
Computational Optimization and Applications, 17, 329–341, 2000

Accurate Solution to Overdetermined Linear Equations with Errors Using L_1 Norm Minimization

J. BEN ROSEN jbrosen@cs.ucsd.edu
Computer Science Department, University of Minnesota, Minneapolis, MN 55455, USA; Department of Computer Science and Engineering, University of California, San Diego, La Jolla, CA 92093

HAESUN PARK hpark@cs.umn.edu
Computer Science Department, University of Minnesota, Minneapolis, MN 55455, USA

JOHN GLICK glick@acusd.edu
Department of Mathematics and Computer Science, University of San Diego, San Diego, CA 92110, USA

LEI ZHANG lzhang@cs.umn.edu
Computer Science Department, University of Minnesota, Minneapolis, MN 55455, USA

Received April 6, 1999; Accepted July 6, 1999

Abstract. It has been known for many years that a robust solution to an overdetermined system of linear equations $Ax \approx b$ is obtained by minimizing the $L1$ norm of the residual error. A correct solution x to the linear system can often be obtained in this way, in spite of large errors (outliers) in some elements of the $(m \times n)$ matrix A and the data vector b. This is in contrast to a least squares solution, where even one large error will typically cause a large error in x. In this paper we give necessary and sufficient conditions that the correct solution is obtained when there are some errors in A and b. Based on the sufficient condition, it is shown that if k rows of $[A\ b]$ contain large errors, the correct solution is guaranteed if $(m - n)/n \geq 2k/\sigma$, where $\sigma > 0$, is a lower bound of singular values related to A. Since m typically represents the number of measurements, this inequality shows how many data points are needed to guarantee a correct solution in the presence of large errors in some of the data. This inequality is, in fact, an upper bound, and computational results are presented, which show that the correct solution will be obtained, with high probability, for much smaller values of $m - n$.

Keywords: overdetermined linear systems, robust approximation, minimum error, zero error conditions, outliers, 1-norm, linear programming

1. Introduction

The benefits of using minimization of the L_1 norm in approximation problems have been known for many years [3–5, 10, 16]. The advantages are likely to be greatest when the problem data contains a relatively small number of large errors (outliers). Examples of applications where the L_1 norm minimization is used to advantage are described in [8, 9, 13, 14, 20]

We present an analysis for the linear, overdetermined system of equations

$$Ax \approx b, \tag{1}$$

where there may be large errors in some elements of the $(m \times n)$ matrix A, $m > n$, and in the data vector $b \in \mathbf{R}^m$.

Overdetermined linear equations arise in many practical applications [2, 6, 7, 15, 22]. A typical such application is the representation of a measured signal $\beta(t)$ as a linear combination of n specified functions

$$\sum_{j=1}^{n} x_j \varphi_j(t_i) = \beta(t_i), \quad i = 1, \ldots, m \tag{2}$$

where the functions $\varphi_j(t)$ are known (but with possible errors) and the output signal $\beta(t)$ is measured (also with possible errors) at $m(>n)$ discrete times t_i, $i = 1, \ldots, m$. The amplitudes x_j are to be determined. Letting $b_i = \beta(t_i)$ and $A_{ij} = \varphi_j(t_i)$ gives the system (1). If there are no errors in A or b, then the correct value x_c of x is easily obtained by choosing any set of n linearly independent rows of A, and solving the resulting $(n \times n)$ system for x_c.

When there are errors, an important practical question is how to choose the number m, of measurements needed to obtain an accurate solution, given n, and the number k of rows of $[A\ b]$ which may contain errors. An answer to this question is given in this paper.

Given a full rank matrix A and vector b, with possible errors in both, it is desired to get a solution vector x as close as possible to the correct (but unknown) vector x_c. A basic assumption is that there is a correct (but unknown) matrix A_c and data vector b_c such that

$$A_c x_c = b_c. \tag{3}$$

A solution x^* is obtained by solving

$$\min_x \|Ax - b\|_1. \tag{4}$$

In the next section we give necessary and sufficient conditions that $x^* = x_c$. The minimum norm solution to (4) is given by

$$x^* = B^{-1} b_B, \tag{5}$$

where B is an $(n \times n)$ basis matrix selected from the rows of A, and b_B consists of the corresponding elements of b. If the n selected rows of A and corresponding elements of b contain no errors, then the result will be that $x^* = x_c$. Clearly, if there are k rows with errors, there must be at least $m \geq n + k$ rows in order to get the correct solution. A sufficient condition which guarantees that such a selection will be made is given in the next section. It is shown there (Theorem 2.1) that $x^* = x_c$, provided $(m - n)/n \geq 2k/\sigma$, where $\sigma > 0$, is a minimum singular value of an $(n \times n)$-submatrix of $B^{-1}A$. This, in fact, gives an upper bound to the value of m needed to guarantee an accurate solution. More practically useful computational results are given in §3. The closely related case when the rows of B and the elements of b_B contain only small errors is also analyzed in the next section.

It is also shown there that the condition is invariant with respect to the magnitude of the errors in b. This property of the L_1 norm minimization is compared to the approximation $\tilde{x}$ obtained from the solution to the least squares problem

$$\min_x \|Ax - b\|_2, \tag{6}$$

where the error in $\tilde{x}$ will typically be proportional to the errors in b.

In order to obtain more practically useful relationships, extensive computational tests were carried out. The results of these tests give an empirical formula and curves which show how many data points (m) are needed to give $x^* = x_c$ with any specified probability P, assuming that there are no more than k rows with large errors. These results are presented in §3. The computational results show that for k rows with errors, $k \leq 50$ and $n \leq 40$, (4) will give $x^* = x_c$ with probability $P \geq 0.995$ when $m - n \geq 22 + 2k$. Note that this value of $m - n$ depends linearly on k, as does the upper bound in Theorem 2.1, but it is substantially smaller, and therefore more useful in practice.

Given an optimal solution to (4), and the corresponding basis B, it is also of interest to determine what changes can be made in b without changing the optimality of the basis B. This situation can be analyzed by a sensitivity analysis of the linear program corresponding to (4). This linear program is given by (31) with $w = x$ and $d = b$. Such an analysis is described in [10].

The use of the L_1 norm minimization has also been investigated for the more difficult parameter estimation problem

$$A(\alpha)x \approx b. \tag{7}$$

For this type of problem it is necessary to determine the parameter vector α as well as the vector x when there are errors in the data vector b. Many signal identification problems can be formulated in this way [1, 11, 12, 18, 19, 21]. The minimization problem now becomes

$$\min_{\alpha,x} \|A(\alpha)x - b\|_1. \tag{8}$$

The theoretical results presented here for (4) are relevant for (8), but are not directly applicable. An iterative algorithm for the solution of (8) is described and analyzed in [18]. It has been applied to the identification of a complex signal with some large errors in the data. The computational results presented in [19] show that, as with the linear case (4), the solution obtained using the L_1 norm is very robust with respect to large data errors.

2. Conditions for correct solution

2.1. Necessary and sufficient conditions

In this section we investigate when the L_1 minimization (4) will give the correct solution x_c when there may be large errors in some rows of A and in some elements of b. A condition is formulated in terms of a scalar β, and a vector d. It is then shown that $\|d\|_1 \leq \beta$ is a

necessary condition and $\|d\|_1 < \beta$ is a sufficient condition that $x^* = x_c$. A closely related situation, which is more likely to occur in practice, is where there may be large errors in some rows of A and in some elements of b, and in addition, there are small errors in all elements of b. A sufficient condition that $x^* \approx x_c$ for this situation is also given.

We are given a full rank $(m \times n)$ matrix A and a data vector $b \in \mathbf{R}^m$. The minimum L_1 norm solution x^* is given by (4). There may be large errors in some rows of A and in some elements of b. However, as given by (3) in §1, we assume that there is a correct matrix A_c and correct data vector b_c such that all m equations are satisfied by a correct vector x_c. The errors in A are represented by the $(m \times n)$ matrix E, and the errors in b by the vector $d_e \in \mathbf{R}^m$, so that

$$\begin{aligned} A &= A_c + E, \\ b &= b_c + d_e. \end{aligned} \tag{9}$$

Also, let w represent the error in x, so that

$$x = x_c + w. \tag{10}$$

Finally, let

$$d = d_e - Ex_c. \tag{11}$$

It follows from (3), (9), (10) and (11) that

$$Ax - b = Aw - d. \tag{12}$$

Therefore, (4) is equivalent to

$$\min_w \|Aw - d\|_1. \tag{13}$$

A linear programming formulation for (13) is given by (31) in §3. It follows directly from the linear programming formulation that the minimum L_1 norm solution to (13) is given by

$$w = B^{-1}d_B, \tag{14}$$

where B is a nonsingular $(n \times n)$ basis matrix consisting of selected rows of A. Each such selection partitions A into B and an $(m - n) \times n$ matrix $\hat{A}$. The vector d is partitioned in a corresponding manner to give $d_B \in \mathbf{R}^n$ and $d_{NB} \in \mathbf{R}^{m-n}$. For each such possible partition, we have

$$\|Aw - d\|_1 = \|\hat{A}B^{-1}d_B - d_{NB}\|_1, \tag{15}$$

since for n of the rows we have $Bw - d_B = 0$.

It follows immediately from (14) that if $d_B = 0$, then $w = 0$ and $x^* = x_c$. For $d_B = 0$, we also have the value of (15) given by $\|d_{NB}\|_1 = \|d\|_1$. If $d_B \neq 0$ for all possible partitions,

then we always get $w \neq 0$. This will always occur if more than $m - n$ elements of d are nonzero.

Directly from (15), the minimization (13) is equivalent to

$$\min_B \|\hat{A}B^{-1}d_B - d_{NB}\|_1. \tag{16}$$

Let $\mathbf{S}$ denote the set of basis matrices B, such that $d_B \neq 0$, and define

$$\beta = \min_{B \in \mathbf{S}} \|\hat{A}B^{-1}d_B - d_{NB}\|_1. \tag{17}$$

If $\|d\|_1 < \beta$, then the minimum for (16) will be attained with $d_B = 0$. If $\|d\|_1 = \beta$, then we may have $d_B \neq 0$, while for $\|d\|_1 > \beta$, we will always have $d_B \neq 0$. This is summarized by the following conditions for $x^* = x_c$:

NC: A necessary condition that $w = 0$ and $x^* = x_c$ is that $\|d\|_1 \leq \beta$.
SC: A sufficient condition that $w = 0$ and $x^* = x_c$ is that $\|d\|_1 < \beta$.

We now show that, for any fixed number k of errors, the number of data points m can be chosen so as to guarantee that the correct solution x_c is computed. To show this, we choose $m = (q + 1)n$, where q is a positive integer. The matrix A is then partitioned so that

$$A^T = \begin{pmatrix} B^T & A_1^T & A_2^T, \ldots, A_q^T \end{pmatrix}. \tag{18}$$

Also let

$$\begin{aligned} \hat{A}^T &= \begin{pmatrix} A_1^T & A_2^T, \ldots, A_q^T \end{pmatrix}, \\ C^T &= \begin{pmatrix} C_1^T & C_2^T, \ldots, C_q^T \end{pmatrix}, \end{aligned} \tag{19}$$

where $C_j = A_j B^{-1}$, $j = 1, \ldots, q$, are $(n \times n)$ matrices. Let σ_j be the minimum singular value of C_j, when B is determined by (17), and assume that $\sigma_j \geq \sigma > 0$, $j = 1, \ldots, q$. Note that if $A_j = B$, then $\sigma_j = 1$. Let k elements of d be nonzero, and for simplicity assume they have the values ± 1.

Theorem 2.1. *A sufficient condition that $x^* = x_c$ is that*

$$\frac{m - n}{n} \geq \frac{2k}{\sigma}. \tag{20}$$

Proof: We have $\|d\|_1 = k$, and for $d_B \neq 0$, $\|d_B\|_2 \geq 1$, and $\|d_{NB}\|_1 \leq k - 1$. We consider $d \neq 0$, since otherwise $x^* = x_c$. From (17),

$$\begin{aligned} \beta &= \min_{B \in \mathbf{S}} \|Cd_B - d_{NB}\|_1 \\ &\geq \min_{B \in \mathbf{S}} \|Cd_B\|_1 - \max_{B \in \mathbf{S}} \|d_{NB}\|_1 \\ &\geq \sum_{j=1}^{q} \|C_j d_B\|_1 - k + 1. \end{aligned} \tag{21}$$

Now, $\|C_j d_B\|_1 \geq \|C_j d_B\|_2 \geq \sigma_j \|d_B\|_2 \geq \sigma$, $j = 1, \ldots, q$. Then from (21)

$$\beta \geq q\sigma - k + 1 > q\sigma - k = \frac{(m-n)\sigma}{n} - k.$$

Therefore, the sufficient condition $\beta > \|d\|_1 = k$ is satisfied whenever (20) is satisfied. □

This result gives an upper bound to the number of data points needed to guarantee a correct solution when there are k errors of ± 1. It can be generalized to k errors δ_i, $i = 1, \ldots, k$, with $\delta_{\min} \leq |\delta_i| \leq \delta_{\max}$, by replacing σ in (20) with $\sigma' = \sigma(\delta_{\min}/\delta_{\max})$.

These conditions are invariant with respect to scaling of the error in b if there is no error in A. For this situation, replacing d_e by λd_e for any finite scalar $\lambda \neq 0$ simply scales both $\|d\|_1$ and β by λ, so that the conditions are unchanged. It is therefore the number and relative size of the nonzero elements in d_e that determines when $x^* = x_c$, independent of the magnitude of $\|d_e\|_1$.

For comparison, consider the situation when the least squares minimization (6) is used. As before, (6) is equivalent to

$$\min_w \|Aw - d\|_2. \tag{22}$$

This has the solution $(A^T A)w = A^T d$, so that

$$\|w\| \geq \|A^T d\| / \|A^T A\|. \tag{23}$$

Therefore, $\|w\| \neq 0$, unless $A^T d = 0$. Furthermore, for no error in A, the magnitude of $\|w\|$ will be proportional to $\|d_e\|$.

2.2. Modified sufficiency conditions

In most practical applications there are likely to be small random errors in all elements of b in addition to possible large errors in only some rows of $[A\ b]$. We now consider this situation and give a sufficient condition such that the difference $\|x^* - x_c\|_1$ will be proportional to the size of the small random errors and will not depend on the large errors.

We give a sufficient condition such that

$$\min_x \|Ax - b\|_1 \tag{24}$$

gives an x with an error depending only on the small random errors in every element of b. Let $b = b_c + d + \epsilon$, where d represents large errors in some elements and ϵ represents small errors in every element, with $\|\epsilon\|_\infty \leq \delta$. As before, $x = x_c + w$, so that the minimization (24) is equivalent to

$$\min_w \|Aw - d - \epsilon\|_1. \tag{25}$$

Let **B** denote the set of nonsingular $(n \times n)$ basis matrices B selected from the rows of A. Corresponding to each $B \in \mathbf{B}$ are the partitions:

$$A = \begin{pmatrix} B \\ \hat{A} \end{pmatrix}, \qquad d = \begin{pmatrix} d_B \\ d_{NB} \end{pmatrix}, \quad \text{and} \quad \epsilon = \begin{pmatrix} \epsilon_B \\ \epsilon_{NB} \end{pmatrix}.$$

The solution to (24) or (25) will give a basis $B_M \in \mathbf{B}$ and corresponding $\hat{A}_M$, d_B and ϵ_B:

$$\begin{aligned} \min_w \|Aw - d - \epsilon\|_1 &= \min_{B \in \mathbf{B}} \|\hat{A}B^{-1}(d_B + \epsilon_B) - d_{NB} - \epsilon_{NB}\|_1 \\ &= \left\|\hat{A}_M B_M^{-1}(d_B + \epsilon_B) - d_{NB} - \epsilon_{NB}\right\|_1, \end{aligned} \tag{26}$$

with $w^* = B^{-1}(d_B + \epsilon_B)$. Also let $\mathbf{S} \subset \mathbf{B}$ denote the subset of basis matrices for which $d_B \neq 0$. Then the related minimization

$$\min_{B \in \mathbf{S}} \|\hat{A}B^{-1}(d_B + \epsilon_B) - d_{NB} - \epsilon_{NB}\|_1 = \left\|\hat{A}_S B_S^{-1}(d_B + \epsilon_B) - d_{NB} - \epsilon_{NB}\right\|_1 \tag{27}$$

gives the matrix B_S^{-1} and corresponding $\hat{A}_S$. Now define the two quantities

$$\rho_M = \left\|\hat{A}_M B_M^{-1}\right\|_1 \quad \text{and} \quad \rho_S = \left\|\hat{A}_S B_S^{-1}\right\|_1.$$

Note that although the $\|\cdot\|_1$ given by (27) is greater than or equal to that given by (26), we may have $\rho_S < \rho_M$.

We want a sufficient condition that (26) gives $d_B = 0$, so that

$$w^* = B_M^{-1}\epsilon_B$$

and

$$\|w^*\|_1 \leq \left\|B_M^{-1}\right\|_1 n\delta. \tag{28}$$

This will be true if the minimum of (26) is obtained with $B \notin \mathbf{S}$, so that from (27) we want

$$\left\|\hat{A}_M B_M^{-1}\epsilon_B - d_{NB} - \epsilon_{NB}\right\|_1 < \left\|\hat{A}_S B_S^{-1}(d_B + \epsilon_B) - d_{NB} - \epsilon_{NB}\right\|_1 \tag{29}$$

The left side of (29) is overestimated by

$$\|d_{NB}\|_1 + \left\|\hat{A}_M B_M^{-1}\epsilon_B\right\|_1 + \|\epsilon_{NB}\|_1 \leq \|d\|_1 + \rho_M n\delta + (m - n)\delta.$$

The right side of (29) is underestimated by

$$\min_{B \in \mathbf{S}} \|\hat{A}B^{-1}d_B - d_{NB}\|_1 - \rho_S n\delta - (m - n)\delta,$$

where we have used the fact that

$$\min_{B\in\mathbf{S}} \|\hat{A}B^{-1}d_B - d_{NB}\|_1 \le \left\|\hat{A}_S B_S^{-1} d_B - d_{NB}\right\|_1.$$

Therefore, a sufficient condition that $\|w^*\|_1$ satisfies (28) is that

$\mathbf{SC}_\delta$: $\|d\|_1 + (\rho_M + \rho_S)n\delta + 2(m-n)\delta < \beta$.

The sufficient condition $\mathbf{SC}_\delta$ reduces to the earlier sufficient condition (**SC**) as $\delta \to 0$. Furthermore, for fixed n, the error norm $\|w^*\|_1$ goes to zero linearly with δ, as given by (28).

3. Computational verification of robust performance

In this section we present the results of extensive computational tests to determine how the problem size (m, n) determines the ability of the L_1 norm minimization to recover x_c, in spite of multiple large errors in the data. Specifically, we investigated the probability P, that $x^* = x_c$, as a function of $m - n$, for specified values of k, the number of rows with large data errors.

In the previous section, the Theorem 2.1 shows how large a value of m is needed in order to guarantee a correct solution, when there are k large errors in the combined data vector d (which includes possible errors in A). The value of m given by (20) is an upper bound, and the actual value which will give $x^* = x_c$, with a very high probability, may be much smaller. In order to determine smaller, more useful requirements for m, a computational study was carried out. The results of this study give a practical basis for deciding how many data points (m) will be needed to obtain the correct solution with a desired probability. Alternatively, for fixed n, m and k, the probability of obtaining a correct solution can be estimated.

The computational study, and its results will now be summarized. A total of 9450 linear programs were solved. Each such linear program solved the problem

$$\min_w \|Aw - d\|_1, \tag{30}$$

where A is $(m \times n)$ and d has $k < m - n$ nonzero elements. The elements a_{ij} of A were randomly generated with $-9 \le a_{ij} \le 9$. The first k elements of d each had the value unity, and the remaining $m - k$ elements were zero. Since the solution to (30) is independent of the row order or of changing the sign of all elements in a given row, this is equivalent to choosing k elements of d as ± 1 in random positions.

The problem (30) is formulated as the following linear program

$$\begin{aligned} \min \quad & \sum_{i=1}^{m} \gamma_i, \\ \text{subject to} \quad & -\gamma \le Aw - d \le \gamma \end{aligned} \tag{31}$$

with $m + n$ variables and $2m$ inequality constraints. The solution is given by $w = B^{-1}d_B$, with

$$\sum_{i=1}^{m} \gamma_i = \|Aw - d\|_1, \qquad \text{and} \quad \gamma_i \geq 0. \tag{32}$$

The desired result, $(x^* = x_c)$, therefore corresponds to $w = 0$.

For each fixed value of (m, n, k) a total of 50 LPs were solved, each with a different matrix A. The probability P that $w = 0$, corresponding to (m, n, k), was then taken to be

$$P = P(m, n, k) = l/50, \tag{33}$$

where l is the number of LPs with the solution $w = 0$.

For each fixed value of n and k, 7 values of m were chosen so that the corresponding 7 values of l covered the range $[0,\ 50]$. For each of the values $n = 10, 20$, a total of 8 values of $k \in [5,\ 40]$ were used, and for $n = 40$, a total of 11 values of $k \in [1,\ 40]$ were used. This gave a total of 9450 LPs solved. The largest of these $(m = 140, n = 40)$ required approximately 10 seconds to solve on a Sun Sparcstation 4. The results obtained are summarized in figures 1–3. The curves presented in figures 1–3 show how the probability P

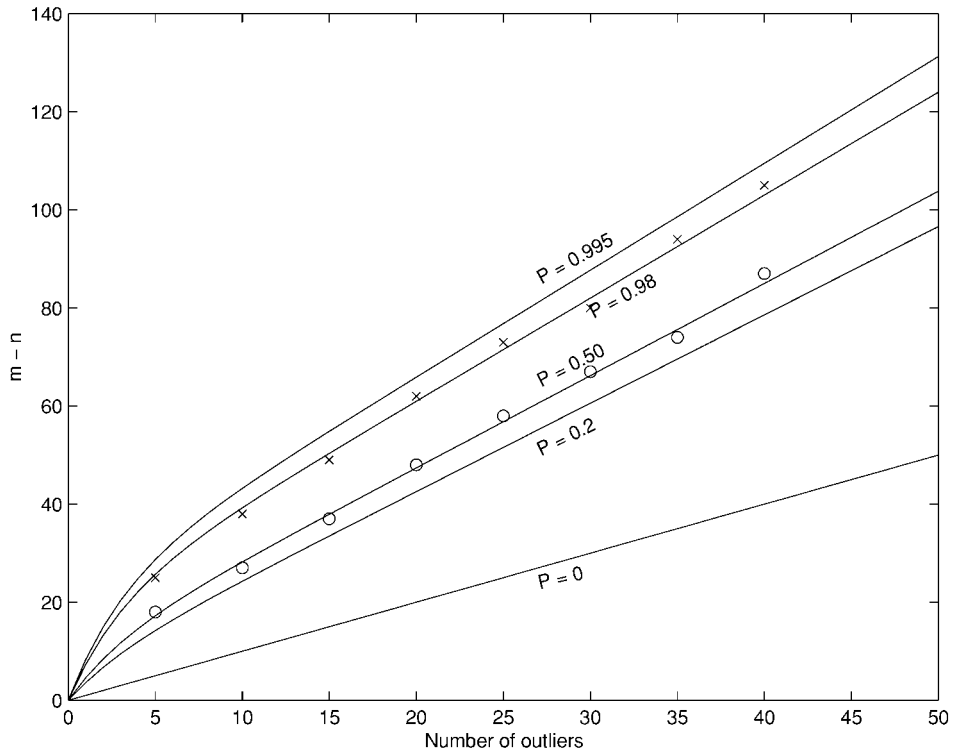

Figure 1. Probability of zero outliers in the basis for $n = 40$.

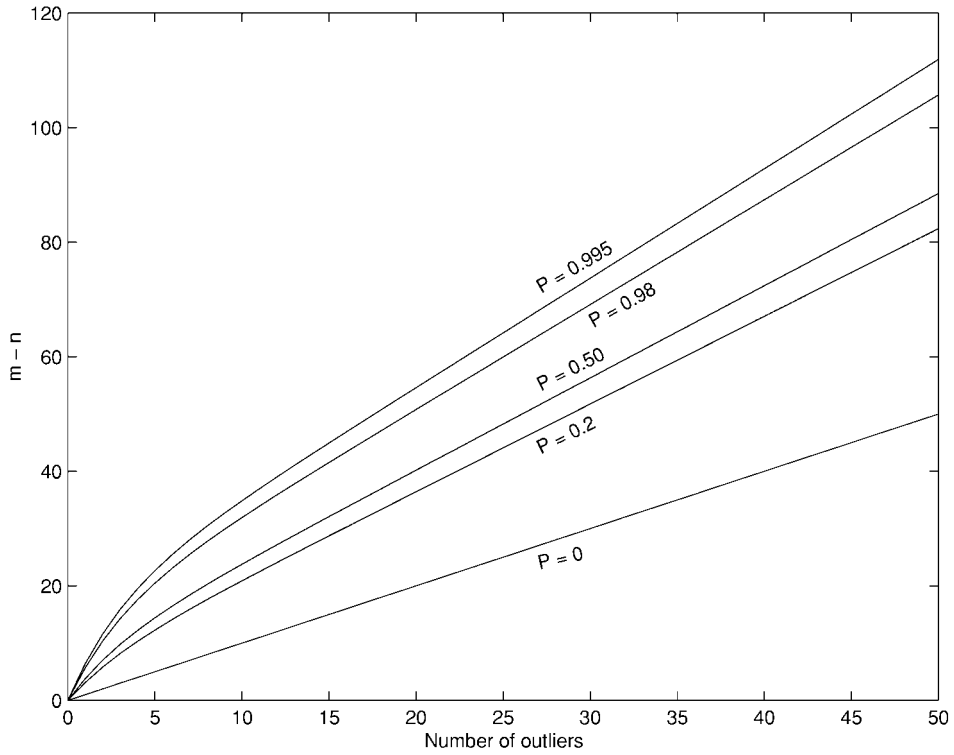

Figure 2. Probability of zero outliers in the basis for $n = 20$.

that $x^* = x_c$ depends on m, n, and k. These curves are valid for $n \in [10,\ 40]$, $k \in [0,\ 40]$, and $m \in [n + k,\ 140]$. They show, in a direct way, the number of data measurements (m) needed to get the correct solution, with probability P, when the measured data may contain k large errors. Any value of $m - n$ above the upper curve will give the correct solution with probability $P \geq 0.995$. It should be noted that P increases rapidly with $m - n$ from $P = 0.5$ to $P = 0.995$ for any fixed k.

These curves are obtained from an empirical formula which gives $m - n$ as a function of P and k. This formula contains 5 parameters, whose values are determined by fitting the computational data. Two of these parameters are independent of n, and the other 3 are slowly varying functions of n, for $n \in [10,\ 40]$.

The empirical formula is valid for $m - n \geq k$, and is given by

$$m - n = \gamma_1\left(1 - e^{\gamma_3 k}\right) + \gamma_2 k + \frac{1}{3.89}\ln\left(\frac{P}{1-P}\right)\left(\Delta\gamma_1\left(1 - e^{\gamma_3 k}\right) + \Delta\gamma_2 k\right). \quad (34)$$

The values of the 5 parameters, as functions of n, are given in Table 1, for $n = 10$, 20, and 40. This formula was chosen so as to have certain important properties corresponding to

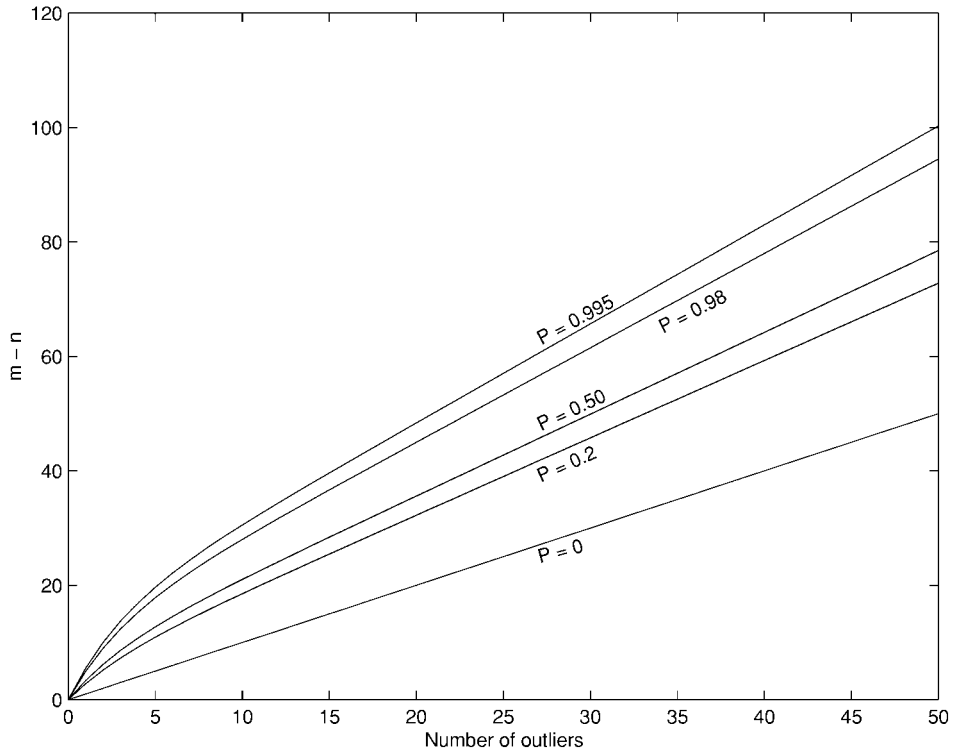

Figure 3. Probability of zero outliers in the basis for $n = 10$.

the computational results. Specifically, for any fixed P, $m - n$ is essentially linear in k, for $\gamma_3 k \geq 3$ or $k \geq 10$. Also, $m - n = 0$ for $k = 0$. Finally, $k \geq 5$ and $m - n \geq 22 + 2k$, gives $P \geq 0.995$.

The parameters γ_1 and γ_2 were determined by setting $P = 0.5$ and fitting the resulting linear function $m - n = \gamma_1 + \gamma_2 k$ to the data for $P = 0.5$ and $k \geq 0$. The values of $\Delta\gamma_1$ and $\Delta\gamma_2$ were then determined to give a best linear fit to the data for $P = 0.98$ and $k \geq 10$. Note that $\ln(\frac{P}{1-P}) \approx 3.89$ for $P = 0.98$, so that for $P = 0.98$ and $k \geq 10$ we have

$$m - n \approx \gamma_1 + \Delta\gamma_1 + (\gamma_2 + \Delta\gamma_2)k. \tag{35}$$

Table 1. Parameters for Eq. (34).

	$n = 10$	$n = 20$	$n = 40$
γ_1	7.0	8.0	9.8
γ_2	1.43	1.61	1.88
γ_3	0.32	0.32	0.32
$\Delta\gamma_1$	5.0	6.2	9.2
$\Delta\gamma_2$	0.22	0.22	0.22

The value of γ_3 was chosen so that $m-n$ is essentially linear in k for $k \geq 10$, but decreases smoothly to zero as k decreases to zero.

The formula (34) can be solved to give P as a function of $m-n$. This gives

$$P = e^{\mu}/(1+e^{\mu}), \tag{36}$$

where $\mu = (m-n-\rho_1)/\rho_2$, $\rho_1 = \gamma_1 + \gamma_2 k$, and $\rho_2 = (\Delta\gamma_1 + \Delta\gamma_2 k)/3.89$. It follows that for any fixed k, $P \to 1.0$ as $m-n$ increases, and $P \to 0$ as $m-n$ decreases. Furthermore using the values in Table 1 we get for $k = m-n \geq 10$ that $P \approx 0.0017$. This shows that (36) closely approximates the required value $P = 0$ when $m-n < k$.

Figures 1–3 make it very easy to determine the number of data points which should be used to get correct solution x_c with probability P, when the number of large errors is believed to be no more than k. For the known n and any desired value of P, the required $m-n$ is given directly by the corresponding curve in figures 1–3. The simplest choice is to use $m-n \geq 22+2k$, which will insure that the probability of a correct solution is at least 0.995.

The results presented here are valid when there are errors in the matrix A, as well as in the vector b. As shown in §2.1, when $A = A_c + E$ and $b = b_c + d_e$, the problem is equivalent to

$$\min_w \|Aw - d\|_1, \tag{37}$$

where $d = d_e - Ex_c$. Therefore, the relevant value of k is the number of nonzero elements in d. Thus k is given by the number of rows in which either A or b has at least one error. The value of k does not depend on the number (greater than one) of errors in a given row of A. Therefore, the value of $m-n$ needed to get $x^* = x_c$ with a specified probability will depend essentially on the number of rows of $[A\ b]$ with error.

Acknowledgment

The work of all four authors was supported in part by the National Science Foundation grant CCR-9509085.

References

1. T.J. Abatzoglou, J.M. Mendel, and G.A. Harada, "The constrained total least squares technique and its application to harmonic superresolution," IEEE Trans. Signal Process., vol. 39, pp. 1070–1087, 1991.
2. H. Barkhuijsen, R. De Beer, and D. Van Ormondt, "Improved algorithm for noniterative time-domain model fitting to exponentially damped magnetic resonance signals," J. Magnetic Resonance, vol. 73, pp. 553–557, 1987.
3. I. Barrodale, "L_1 approximation and the analysis of data," Appl. Stat., vol. 17, pp. 51–57, 1968.
4. I. Barrodale and F.D.K. Roberts, "An improved algorithm for discrete L_1 linear approximation," SIAM J. Numer. Anal., vol. 10, pp. 839–848, 1973.
5. I. Barrodale and A. Young, "Algorithms for best L_1 and L_∞ linear approximations on a discrete set," Numer. Math., vol. 8, pp. 295–306, 1966.

6. A. Bjorck, Numerical Methods for Least Squares Problems, SIAM: Philadelphia, PA, 1996.
7. R. De Beer and D. Van Ormondt, "Analysis of NMR data using time-domain fitting procedures," in In-vivo Magnetic Resonance Spectroscopy I: Probeheads, Radiofrequency Pulses, Spectrum Analysis, NMR Basic Principles and Progress 26, M. Rudin (Ed.), Springer-Verlag: Berlin, 1992, pp. 201–248.
8. T.E. Dielman, "Least absolute value estimation in regression models: An annotated bibliography," Commun. Statistical-Theor. Meth., vol. 13, pp. 513–541, 1984.
9. T.E. Dielman and E.L. Rose, "Forecasting in least absolute value regression with autocorrelated errors: A small sample study," Int. J. Forecast, vol. 10, pp. 539–547, 1994.
10. J. Dupacova, "Robustness of L_1 regression in the light of linear programming," in L_1-statistical Analysis and Related Methods, Y. Dodge (Ed.), North-Holland: Amsterdam, 1992.
11. R. Kumaresan and D.W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," IEEE Trans. on Acoust., Speech, and Signal Process., vol. 30, pp. 833–840, 1982.
12. L. Ljung, System Identification Theory for the User, Prentice-Hall: Englewood Cliffs, NJ, 1987.
13. P.W. Mielke, Jr and K.J. Berry, "Permutation-based multivariate regression analysis: The case for least sum of absolute deviations regression," Ann. Oper. Res., vol. 74, pp. 259–268, 1997.
14. S.C. Narula and J.F. Wellington, "The minimum sum of absolute errors regression: A state of the art survey," Int. Statist. Rev., vol. 50, pp. 317–326, 1982.
15. M.A. Rahman and K.B. Yu, "Total least squares approach for frequency estimation using linear prediction," IEEE Trans. Acous. Speech Signal Process., vol. 35, pp. 1440–1454, 1987.
16. J.R. Rice and J.S. White, "Norms for smoothing and estimation," SIAM Rev., vol. 6, pp. 243–256, 1964.
17. J.B. Rosen, H. Park, and J. Glick, "Total least norm formulation and solution for structured problems," SIAM J. Matrix Anal. Appl., vol. 17, pp. 110–128, 1996.
18. J.B. Rosen, H. Park, and J. Glick, "Structured total least norm for nonlinear problems," SIAM J. Matrix Anal. Appl., vol. 20, pp. 14–30, 1999.
19. J.B. Rosen, H. Park, and J. Glick, "Signal identification using a least L_1 norm algorithm," Optimization & Engineering, vol. 1, pp. 51–65, 2000.
20. P.J. Rousseeuw and A.M. Leroy, Robust Regression and Outlier Detection, John Wiley: New York, 1987.
21. S. Van Huffel, H. Park, and J.B. Rosen, "Formulation and solution of structured total least norm problems for parameter estimation," IEEE Trans. Signal Process., vol. 44, pp. 2464–2474, 1996.
22. S. Van Huffel and J. Vandewalle, The Total Least Squares Problem, Computational Aspects and Analysis, SIAM: Philadelphia, PA, 1991.