sum_{i=1}^m (q_i - \gamma/2) \\ \text{s.t.} & \max_{\|u\|_2\leq 1}\{\bar{a}_i^T x - b_i + u^T P_i x\} \leq p_i + q_i, \quad \forall i = 1,\dots,m, \\ & \max_{\|u\|_2\leq 1}\{-\bar{a}_i^T x + b_i - u^T P_i x\} \leq p_i + q_i, \quad \forall i = 1,\dots,m, \\ & 0 \leq p \leq \gamma e, \qquad q \geq 0, \end{array}$$

which yields the program **RHQP**

$$\begin{array}{ll} \min & \frac{1}{2\gamma}\sum_{i=1}^m p_i^2 + \sum_{i=1}^m (q_i - \gamma/2) \\ \text{s.t.} & |\bar{a}_i^T x - b_i| + \|P_i x\|_2 \leq p_i + q_i, \quad \forall i = 1,\dots,m, \\ & 0 \leq p \leq \gamma e, \qquad q \geq 0, \end{array}$$

which is equivalent to a Huber estimation problem where every residual $a_i^T x - b_i$ is replaced by $|\bar{a}_i^T x - b_i| + \|P_i x\|$.

Thus far, we have assumed that the uncertainty is restricted to A, and that the rows of A are subject to independent errors confined to ellipsoids. A variant of the problem is to consider the case where the elements of b are also subject to independent errors as follows. Let

$$\begin{pmatrix} a_i \\ b_i \end{pmatrix} \in \mathcal{E}_i = \left\{ \begin{pmatrix} a_i \\ b_i \end{pmatrix} + Q_i u : \|u\|_2 \leq 1 \right\}$$

with symmetric $Q_i \in \mathbf{R}^{(n+1)\times(n+1)}$. Partition the $(n+1)\times(n+1)$ matrix Q_i as $Q_i = [P_i : d_i]$ where $P_i \in \mathbf{R}^{(n+1)\times n}$ and $d_i \in \mathbf{R}^{n+1}$. It is easy to verify that the robust counterpart problem is

$$\begin{array}{ll} \min & \frac{1}{2\gamma}\sum_{i=1}^m p_i^2 + \sum_{i=1}^m (q_i - \gamma/2) \\ \text{s.t.} & |\bar{a}_i^T x - b_i| + \|P_i x - d_i\|_2 \leq p_i + q_i, \quad \forall i = 1,\dots,m, \\ & 0 \leq p \leq \gamma e, \qquad q \geq 0. \end{array}$$

3 Optimality conditions and duality.

In this section we investigate optimality conditions to characterize minimizers to the robust counterpart problems we dealt with. Interestingly, all the robust counterpart problems corresponding to 1, ∞, and 2 norms and the Huber criterion have an optimal value bounded below by zero, and satisfy trivially the Slater condition, and thus lead to duals where strong duality is attained, i.e., the optimal values of the respective primal and dual problems are equal; see Theorem 2.4.1 of [4]. We first rewrite the robust counterpart of the Huber problem as follows.

$$\begin{array}{lll} \min & \frac{1}{2\gamma}\sum_{i=1}^{m} p_i^2 + \sum_{i=1}^{m}(q_i - \gamma/2) & \\ \text{s.t.} & t_i + |\bar{a}_i^T x - b_i| \le p_i + q_i, & \forall i = 1,\ldots,m, \\ & \|u_i\|_2 \le t_i, & \forall i = 1,\ldots,m, \\ & P_i x = u_i, & \forall i = 1,\ldots,m, \\ & 0 \le p \le \gamma e, & q \ge 0. \end{array}$$

Define the Lagrange function with multiplier vectors $y \in \mathbf{R}_+^m$, $z \in \mathbf{R}_+^m$ and $w_i \in \mathbf{R}^n$ for $i = 1,\ldots,m$

$$\begin{aligned} L(p,q,x,y,z,w_i) \\ = \frac{1}{2\gamma}\sum_{i=1}^{m} p_i^2 + \sum_{i=1}^{m}(q_i - \gamma/2) + \sum_{i=1}^{m} y_i(a_i^T x - b_i + t_i - p_i - q_i) \\ + \sum_{i=1}^{m} z_i(-a_i^T x + b_i + t_i - p_i - q_i) + \sum_{i=1}^{m} w_i^T(u_i - P_i x). \end{aligned}$$

The minimization of the Lagrange function in p_i over $0 \le p_i \le \gamma$ yields the requirement that y_i and z_i satisfy

$$0 \le y_i + z_i \le 1, \tag{3.1}$$

for all $i = 1,\ldots,m$. The minimization over x yields the equality

$$\bar{A}^T(y - z) = \sum_{i=1}^{m} P_i w_i.$$

Finally, for $i = 1,\ldots,m$, we have the term

$$\min_{u_i,t_i,\|u_i\|_2 \le t_i} w_i^T u_i + t_i(y_i + z_i).$$

This minimization yields the requirement

$$|w_i\|_2 \le y_i + z_i$$

for all $i = 1,\ldots,m$. To see why this is true, fix $t_i > 0$. Then, we have

$$\min_{u_i:\|u_i\|_2 \le t_i} w_i^T u_i + t_i(y_i + z_i) = -t_i\|w_i\|_2 + t_i(y_i + z_i).$$

For the minimization over t_i to yield a finite value (zero), it suffices that $\|w_i\|_2 \le$

$y_i + z_i$. Hence, we have obtained the following dual program:

$$\begin{array}{ll} \max & -\frac{1}{2}\gamma\sum_{i=1}^{m}(y_i+z_i)^2 - b^T(y-z) - m\frac{\gamma}{2} \\ \text{s.t.} & \bar{A}^T(y-z) = \sum_{i=1}^{m} P_i w_i, \\ & \|w_i\|_2 \le y_i + z_i, \qquad \forall i = 1,\ldots,m, \\ & 0 \le y_i + z_i \le 1, \qquad \forall i = 1,\ldots,m, \\ & y_i \ge 0, \qquad \forall i = 1,\ldots,m, \\ & z_i \ge 0, \qquad \forall i = 1,\ldots,m, \end{array}$$

which is again a second-order cone programming problem. Hence, for x to be an optimal solution in the robust counterpart RHQP it is necessary and sufficient that it exist $(y, z, \mathcal{W})$, where $\mathcal{W}$ is the $n \times m$ matrix with columns w_i, $i = 1, \ldots, m$, which satisfy the constraints of the dual, for which equality between the primal and dual objective functions is observed. It is easy to verify that setting $\gamma = 0$ in the dual program above, we obtain the following second-order cone program

$$\begin{array}{ll} \max & -b^T(y-z) \\ \text{s.t.} & \bar{A}^T(y-z) = \sum_{i=1}^{m} P_i w_i, \\ & \|w_i\|_2 \le y_i + z_i, \qquad \forall i = 1,\ldots,m, \\ & 0 \le y_i + z_i \le 1, \qquad \forall i = 1,\ldots,m, \\ & y_i \ge 0, \qquad \forall i = 1,\ldots,m, \\ & z_i \ge 0, \qquad \forall i = 1,\ldots,m. \end{array}$$

which is nothing else than the dual program to L1R.

4 A numerical example.

In this section we illustrate the utility of the robust counterpart approach in the context of Huber M-estimation using a numerical example inspired from [4]. We consider a linear regression problem of the form (2.1) where the matrix A and the vector b[1] have dimensions 21×10 and 21, respectively. We take $\gamma = 0.001$. The optimal solution returned by the nonlinear programming solver FILTER [8] has value 0.06509 while the optimal coefficients are

$$x^* = (-459.89, 395.99, -294.62, 195.2, \\ -89.54, 33.01, -12.87, 2.88, -0.4455, 0.0348)^T.$$

Now, consider a random perturbation of the optimal solution obtained as $x_j^{\text{pert}} = x_j^* \eta_j$ where η_j is a normally distributed random variable with mean 1 and variance 0.00001 for all $j = 1, \ldots, n$. Computing the objective function value in problem

[1]Data are available at `http://www.ie.bilkent.edu.tr/~mustafap/data`

(2.1) corresponding to x^{pert} gives 110.66! The optimal solution we computed seems to be extremely unstable with respect a small perturbation. To remedy this instability we can use the robust counterpart approach as follows. Consider the residuals $r_i = \sum_{j=i}^n a_{ij}x_j - b_i$. The random perturbation we introduced to the optimal solution x^* can be thought of as a perturbation of the coefficients a_{ij}. Hence, for fixed x our actual residuals are of the form

$$\xi_i(x) = \sum_{j=1}^n a_{ij}\eta_j x_j - b_i,$$

where η_j is a random variable with variance 0.00001. Since for fixed x, $\xi_i(x)$ is now a random variable for all $i = 1, \ldots, m$, it has expected value equal to

$$\xi_i^*(x) = \sum_{j=1}^n a_{ij}x_j,$$

and standard deviation

$$\sigma_i(x) = \sqrt{E\{(\xi_i(x) - \xi_i^*(x))^2\}} = \sqrt{\sum_{j=1}^n x_j^2 a_{ij}^2 E\{(\eta_j - 1)^2\}} = 0.001\sqrt{\sum_{j=1}^n x_j^2 a_{ij}^2}.$$

Now, using the methodology of Ben-Tal and Nemirovski we can act as an engineer, who believes that a random variable will never differ from its mean value by more than a constant, say two or three, times its standard deviation. Therefore, we can choose a safety parameter ω and ignore all events which result in $|\xi_i(x) - \xi_i^*(x)| > \omega\sigma_i(x)$. As a result, we obtain as robust versions of the constraints

$$\begin{aligned} a_i^T x - b_i &\leq p_i + q_i, \\ -p_i - q_i &\leq a_i^T x - b_i \end{aligned}$$

the constraints

$$\begin{aligned} a_i^T x + \omega\sigma_i(x) - b_i &\leq p_i + q_i, \\ -p_i - q_i &\leq a_i^T x - \omega\sigma_i(x) - b_i, \end{aligned}$$

for all $i = 1, \ldots, m$. Therefore we obtain a robust problem of the form

$$\begin{array}{ll} \min & \dfrac{1}{2\gamma}\displaystyle\sum_{i=1}^m p_i^2 + \sum_{i=1}^m (q_i - \gamma/2) \\ \text{s.t.} & |a_i^T x - b_i| + \|P_i x\|_2 \leq p_i + q_i, \ \forall i = 1, \ldots, m, \\ & \qquad 0 \leq p \leq \gamma e, \qquad\qquad q \geq 0, \end{array}$$

where $P_i = 0.001\omega \mathrm{Diag}(a_{i1}, a_{i2}, \ldots, a_{in})$. On the other hand, it is easy to see that the above robust problem can be obtained as the robust counterpart of the Huber M-estimation problem corresponding to the ellipsoidal uncertainty set:

$$E_i = \{a_i + 0.001\omega Q_i u | \|u\|_2 \leq 1\},$$

where $Q_i = \mathrm{Diag}(a_{i1}, a_{i2}, \ldots, a_{in})$, and uncertainty ellipsoids affect each row i of the matrix A.

We solve the robust counterpart for values of $\omega = 1, 2, 3$. For $\omega = 1$ the optimal solution has value 0.22669 while the coefficients x have the following optimal values:

$$(-0.00119, 0.00077, 0.00148, -0.00019, -0.00101,\\ 0.00043, 0.00055, -0.00074, 0.00034, -5.66758 \times 10^{-5})^T.$$

We apply the same random (normal with mean 1 and variance 0.00001) perturbation to this solution, and find that in 10 replicates the objective function value varies between 0.22664 and 0.22671. For $\omega = 2$, we obtain a robust value equal to 0.256 along with optimal coefficients

$$x = (-0.00013, 0.00067, 0.00096, -2.32751 \times 10^{-5}, -0.00058\\ 0.00013, 0.00034, -0.000258, 3.35051 \times 10^{-5}, 2.79662 \times 10^{-5})^T.$$

The objective function values obtained from 10 random perturbations range from 0.25598 to 0.25603. Finally, for $\omega = 3$, we get an optimal value equal to 0.26489 along with an optimal solution vector

$$x = (0.00012, 0.00065, 0.00083, 1.37335 \times 10^{-5}, -0.00046\\ 6.12315 \times 10^{-5}, 0.00028, -0.00013, -4.30556 \times 10^{-5}, 4.65422 \times 10^{-5})^T.$$

The objective function values fluctuate between 0.26487 and 0.26492 in this case. We can conclude that the three solutions reported above are very stable with respect to the random perturbations introduced above, and thus to the particular form of ellipsoidal uncertainty considered in our numerical example although the robust objective function value represents an increase to around 0.25 from an optimal value of 0.065.

An important question is to ask what would happen to the robust optimal value if we were to introduce normal perturbations with a variance equal to 0.0001. It turns out that, although we hedged ourselves against perturbations with variance equal to 0.00001, the objective function value (for $\omega = 3$) fluctuates only between 0.26474 and 0.26528 in 10 replications. Our robust solution is indeed quite insensitive to even larger perturbations! If we use a normal perturbation with variance equal to 0.001 (one hundred times larger) the robust solution fluctuates between an objective value of 0.27319 and 0.26547 only. For $\omega = 1$ and normal random perturbations with variance 0.001 the fluctuation in objective function value is only between 0.2285 and 0.25677 in 10 trials. These results demonstrate the stability of the robust solution. The choice of ω does not seem to influence stability much.

As a comparison to our method, we used a straightforward Tikhonov regularization [19] which consists in solving the following problem:

$$\min F(x) + \mu \|x\|_2^2,$$

where F is the Huber function. We tried the values $\mu = 0.1$, 0.5, 1, 2, 5, 10, 20, 200. When we solved this problem we obtained a solution which seems robust at first sight. The objective function value varies between 0.10035 (for $\mu = 0.1$) and 0.22049 (for $\mu = 200$) where the increase is monotonic with increasing values of μ.

We observed that for normal perturbations of variance 0.00001 this value changes little. However, when we use normal perturbations with variance 0.001, the objective function value of the regularized solution for $\mu = 1$ takes the following values in 10 replicates:

$$(2.25197, 1.42348, 1.78415, 1.22244, 2.75104, 1.32303, 2.31127, 2.73932, 1.67131, 0.54571).$$

For $\mu = 0.1$ where we obtained the smallest objective function value in our sample, the variation of objective function values under perturbations with variance 0.001 in 10 replicates is

$$(3.31033, 1.24997, 1.054, 1.76022, 2.50652, 1.49002, 3.44534, 2.53109, 1.60482, 0.81992).$$

For $\mu = 10$, the variation of objective function values under perturbations with variance 0.001 in 10 replicates is

$$(1.23913, 0.80973, 1.23441, 0.742178, 1.67835, 0.84139, 1.34522, 1.57795, 1.05429, 0.45797).$$

It appears that the deviation in the objective function value under random perturbations decreases as μ is increased. We obtained our best result with Tikhonov regularization with $\mu = 200$ where the solution has indeed small variance under random perturbations: the objective function value varied only between 0.22653 and 0.31784 in 10 trials. Increasing μ beyond this value does not change the situation.

Therefore, we can conclude that our method matches the power of Tikhonov regularization in this particular example without having to tune a regularization parameter. The choice of the regularization parameter is an active area of research; for a recent coverage of the subject the interested reader is referred to [10].

5 Conclusions.

We considered the robust counterpart of Huber's M-estimation problem in the sense of Ben-Tal and Nemirovski [2] for linear data fitting problems where the data is subject to ellipsoidal uncertainty. We derived a robust problem which is a second-order cone programming problem, investigated duality issues and optimality conditions, and finally gave a numerical example illustrating the effectiveness of the robust counterpart approach in the presence of severe instability of optimal solutions in a Huber M-estimation problem.

Acknowledgement.

The manuscript benefited from the comments of two anonymous reviewers.

REFERENCES

1. H. Y. Benson and R. Vanderbei, *Using LOQO to solve second-order cone problems*, Technical Report 98–09, Princeton University, Princeton, NJ, 1999.
2. A. Ben-Tal and A. Nemirovski, *Robust convex optimization*, Math. Oper. Res., 23 (1998), pp. 769–805.

3. A. Ben-Tal and A. Nemirovski, *Robust solutions to uncertain linear programs via convex programming*, Oper. Res. Lett., 25 (1999), pp. 1–13.
4. A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms and Engineering Applications*, SIAM, Philadelphia, PA, 2001.
5. S. P. Boyd, *EE 364 Final examination*, Stanford University, Stanford, CA, 1999.
6. S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, *Parameter estimation in the presence of data uncertainties*, SIAM J. Matrix Anal. Appl., 19 (1998), pp. 235–252.
7. S. Chandrasekaran, G. H. Golub, M. Gu, and A. H. Sayed, *Efficient algorithms for least squares type problems with bounded uncertainties*, in Recent Advances in Total Least-Squares Techniques and Errors-in-Variables Modeling, S. Van Huffel, ed., SIAM, Philadelphia, PA., 1998, pp. 171–180.
8. R. Fletcher and S. Leyffer, *User manual for FILTER/SQP*, Numerical Analysis Report NA-181, University of Dundee, Dundee, UK, 1998.
9. L. El-Ghaoui and H. Lebret, *Robust solutions to least squares problems with uncertain data*, SIAM J. Matrix Anal. Appl., 18 (1997), pp. 1037–1064.
10. P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*, SIAM, Philadelphia, PA, 1998.
11. H. A. Hindi and S. P. Boyd, *Robust solutions to ℓ_1, ℓ_2 and ℓ_∞ linear approximation problems using convex optimization*, Tech. Report, Department of Electrical Engineering, Stanford University, Stanford, CA, 1999.
12. P. Huber, *Robust Statistics*, Wiley, New York, 1981.
13. W. Li and J. Swetits, *Linear ℓ_1 estimator and Huber estimator*, SIAM J. Optim., 8 (1998), pp. 457–475.
14. M. S. Lobo, L. Vandenberghe, S. P. Boyd, and H. Lebret, *Applications of second-order cone programming*, Linear Algebra Appl., 284 (1998), pp. 193–228.
15. M. S. Lobo, L. Vandenberghe, and S. P. Boyd, *SOCP: Primal-dual potential reduction method for solving second-order cone programming problems*, Tech. Report, Stanford University, Stanford, CA, 1998.
16. K. Madsen and H. B. Nielsen, *Finite algorithms for Huber's M-estimator*, BIT, 30 (1990), pp. 682–699.
17. J. Sturm, *Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones*, Optim. Meth. Software, 11–12 (1999), pp. 625–653.
18. G. A. Watson, *Solving data fitting problems in ℓ_p norms with bounded uncertainties in the data*, Tech. Report, University of Dundee, Dundee, UK, 1999.
19. H. Zha and P. C. Hansen, *Regularization and the general Gauss–Markov linear model*, Math. Comp., 55 (1990), pp. 613–624.