Available online at www.sciencedirect.com
ScienceDirect
European Journal of Operational Research 188 (2008) 460–484
EUROPEAN JOURNAL OF OPERATIONAL RESEARCH
www.elsevier.com/locate/ejor

Stochastics and Statistics

Dealing with the multiplicity of solutions of the ℓ_1 and ℓ_∞ regression models

Enrique Castillo [a,b,*], Roberto Mínguez [b], Carmen Castillo [c], Antonio S. Cofiño [a]

[a] Department of Applied Mathematics and Computational Sciences, University of Cantabria, Spain
[b] Department of Applied Mathematics, University of Castilla-La Mancha, Spain
[c] Department of Civil Engineering, University of Castilla-La Mancha, Spain

Received 25 June 2006; accepted 16 April 2007
Available online 21 April 2007

Abstract

It is well known that the least absolute value (ℓ_∞) and the least sum of absolute deviations (ℓ_1) algorithms produce estimators that are not necessarily unique. In this paper it is shown how the set of all solutions of the ℓ_1 and ℓ_∞ regression problems for moderately large sample sizes can be obtained. In addition, if the multiplicity of solutions wants to be avoided, two new methods giving the same optimal ℓ_1 and ℓ_∞ values, but supplying unique solutions, are proposed. The idea consists of using two steps: in the first step the optimal values of the ℓ_1 and ℓ_∞ errors are calculated, and in the second step, in case of non-uniqueness of solutions, one of the multiple solutions is selected according to a different criterion. For the ℓ_∞ the procedure is used sequentially but removing, in each iteration, the data points with maximum absolute residual and adding the corresponding constraints for keeping these residuals, and this process is repeated until no change in the solution is obtained. In this way not only the maximum absolute residual values are minimized in the modified method, but also the maximum absolute residual values of the remaining points sequentially, until no further improvement is possible. In the ℓ_1 case a least squares criterion is used but restricted to the ℓ_1 residual condition. Thus, in the modified ℓ_1 method not only the ℓ_1 residual is minimized, but also the sum of squared residuals subject to the ℓ_1 residual. The methods are illustrated by their application to some well known examples and their performances are tested by some simulations, which show that the lack of uniqueness problem cannot be corrected for some experimental designs by increasing the sample size.

Keywords: Least squares; ℓ_1-norm; ℓ_∞-norm; Parameter estimation; Set of all solutions; Asymptotic properties

1. Introduction

Consider the regression model

$$y = f(\boldsymbol{x}; \boldsymbol{\beta}) + \varepsilon, \tag{1}$$

* Corresponding author. Tel.: +34 942201722; fax: +34 942201703.
E-mail address: castie@unican.es (E. Castillo).

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2007.04.020

where y is the response variable, $\boldsymbol{x}$ is the vector of p predictor variables, $\boldsymbol{\beta}$ is the parameter vector, ε is the error term, and we assume that the Jacobian determinant $|J| = |\partial f(\boldsymbol{x};\boldsymbol{\beta})/\partial\boldsymbol{\beta}|$ is non-null. If $f(x;\boldsymbol{\beta}) = \boldsymbol{X}\boldsymbol{\beta}$ with $\boldsymbol{X}$ the design matrix, we have the linear regression model.

Several estimation methods are used for estimating the parameters of linear and non-linear regression models of the form (1). Among them, the least squares (ℓ_2), the least sum of absolute deviations (ℓ_1) and the least absolute value (ℓ_∞) are the most common (see, for example, Laplace, 1789; Meketon, 1986; Belsley et al., 1980; Bloomfield and Steiger, 1980; Dodge, 1987, 1992, 1997; Dodge and Falconer, 2002; Rao and Toutenburg, 1999 or Chatterjee et al., 2000 for recent papers).

The most popular method for estimating the regression parameters $\boldsymbol{\beta}$ of the models in (1) is the ℓ_2 method, where the sum of squared distances between observed and predicted values is minimized, that is,

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad Z_{\ell_2} = \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta}))^2. \tag{2}$$

Though ℓ_1 and ℓ_∞ methods had initially a great success, they were obscured by the appearance of the ℓ_2 method. Later they recovered some prestige (see Edgeworth, 1887, 1888), when it was discovered that they correspond to maximum likelihood estimators for the uniform and double exponential residuals, respectively, and gave iterative methods for finding the solution, but soon they returned to obscurity mainly due to their associated computational complexities.

A posterior prestige recovery of these methods and a more frequent use of them took place because of the important contribution of Mosteller et al. (1950), who discovered the possibility of stating these problems as linear programming problems. Since then, the advances of mathematical programming were applied to these regression problems and many new results appeared. Recently, Portnoy and Koenker (1997) have shown the interesting result that there are algorithms that make them competitive with the ℓ_2 method, and even superior for some sample sizes.

One important property of the ℓ_1 and ℓ_∞ methods is that they are less sensitive to extreme errors (outliers) than the ℓ_2 method, as already pointed out by Bowditch (see Eisenhart, 1961). Some interesting sensitivity measures are given, for example, in Chatterjee and Hadi (1988), and a recent sensitivity analysis of the three regression methods has been presented by Castillo et al. (2004).

All three regression models can be seen as particular cases of the weighted regression model

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad Z = \sum_{i=1}^{n} w_i |y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})|, \tag{3}$$

where the weights are $w_i = |y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})|$ for the least squares, $w_i = 1$ if $|y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})| = \max_j |y_j - f(\boldsymbol{x}_j;\boldsymbol{\beta})|$ and $w_i = 0$, otherwise, for the ℓ_∞ method, and $w_i = 1$ for the ℓ_1 method. Thus, the least squares method gives much more weight to large residuals, the ℓ_∞ method gives only weight to the maximum residual, and the ℓ_1 method gives equal weight to all residuals. This immediately suggests when each method should be used in a particular application.

1.1. The ℓ_1 regression method

In the ℓ_1 regression problem, the sum of absolute residuals is minimized, i.e.

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad Z_{\ell_1} = \sum_{i=1}^{n} |y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})|. \tag{4}$$

The ℓ_1 method is very old. The first idea seems to be attributed to Boscovich in 1760 (see Stigler, 1984, 1986), who added to (4) the condition of the regression line to pass through the centroid of the mass of points. However, the first written solution to this problem is due to Laplace (1789).

One century later, Edgeworth (1887, 1888) removed the constraint and stated the unconstrained problem. So, we must give the credit for the first proposal of the ℓ_1 method to Edgeworth. He also proposed a numerical method for solving problem (4). However, as indicated by Hawley and Gallagher (1994), his method cycles when the data have some special degeneracies. Other efficient methods for solving this problem were given

by Bloomfield and Steiger (1980) and Wesolowsky (1981). For some detailed analysis of the ℓ_1 method and some applications (see Dodge, 1987, 1992, 1997; Dodge and Falconer, 2002).

It can be shown that the ℓ_1 method is a maximum likelihood approach for the case of a double exponential distribution (see, for example, Yinbo and Arce, 2004).

Since the important work of Mosteller et al. (1950), one can state this problem as a linear programming problem, and nowadays it is well known that the ℓ_∞ problem (4) can be written as

$$\underset{\boldsymbol{\beta},\varepsilon_i}{\text{minimize}} \quad Z_{\ell_1} = \sum_{i=1}^{n} \varepsilon_i \tag{5}$$

subject to

$$y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta}) \leqslant \varepsilon_i; \quad i = 1,\ldots,n, \tag{6}$$

$$f(\boldsymbol{x}_i;\boldsymbol{\beta}) - y_i \leqslant \varepsilon_i; \quad i = 1,\ldots,n, \tag{7}$$

where ε_i are the residuals.

1.2. The ℓ_∞ regression method

This method has a long history. It was proposed even before the least squares method (Euler, 1749) and has also been used in other areas of research as artificial intelligence (see, for example, Castillo et al., 2000a). It can also be seen as a maximum likelihood method for the uniform distribution of errors.

The ℓ_∞ method minimizes the maximum absolute residual, that is,

$$\underset{\boldsymbol{\beta}}{\text{minimize}} \quad Z_{\ell_\infty} = \max_i |y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta})|, \tag{8}$$

which, even though it is an awful model because of the absolute value and maximum functions, can be written as

$$\underset{\boldsymbol{\beta},\varepsilon}{\text{minimize}} \quad Z_{\ell_\infty} = \varepsilon \tag{9}$$

subject to

$$y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta}) \leqslant \varepsilon : \mu_i^{(1)} \quad i = 1,\ldots,n, \tag{10}$$

$$f(\boldsymbol{x}_i;\boldsymbol{\beta}) - y_i \leqslant \varepsilon : \mu_i^{(2)} \quad i = 1,\ldots,n, \tag{11}$$

where ε is the maximum error, and $\mu_i^{(1)}, \mu_i^{(2)};\ i = 1,\ldots,n$ are the dual variables corresponding to constraints (10),(11), respectively (Luenberger, 1989).

Note that this method gives and infinite weight to the largest error. In other words, users of this method are only concerned of the maximum error. This gives an important hint on when this method must be used instead of other alternative regression methods.

Note that the constraints $\varepsilon_i > 0$ and $\varepsilon > 0$ have not been included in the problems (5),(6),(7) and (9),(10),(11), respectively, because these constraints are implied by (6), (7) and (10),(11), respectively (see Castillo et al., 2001). This saves one-third of the number of constraints utilized by many researchers.

1.3. The set of all solutions and the uniqueness problem

The fact that more than one solution exists has been the concern of many researchers and users of these regression methods, because different people using the same model, data and method can obtain different results, that even though they share the same value of the objective function, lead to different models. This concern has led to some proposals for having a unique solution in the past for the ℓ_1 and the ℓ_∞ problems. For example, Giloni and Padberg (2002) proposed the center of gravity of all extreme points of the solution to solve the uniqueness problem. However, this is very costly computationally.

Though uniqueness of solution is in general very convenient, in some cases, one could be more interested in the information about all possible feasible solutions than in one particular solution. Thus, another interesting problem consists of identifying the set of all possible solutions to the ℓ_1 and ℓ_∞ regression problems.

To illustrate the lack of uniqueness and the set of feasible solutions when they exist, the following data set is considered

$$\{(1,7),(1,5),(2,13),(2,9),(3,22),(3,14)\} \tag{12}$$

and the simple regression model

$$y_i = a + bx_i + cx_i^2 + \varepsilon_i, \tag{13}$$

where $\{(x_i, y_i);\ i = 1, 2, \ldots, n\}$ is the sample. Fig. 1 shows this simple set with six data points and the corresponding ℓ_2 regression model, which is unique.

Fig. 2a shows some of the infinitely many optimal ℓ_∞ models. In fact, as it will be shown in Section 3, it shows the four solutions generating, by linear convex combinations, all feasible optimal solutions. Thus, all of them have the same optimal value for the ℓ_∞ error $\varepsilon = 4$. The black squares in the figures show the points defining the four regression models, apart from the data points; they are at a distance of $\varepsilon = 4$ from some particular data points.

Finally, Fig. 2b illustrates the multiplicity of solutions of the ℓ_1 regression model. In fact, as it will be shown in Section 3, it shows the eight solutions generating, by linear convex combinations, all the feasible optimal solutions. Note that the optimal value of the objective function $Z^*_{\ell_1} = \sum_{i=1}^{n} \varepsilon_i = 14$ coincides with the sum of the absolute values of the differences of the ordinates associated with the pairs of points with abscissas 1, 2 and 3.

So, in this paper we do not deal with the properties of the ℓ_1 and ℓ_∞ methods, which have been sufficiently described in the existing literature (see Sielken and Hartley, 1973; Meketon, 1986 or Portnoy and Koenker, 1997, for example). The aim of this paper is to show some uniqueness problems associated with the ℓ_∞ and the ℓ_1 regression methods, and to provide; (a) some methods for obtaining all possible solutions, (b) additional constraints to have unique solutions, and (c) efficient methods to obtain them.

To avoid this multiplicity of solutions without modifying the optimal value of the ℓ_∞ objective function, we propose in Sections 4 and 5 of this paper the sequential ℓ_∞ (SLI) and the revised ℓ_1 methods, which for the above simple example lead to the same parameter estimates as the ℓ_2 method in Fig. 1. The main idea is that if several solutions minimize the objective function, one can benefit by adding extra desirable conditions to chose only one and get uniqueness.

The proposed methods allow us to: (a) obtain a unique solution for the ℓ_1 and ℓ_∞ models, and (b) determine if the ℓ_1 and ℓ_∞ methods have more than one solution for a given set of data.

The proposed methods lead to uniqueness of the modified ℓ_1 and ℓ_∞ methods and can be solved by efficient methods. The fact that extra desirable conditions can be enforced in addition to the minimization of the ℓ_1 and ℓ_∞ errors make them competitive with alternative uniqueness proposals.

So, the present paper tries to propose better alternatives for obtaining this uniqueness. To this end, we benefit from the important results in the recent paper by Giloni and Padberg (2002), who survey traditional regression techniques using the ℓ_1, ℓ_2 and ℓ_∞ norms and discuss uniqueness and coincidence of models. In par-

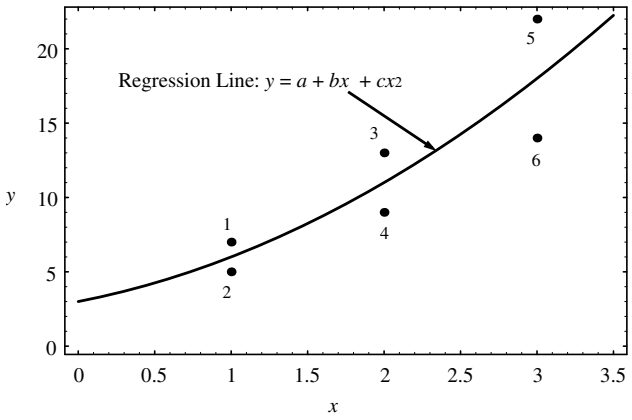

Fig. 1. Scatter plot of y_i versus x_i, for the illustrative example and the resulting ℓ_2 fitted model.

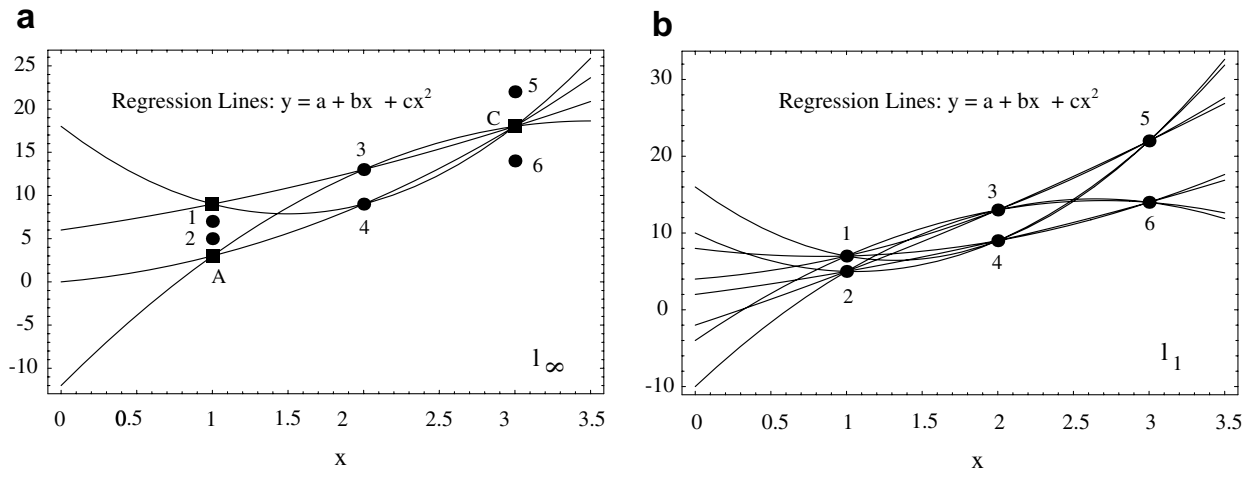

Fig. 2. Scatter plot of y_i versus x_i, for the illustrative example and the regression models generating the infinitely many ℓ_∞ (a) and ℓ_1 (b) resulting models.

ticular they analyze the conditions under which the ℓ_1 and ℓ_∞ solutions coincide, which are relevant to this paper.

The paper is organized as follows. In Section 2 a general method for obtaining the set of all their possible solutions is given. In Section 3 the uniqueness problems of the ℓ_1 and the ℓ_∞ regression methods are described. In Sections 4 and 5 the SLI and the revised ℓ_1 methods aimed at solving the uniqueness problems are introduced, and two algorithms for implementing these methods in a computer are given. Section 6 is devoted to illustrating these methods with the help of some data samples in the existing literature. Section 7 presents some theoretical discussion and the results of a simulation to show the performance of the two methods for small and large samples. Finally, in Section 8, some conclusions are given.

2. Obtaining the set of all solutions for the ℓ_1 and ℓ_∞ methods

In this section we show how the set of all solutions of the problems ℓ_1 and ℓ_∞ can be obtained.

In the case of linear regression, the general solution of the problem (9)–(11) can be obtained by solving the linear system of inequalities (10), (11) and replacing this solution into (9). In addition, since one can know the value of the objective function ε^*, one must add the corresponding constraint $\varepsilon = \varepsilon^*$, because all solutions must share the same objective function value.

Since the methods for solving systems of linear inequalities are not well known by some researchers and references are scarce and difficult to find, we supply the following references: Padberg (1995), Castillo et al. (1999, 2000, 2002), or Castillo and Jubete (2004), where the methods and techniques can be consulted.

The general solution of a system of inequalities is a polyhedron, i.e., the sum of three components: a linear space (generated by linear combinations of vectors), a cone (generated by non-negative linear combinations of vectors) and a polytope (generated by linear convex combinations of vectors). However, in the particular case of the ℓ_1 (6), (7) and ℓ_∞ (10), (11) systems, they lack the linear space component, i.e., they are of the form

$$\begin{pmatrix} \boldsymbol{\beta} \\ \varepsilon \end{pmatrix} = \sum_s \pi_s \begin{pmatrix} \boldsymbol{\beta}^s \\ \varepsilon^s \end{pmatrix} + \sum_j \lambda_j \begin{pmatrix} \boldsymbol{\beta}^j \\ \varepsilon^j \end{pmatrix}; \quad \pi_s \geqslant 0 \ \forall s; \ \lambda_j \geqslant 0 \ \forall j; \ \sum_j \lambda_j = 1, \tag{14}$$

where $\boldsymbol{\beta}^s$ and $\boldsymbol{\beta}^j$ are the generators of the cone and the extremes of the polytope, respectively.

2.1. The ℓ_1 method

Replacing (14) into the objective function (5) one gets

$$Z_{\ell_1} = \sum_{i=1}^{n} \varepsilon_i = \sum_s \pi_s \left(\sum_{i=1}^{n} \varepsilon_i^s \right) + \sum_j \lambda_j \left(\sum_{i=1}^{n} \varepsilon_i^j \right); \quad \pi_s \geqslant 0 \ \forall s; \ \lambda_j \geqslant 0 \ \forall j; \ \sum_j \lambda_j = 1. \tag{15}$$

To obtain the π and λ coefficients leading to a minimum of Z_{ℓ_1}, one must look for the polytope generators with minimum value of $(\sum_{i=1}^{n} \varepsilon_i^s)$, which corresponds to $\pi_s = 0 \ \forall s$ and the non-null λ_js such that $\sum_{i=1}^{n} \varepsilon_i^j$ is a minimum. Let Q^* be the set of j indices where the minimum of $\sum_{i=1}^{n} \varepsilon_i^j$ is attained. If Q^* contains a single element, the problem has a unique solution; otherwise it has multiple solutions corresponding to the polytope generated by such vectors, i.e., the solution is

$$\boldsymbol{\beta} = \sum_{j \in Q^*} \lambda_j \boldsymbol{\beta}^j; \quad \lambda_j \geqslant 0 \ \forall j \in Q^*; \ \sum_{j \in Q^*} \lambda_j = 1. \tag{16}$$

As one example, using the data in (12) and the model in (13), the general solution of the ℓ_1 problem (5)–(7) is the polytope

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 4 & -4 & 16 & 8 & -2 & -10 & 10 & 2 \\ \frac{3}{2} & \frac{27}{2} & -\frac{29}{2} & -\frac{5}{2} & \frac{13}{2} & \frac{37}{2} & -\frac{19}{2} & \frac{5}{2} \\ \frac{3}{2} & -\frac{5}{2} & \frac{11}{2} & \frac{3}{2} & \frac{1}{2} & -\frac{7}{2} & \frac{9}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \\ \lambda_6 \\ \lambda_7 \\ \lambda_8 \end{pmatrix}; \quad \lambda_j \geqslant 0; \ j = 1, 2, \ldots, 8; \ \sum_{j=1}^{8} \lambda_j = 1, \tag{17}$$

Fig. 2b shows the eight generating solutions in (17).

Note that the complexity of the problem of obtaining the solution of a system of inequalities can be very high and increases exponentially with the sample size, so solving the problem using this method for very large samples could be intractable.

In Section 6 one example with $n = 50$, which was obtained without any problem with Mathematica and exact precision (rational numbers), is shown. A much larger sample size can be dealt with using, for example, the GNU Multiple Precision Arithmetic Library (`http://www.swox.com/gmp/`) for any arbitrary precision and rational representation limited by the space.

2.2. The ℓ_∞ method

In this case (14) transforms to

$$\begin{pmatrix} \boldsymbol{\beta} \\ \varepsilon \end{pmatrix} = \sum_s \pi_s \begin{pmatrix} \boldsymbol{\beta}^s \\ \varepsilon^s \end{pmatrix} + \sum_j \lambda_j \begin{pmatrix} \boldsymbol{\beta}^j \\ \varepsilon^j \end{pmatrix}; \quad \pi_s \geqslant 0 \ \forall s; \ \lambda_j \geqslant 0 \ \forall j; \ \sum_j \lambda_j = 1, \tag{18}$$

which replaced into the objective function (9) leads to

$$Z_{\ell_\infty} = \varepsilon = \sum_s \pi_s \varepsilon^s + \sum_j \lambda_j \varepsilon^j; \quad \pi_s \geqslant 0 \ \forall s; \ \lambda_j \geqslant 0 \ \forall j; \ \sum_j \lambda_j = 1. \tag{19}$$

Then, the solution is as in (16) and a similar discussion for the uniqueness problem remains valid.

As one example, the general solution of the linear system of inequalities (10) and (11) for the model (13) and data in (12) is the polyhedron:

$$\begin{pmatrix} a \\ b \\ c \\ \varepsilon \end{pmatrix} = \begin{pmatrix} -\frac{5}{2} & \frac{1}{2} & \frac{1}{2} & \frac{7}{2} & -\frac{7}{2} & -\frac{1}{2} & -\frac{1}{2} & \frac{5}{2} & -12 & 0 & 6 & 18 \\ \frac{5}{2} & 0 & -\frac{3}{2} & -4 & 4 & \frac{3}{2} & 0 & -\frac{5}{2} & \frac{35}{2} & \frac{3}{2} & \frac{5}{2} & -\frac{27}{2} \\ -\frac{1}{2} & 0 & \frac{1}{2} & 1 & -1 & -\frac{1}{2} & 0 & \frac{1}{2} & -\frac{5}{2} & \frac{3}{2} & \frac{1}{2} & \frac{9}{2} \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & 4 & 4 & 4 & 4 \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \pi_3 \\ \pi_4 \\ \pi_5 \\ \pi_6 \\ \pi_7 \\ \pi_8\lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix}, \tag{20}$$

where $\pi_i \geqslant 0;\ i = 1, 2, \ldots, 8;\ \lambda_j \geqslant 0;\ j = 1, 2, 3, 4;\ \sum_{j=1}^{4} \lambda_j = 1$.

Replacing now this solution into (9) one gets

$$Z_{\ell_\infty} = \varepsilon = \frac{1}{2}(\pi_1 + \pi_2 + \cdots + \pi_8) + 4(\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4) = \frac{1}{2}(\pi_1 + \pi_2 + \cdots + \pi_8) + 4,$$

which obviously attains a minimum for $\pi_i = 0;\ i = 1, 2, \ldots, 8$. Then, the set of all solutions of the general ℓ_∞ problem (9)–(11), obtained from (20) is:

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} -12 & 0 & 6 & 18 \\ \frac{35}{2} & \frac{3}{2} & \frac{5}{2} & -\frac{27}{2} \\ -\frac{5}{2} & \frac{3}{2} & \frac{1}{2} & \frac{9}{2} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix}; \quad \lambda_j \geqslant 0;\ j = 1, 2, 3, 4;\ \sum_{j=1}^{4} \lambda_j = 1, \tag{21}$$

which is a polytope. Fig. 2(a) shows the four generating solutions in (21).

3. Uniqueness of the ℓ_1 and ℓ_∞ problems

It is well known that the ℓ_2 method shows a good behavior in linear and in some non-linear regression models because it has a unique solution. However, the ℓ_1 and ℓ_∞ regression methods, due to their lack of uniqueness, can have problems. For example, Sielken and Hartley (1973) pointed out the lack of uniqueness problem, and proposed an algorithm for obtaining unbiased estimates when the solution is unique. However, they did not solved the uniqueness problem.

For example, the ℓ_1 estimator of a model consisting of only an intercept is the sample median. When there is an even number of observations, any value in the closed interval defined by the two middle observations would serve as a ℓ_1 estimate. In order to solve the regression problem one must select one of the resulting optimal ℓ_1 models.

As it is shown in Section 6, the uniqueness problem does not disappear increasing the sample size and it is more related to the experimental design (repeated or random samples) than to the sample size itself. More precisely, the experiments presented in Section 6 show the interesting (and surprising?) result that the uniqueness problem can become worse when increasing sample sizes.

If the ℓ_∞ problem has more than one solution, it is convenient to provide a way to select one of them satisfying some extra conditions. However, before proceeding to this selection it is good to have a way to check the uniqueness. Apart from the methods given in Section 2 to check uniqueness of the ℓ_1 and ℓ_∞ methods, there is an interesting proposition in Giloni and Padberg (2002), which allows checking uniqueness once the ℓ_∞ problem has been solved. The proposition is as follows:

Proposition 1 (*Uniqueness of the ℓ_∞-problem*). (Giloni and Padberg, 2002). *Let β^* be an extreme point solution of the ℓ_∞-problem in (9)–(11). Let A be the set*

$$A = \{i \in N; |r_i| = \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^*\|_\infty\},$$

let

$$f_i = \begin{cases} 1 & \text{if} \quad r_i \geqslant 0, \\ -1 & \text{if} \quad r_i < 0, \end{cases}$$

and $\boldsymbol{F}_A$ *be the diagonal matrix with diagonal elements* $f_i; i \in A$, *and* $\boldsymbol{X}_A$ *be the submatrix of* $\boldsymbol{X}$ *containing the rows which indices are in* A.

Then, $\boldsymbol{\beta}^*$ *is unique if and only if the system of inequalities in* $\boldsymbol{\xi}$

$$\boldsymbol{F}_A^* \boldsymbol{X}_A \boldsymbol{\xi} \geqslant 0; \quad \xi_k < 0 \tag{22}$$

is inconsistent for all $k = 1, 2, \ldots, p$.

Though it can appear that Proposition 1 leads to a too laborious computation because it implies checking the compatibility of many systems, it can be done much easily, as suggested by the following proposition.

Proposition 2 (Uniqueness of the ℓ_∞-problem). *Under the conditions of Proposition* 1, β^* *is unique if and only if the set of solutions of* $\boldsymbol{F}_A^* \boldsymbol{X}_A \boldsymbol{\xi} \geqslant 0$ *is an acute cone and all the components of its edges* (*generators*) *are strictly positive.*

Proof 1. The proof of this proposition is immediate if one uses the Γ-algorithm (see Castillo and Jubete, 2004) to solve the system of inequations of the form

$$\boldsymbol{F}_A^* \boldsymbol{X}_A \boldsymbol{\xi} \geqslant 0; \quad \xi_k < 0.$$

In a first step, system $\boldsymbol{F}_A^* \boldsymbol{X}_A \boldsymbol{\xi} \geqslant 0$ is solved and the dual cone of the cone generated by the rows of $\boldsymbol{F}_A^* \boldsymbol{X}_A$ is obtained, then, in the last step, the vector associated with the constraint $\xi_k < 0$ is introduced. Thus, in order to get no solution (incompatibility), the kth components of the generators must be strictly positive, because otherwise, a solution is obtained. Since this must hold for all k, the result in the proposition holds. □

The practical importance of this proposition is that to check uniqueness, one needs only to solve the system $\boldsymbol{F}_A^* \boldsymbol{X}_A \boldsymbol{\xi} \geqslant 0$. This is illustrated in the following example.

Example 1. Consider the data and the simple regression model in (12) and (13), respectively. One ℓ_∞ solution is the model in Fig. 2(a) passing through points A, 4 and C, which data estimated values and residuals are given in Table 1. Then one has

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 3 & 9 \end{pmatrix},$$

and $A \equiv \{1, 3, 5, 6\}$, $\boldsymbol{F}_A^* = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$ and $\boldsymbol{X}_A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 3 & 9 \end{pmatrix}$, and Eq. (22) becomes

Table 1
Data, estimated y-values, residuals and f-values, for the example in Fig. 1

i	x_i	y_i	$\hat{y}_i$	r_i	f_i
1	1	7	3	4	1
2	1	5	3	2	1
3	2	13	9	4	1
4	2	9	9	0	1
5	3	22	18	4	1
6	3	14	18	−4	−1

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 3 & 9 \end{pmatrix} \boldsymbol{\xi} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ -1 & -3 & -9 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} \geqslant 0; \quad \xi_k < 0, \ k = 1,2,3.$$

The cone $\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ -1 & -3 & -9 \end{pmatrix} \begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} \geqslant 0$ can be written as

$$\begin{pmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} = \pi_1 \begin{pmatrix} -3 \\ 4 \\ -1 \end{pmatrix} + \pi_2 \begin{pmatrix} 6 \\ -5 \\ 1 \end{pmatrix}; \quad \pi_1, \pi_2 \in \mathbb{R}^+,$$

which is not an acute cone, and then it is obvious that the system (22) is compatible for $k = 1,2,3$. Thus, this ℓ_∞ problem has not a unique solution.

Similarly, Giloni and Padberg (2002), provide a proposition that allows checking uniqueness once the ℓ_1 problem has been solved. This proposition uses the fact that for any optimal extreme point solution of the ℓ_1-problem, there exists a non-singular $p \times p$ submatrix $\boldsymbol{X}_B$ of $\boldsymbol{X}$ such that

$$\boldsymbol{\beta}^* = \boldsymbol{X}_B^{-1} \boldsymbol{y}^B,$$

i.e., the regression surface passes through p points and the parameters of the model can be calculated from them.

The proposition is as follows:

Proposition 3 (*Uniqueness of the ℓ_1-problem*). (Giloni and Padberg, 2002).*Let β^* be a solution of the ℓ_1-problem in (5)–(7).*

Then β^* is unique if and only if

$$-\boldsymbol{e}_p^{\mathrm{T}} - \sum_{i \in D} |\boldsymbol{x}^i \boldsymbol{X}_B^{-1}| < (\boldsymbol{e}_U^{\mathrm{T}} \boldsymbol{X}_U - \boldsymbol{e}_L^{\mathrm{T}} \boldsymbol{X}_L) \boldsymbol{X}_B^{-1} < \boldsymbol{e}_p^{\mathrm{T}} + \sum_{i \in D} |\boldsymbol{x}^i \boldsymbol{X}_B^{-1}|, \tag{23}$$

where e_p, e_U and e_L are vectors with all unit components, $D = Z - B$, $|\boldsymbol{x}| = (|x_1|, |x_2|, \ldots, |x_p|)^{\mathrm{T}}$, Z, U and L are the sets

$$Z \equiv \{i \in N; r_i = 0\}; \quad U \equiv \{i \in N; r_i > 0\}; \quad L \equiv \{i \in N; r_i < 0\}$$

and x_i is the row i of X.

This proposition, once the solution to the ℓ_1 problem is known allows testing its uniqueness of solution. Note that in order to apply it, the sets Z, U, L and B, which cannot be identified without knowledge of one solution, must be known.

This proposition explains also that grouping data (coincidence of the predicting variable values) is not convenient under the point of view of uniqueness, because it leads to large values of the term $(\boldsymbol{e}_U^{\mathrm{T}} \boldsymbol{X}_U - \boldsymbol{e}_L^{\mathrm{T}} \boldsymbol{X}_L) \boldsymbol{X}_B^{-1}$ for extreme point solutions and makes it difficult to satisfy (23), i.e., the uniqueness (see the example below).

Example 2. Consider again the data and the model in (12) and (13), respectively, and the extreme point solution to the ℓ_1 problem passing through points 1, 4 and 5 in Fig. 2b. Then, the following sets and matrices result (see Table 2)

$$Z \equiv B \equiv \{1,4,5\}; \quad U \equiv \{3\}; \quad L \equiv \{2,6\}; \quad \boldsymbol{X}_U = (1 \quad 2 \quad 4); \quad \boldsymbol{X}_L = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 9 \end{pmatrix}$$

Table 2
Data, estimated y-values, residuals and f values, for the example in Example 2

i	x_i	y_i	$\hat{y}_i$	r_i
1	1	7	7	0
2	1	5	7	−2
3	2	13	9	4
4	2	9	9	0
5	3	22	22	0
6	3	14	22	−8

and

$$\boldsymbol{X}_B = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}; \quad \boldsymbol{\beta}^* = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}^{-1} \begin{pmatrix} 7 \\ 9 \\ 22 \end{pmatrix}$$

and then, since $D \equiv \emptyset$, condition (23) becomes

$$(-1 \quad -1 \quad -1) \not< 1(1 \quad 2 \quad 4) - (1 \quad 1) \begin{pmatrix} 1 & 1 & 1 \\ 1 & 3 & 9 \end{pmatrix} < (1 \quad 1 \quad 1),$$

that is,

$$(-1 \quad -1 \quad -1) \not< (-1 \quad -2 \quad -6) < (1 \quad 1 \quad 1),$$

which proves the non-uniqueness of the solution.

4. The SLI method

If the ℓ_∞ has more than one solution, it is convenient to provide a way to select one of them satisfying some extra conditions. In this section we present the sequential ℓ_∞ model. The name sequential ℓ_∞ (SLI) comes from the fact that the ℓ_∞ method is applied sequentially to different sets of decreasing number of data points and an increasing number of constraints are considered to keep the residuals of the remaining points at their corresponding values. In this way not only the maximum absolute residual values are minimized, but also the absolute values of the residuals of the remaining points sequentially, until no further improvement is possible.

The method is described in detail in the following algorithm.

Algorithm 1 (*The SLI model*). The SLI estimation method is as follows:

Input. A data set and a parametric regression model.
Output. The estimates of the corresponding parameters using the SLI method.

Step 1: **Solve the ℓ_∞ problem**. The standard ℓ_∞ problem is solved using all data points.
Step 2: **Find the set of data points that leads to the ℓ_∞ error**. To this end, we find the set of data associated with the non-null dual variables ($\mu^{(1)}$ and $\mu^{(2)}$), which correspond to the data points whose absolute values of the residuals coincide with the actual value of the objective ℓ_∞ function for all existing solutions of the ℓ_∞ problem.
Step 3: **Remove points from the actual data set.** This set of data points are removed from the actual data set (sample).

Step 4: **Test for convergence.** Compare the actual solution with the previous solution. If there is coincidence or there are no points remaining in the actual sample, stop the process and return the parameter estimates. Otherwise continue with Step 5.

Step 5: **Add the corresponding residual constraints.** Add to the optimization problem one constraint per removed data point forcing the corresponding residuals to remain constant and equal to the actual optimal value of the objective function.

Step 6: **Solve the constrained ℓ_∞ problem.** The standard ℓ_∞ problem is solved again but considering only the actual data set and the actual constraints. Next, go to Step 2.

Note that this algorithm allows us to determine if the standard ℓ_∞ method has no unique solution. This occurs when the solutions of the ℓ_∞ and the SLI methods do not coincide, i.e., when the algorithm stops later than the second iteration.

Since the analytical definition of the SLI estimates is complicated and involves iteration and optimization problems, it is practically impossible to give closed formulas for the confidence intervals or the variances of the estimates, even for the asymptotic case. However, the bootstrap method can be easily used to derive these variances and confidence intervals. Note that the bootstrap method cannot behave well for the ℓ_∞ estimators, because of its lack of uniqueness, however, the uniqueness of the SLI solves this problem.

4.1. Illustrative example

To illustrate the proposed method, the example given in Section 1 for the regression model is used

$$y_i = a + bx_i + cx_i^2 + \varepsilon_i.$$

The SLI algorithm in this case proceeds as follows:

First iteration

Step 1: **Solve the initial ℓ_∞ problem.** The standard ℓ_∞ problem is solved with all data points and an optimal value $\varepsilon_1^* = 4$, where the subindex refers to the iteration number, and the following estimates and non-null values of the dual variables are obtained

$$\hat{a}_1 = 0; \quad \hat{b}_1 = 1.5; \quad \hat{c}_1 = 1.5; \quad \mu_6^{(1)} = -0.5; \quad \mu_5^{(2)} = -0.5. \tag{24}$$

Step 2: **Find the set of data points that leads to the ℓ_∞ error**. The data points with absolute value of the residual $|\varepsilon_i| = 4$ and non-null dual variables are data points 5 and 6.

Step 3: **Remove points from the actual data set.** Data points 5 and 6 are removed, so that the actual data set uses cases $\{1, 2, 3, 4\}$.

Step 4: **Test for convergence.** Since it is the first iteration, we continue.

Step 5: **Add the corresponding residual constraints.** We add the two constraints

$$y_5 - a_1 - b_1 x_5 - c_1 x_5^2 = -4; \quad y_6 - a_1 - b_1 x_6 - c_1 x_6^2 = 4.$$

Second iteration

Step 1: **Solve the constrained ℓ_∞ problem.** The standard ℓ_∞ problem is solved again but considering only the actual set of data and the actual constraints. After this process we get an optimal value $\varepsilon_2^* = 2$ and the following estimates and non-null values of the dual variables are obtained

$$\hat{a}_2 = 1.2; \quad \hat{b}_2 = 3.5; \quad \hat{c}_2 = 0.7; \quad \mu_4^{(1)} = -0.5; \quad \mu_3^{(2)} = -0.5. \tag{25}$$

Step 2: **Find the set of data points that leads to the ℓ_∞ error.** The data points with absolute value of the residual $|\varepsilon_i| = 2$ and non-null dual variables are data points 3 and 4.

Step 3: **Remove points from the actual data set.** Data points 3 and 4 are removed, so that the actual data set uses cases $\{1, 2\}$.

Step 4: **Test for convergence.** Since the process has not converged yet, we continue.

Step 5: **Add the corresponding residual constraints.** We add the constraints

$$y_3 - a_2 - b_2x_3 - c_2x_3^2 = -2; \quad y_4 - a_2 - b_2x_4 - c_2x_4^2 = 2.$$

Third iteration

Step 1: **Solve the constrained ℓ_∞ problem.** The standard ℓ_∞ problem is solved again but considering only the actual set of data and the actual constraints. After this process we get an optimal value $\varepsilon_3^* = 1$ and the following estimates and non-null values of the dual variables are obtained

$$\hat{a}_3 = 3; \quad \hat{b}_3 = 2; \quad \hat{c}_3 = 1; \quad \mu_2^{(1)} = -0.5; \quad \mu_1^{(2)} = -0.5. \tag{26}$$

Step 2: **Find the set of data points that leads to the ℓ_∞ error.** The data points with absolute value of the residual $|\varepsilon_i| = 1$ and non-null dual variables are data points 1 and 2.

Step 3: **Remove points from the actual data set.** Data points 1 and 2 are removed, so that the actual data set becomes empty $\{\}$.

Step 4: **Test for convergence.** Since the actual data set is empty, we stop the process and return the actual estimates:

$$\hat{a}_3 = 3; \quad \hat{b}_3 = 2; \quad \hat{c}_3 = 1. \tag{27}$$

Note that these estimates coincide with those of the ℓ_2 method.

Fig. 3 shows the three regression models corresponding to the three iterations.

5. The revised ℓ_1 regression method

In this section we present the revised least sum of absolute deviations (RLSAD), i.e., the ℓ_1 regression model, which is described in detail in the following algorithm.

Algorithm 2 (*The revised ℓ_1 model*). The revised ℓ_1 estimation method is as follows:

Input. A data set and a parametric regression model.
Output. The estimates of the corresponding parameters using the revised ℓ_1 regression method.

Step 1: **Solve the initial ℓ_1 problem.** The standard ℓ_1 problem is solved using all data points and the resulting ℓ_1 error δ is stored, i.e., solve the problem:

$$\underset{\boldsymbol{\beta},\boldsymbol{\varepsilon}}{\text{minimize}} \quad Z_{\ell_1} = \sum_{i=1}^{n} \varepsilon_i \tag{28}$$

subject to

$$y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta}) \leqslant \varepsilon_i; \quad i = 1,\ldots,n, \tag{29}$$

$$f(\boldsymbol{x}_i;\boldsymbol{\beta}) - y_i \leqslant \varepsilon_i; \quad i = 1,\ldots,n, \tag{30}$$

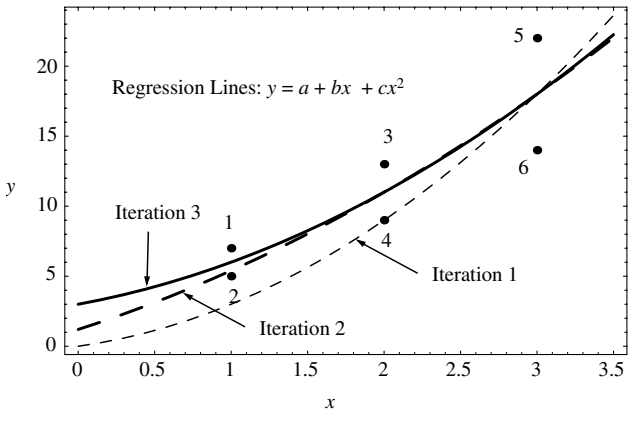

Fig. 3. ℓ_∞ models resulting in the three iterations of the illustrative example.

and store the optimal value $\delta^* = \sum_{i=1}^{n} \varepsilon_i^*$, where ε_i^* are the residuals corresponding to the optimal estimates, and ε is the vector containing all ε_i.

Step 2: **Solve the least squares regression problem subject to the above ℓ_1 error.** In other words, solve the problem:

$$\underset{\boldsymbol{\beta},\boldsymbol{\varepsilon}}{\text{minimize}} \quad Z_{\ell_2} = \sum_{i=1}^{n} (y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta}))^2, \tag{31}$$

subject to

$$y_i - f(\boldsymbol{x}_i;\boldsymbol{\beta}) \leqslant \varepsilon_i; \quad i = 1,\ldots,n, \tag{32}$$

$$f(\boldsymbol{x}_i;\boldsymbol{\beta}) - y_i \leqslant \varepsilon_i; \quad i = 1,\ldots,n, \tag{33}$$

$$\sum_{i=1}^{n} \varepsilon_i = \delta^*. \tag{34}$$

Note that this algorithm allows us to determine if the standard ℓ_1 method has no unique solution. This occurs when the solutions of the ℓ_1 and the RLSAD methods do not coincide.

Since the analytical definition of the revised ℓ_1 estimates involves two optimization problems, it seems difficult to give formulas for confidence intervals or variances of the estimates, even for the asymptotic case. Fortunately, since all ℓ_1 solutions are asymptotically equivalent, the standard ℓ_1 theory applies, because the RLSAD solution is a ℓ_1 solution. Alternatively, the bootstrap method can be easily used to derive these variances and confidence intervals for small samples.

The proposed method can fail if the least squares estimate is also an ℓ_1 estimate. In such a case, it might be possible that the solutions of the ℓ_1 and RLSAD problems coincide and yet the ℓ_1 problem may not have a unique solution. To understand how and when the RLSAD helps to solve the problem, we use the following proposition from Giloni and Padberg (2002):

Proposition 4 (Coincidence of solutions of the ℓ_2 and ℓ_1 problems (Giloni and Padberg, 2002)). *The ℓ_2 estimate β is an optimal solution of the ℓ_1 problem if and only if there exist $v \in \mathbb{R}^{|Z|}$ such that*

$$\boldsymbol{v}\boldsymbol{X}_Z = -\boldsymbol{e}_U^{\mathrm{T}}\boldsymbol{X}_U + \boldsymbol{e}_L^{\mathrm{T}}\boldsymbol{X}_L; \quad -\boldsymbol{e}_Z^{\mathrm{T}} \leqslant \boldsymbol{v} \leqslant \boldsymbol{e}_Z^{\mathrm{T}}. \tag{35}$$

If $Z = \emptyset$, condition (35) simplifies to

$$\boldsymbol{e}_U^{\mathrm{T}}\boldsymbol{X}_U = \boldsymbol{e}_L^{\mathrm{T}}\boldsymbol{X}_L. \tag{36}$$

This proposition allows us not only checking particular cases, but also choosing a $\boldsymbol{X}$ design matrix such that the uniqueness of the proposed algorithm be guaranteed.

In conclusion, apart from very seldom cases in which (35) or (36) hold, Proposition 4 guarantees the uniqueness of the proposed method.

The following example illustrates the use of Proposition 4.

Example 3. Consider again the data and the model in (12) and (13), respectively. To check the coincidence of the regression model with one of the ℓ_1 models, we use Proposition 4, and since we have $Z = \emptyset$, we need to check (36).

For this case we have

$$\boldsymbol{X} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix}; \quad \boldsymbol{X}_L = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix},$$

which implies

$$(1 \quad 1 \quad 1)\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix} = (1 \quad 1 \quad 1)\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{pmatrix},$$

i.e., that the ℓ_2 solution is also a ℓ_1 solution.

This is in agreement with the solution given by the algorithm. In effect, the solution to the problem (28)–(30) in Step 1 for this case is not unique and leads to $\delta^* = Z^*_{\ell_1} = 14$, and the solution to the constrained least squares problem (31)–(34) leads to $Z_{\ell_2} = 42$, and the regression line coincides with the unique unconstrained least squares regression line.

Proposition 4 is also useful to explain how the design matrix $\boldsymbol{X}$ plays an important role in the uniqueness problem. A coincidence of the values of the predictive variables $\boldsymbol{X}$, that is, grouping data, leads to coincidence of the corresponding rows of the matrix $\boldsymbol{X}$, and to redundant equations in the system of equations (35) and (36), which makes the solution of these system possible, and consequently to the coincidence of solutions of the ℓ_2 and ℓ_1 problems. On the contrary, a Random selection of the predictive variables $\boldsymbol{X}$ makes the solution of (35) or (36) difficult or impossible and then it leads to no coincidence of solutions. This is illustrated in the simulation examples in the following section. Thus, Proposition 4 allows us choosing a design matrix $\boldsymbol{X}$ such that the coincidence is impossible.

6. Examples of application

In this section we apply the SLI and the RLSAD methods to some real examples.

6.1. A model for ultrasonic data

The data used here is the result of a NIST study involving ultrasonic calibration. The data consists of 54 observations on two variables. The response variable (y) is ultrasonic response and the predictor variable (x) is metal distance.

The data can be found in http://www.itl.nist.gov/div898/strd/nls/data/LINKS/DATA/Chwirut2.dat.

In this website, a non-linear regression model of the form

$$y_i = f(\boldsymbol{x}_i;\boldsymbol{\beta}) + \varepsilon_i = \frac{\mathrm{e}^{-\beta_1 x_i}}{\beta_2 + \beta_3 x_i} + \varepsilon_i, \quad i = 1, 2 \ldots, n \tag{37}$$

was fitted to the data.

This is not a regular case and the ℓ_∞ regression problem has infinitely many solutions. So, looking for uniqueness the SLI method is used.

The optimal value of the ℓ_∞ error bound is 8.55, which is attained at points 1 and 28. The second ℓ_∞ bound (after removing these points) is 8.003, which is attained at points 6, 36 and 39. The third ℓ_∞ bound (after removing these points) is 7.703, which is attained at point 29. Since the parameter values coincide with the previous iteration ones, the SLI reaches the solution at the third iteration, which proves that the ℓ_∞ estimators do not have a unique solution.

The following ℓ_2 and SLI estimates for the ultrasonic data are obtained:

$$\begin{aligned}
&\ell_2: \quad Z_{\ell_2} = 513.05, \quad \hat{\beta}_1 = 0.16658, \quad \hat{\beta}_2 = 0.00517, \quad \hat{\beta}_3 = 0.01215,\\
&\text{SLI}: \quad Z_{\text{SLI}} = 8.55, \quad \hat{\beta}_1 = -0.01029, \quad \hat{\beta}_2 = 0.00361, \quad \hat{\beta}_3 = 0.01662.
\end{aligned}$$

Fig. 4 shows the ℓ_2 (upper left plot) and the SLI (upper right plot) regression models for the ultrasonic calibration data. In addition to the regression lines, the lower and the upper bands (dashed lines) at a distance $\varepsilon = \pm 8.55$, i.e., the ℓ_∞ error, for the ℓ_2 and sequential ℓ_∞ estimates, are shown.

Note that data point 1 is outside the upper band for the least squares model and that the ℓ_2 method gets a better fit of the data points on the right while the SLI method gets a better fit with respect to the ℓ_∞ fit of the data points on the left.

The RLSAD estimates are

$$Z_{\text{LSAD}} = 105.49, \quad Z_{\ell_2} = 521.16, \quad \hat{\beta}_1 = 0.15110, \quad \hat{\beta}_2 = 0.00499, \quad \hat{\beta}_3 = 0.01283,$$

which coincide with the ℓ_1 estimates. The RLSAD model is shown in Fig. 4 (lower plot).

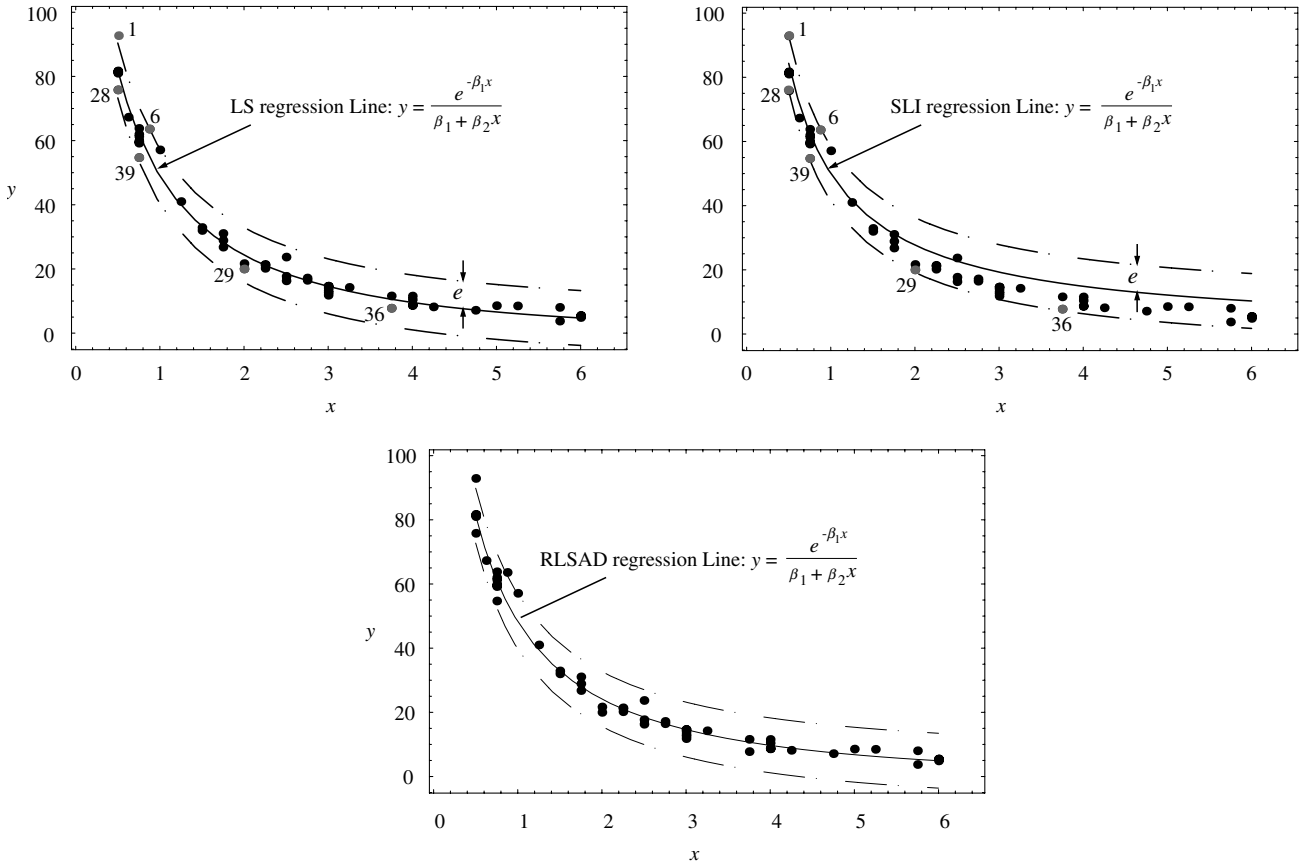

Fig. 4. Scatter plots of y_i versus x_i, for the ultrasonic calibration data set and the ℓ_2 (upper left plot), the SLI fitted models (upper right plot) and the RLSAD (lower plot). Lower and upper bands (dashed lines) at a distance $\varepsilon = \pm 8.55$, i.e., the ℓ_∞ error, for the ℓ_2, sequential ℓ_∞ and RLSAD models are shown.

6.2. A model for calibration data

According to Seber and Wild (1989), Tiede and Pagano (1979) fit the calibration model

$$y_i = \alpha + \beta(1 + \gamma x_i^{\delta})^{-1} + \varepsilon_i, \quad i = 1, 2, \ldots, 14; \quad \alpha, \beta, \gamma, \delta > 0, \tag{38}$$

to the set of data in Table 3. As can be seen from the scatter plot of y_i versus x_i in Fig. 5, observation 9 is a clear outlier.

The optimal value of the ℓ_∞ error bound is 1.041, which is attained at points 9 and 10. The second ℓ_∞ bound (after removing these points) is 1.012, which is attained at points 1, 4, 8 and 11. The third ℓ_∞ bound (after removing these points) is 0.937, which is attained at point 12. Since the parameter values coincide with

Table 3
Calibration data

i	x_i	y_i	i	x_i	y_i
1	0	7.720	8	10	3.208
2	0	8.113	9	20	4.478
3	2	6.664	10	20	2.396
4	2	6.801	11	50	1.302
5	5	4.994	12	50	1.377
6	5	4.948	13	100	1.025
7	10	3.410	14	100	1.096

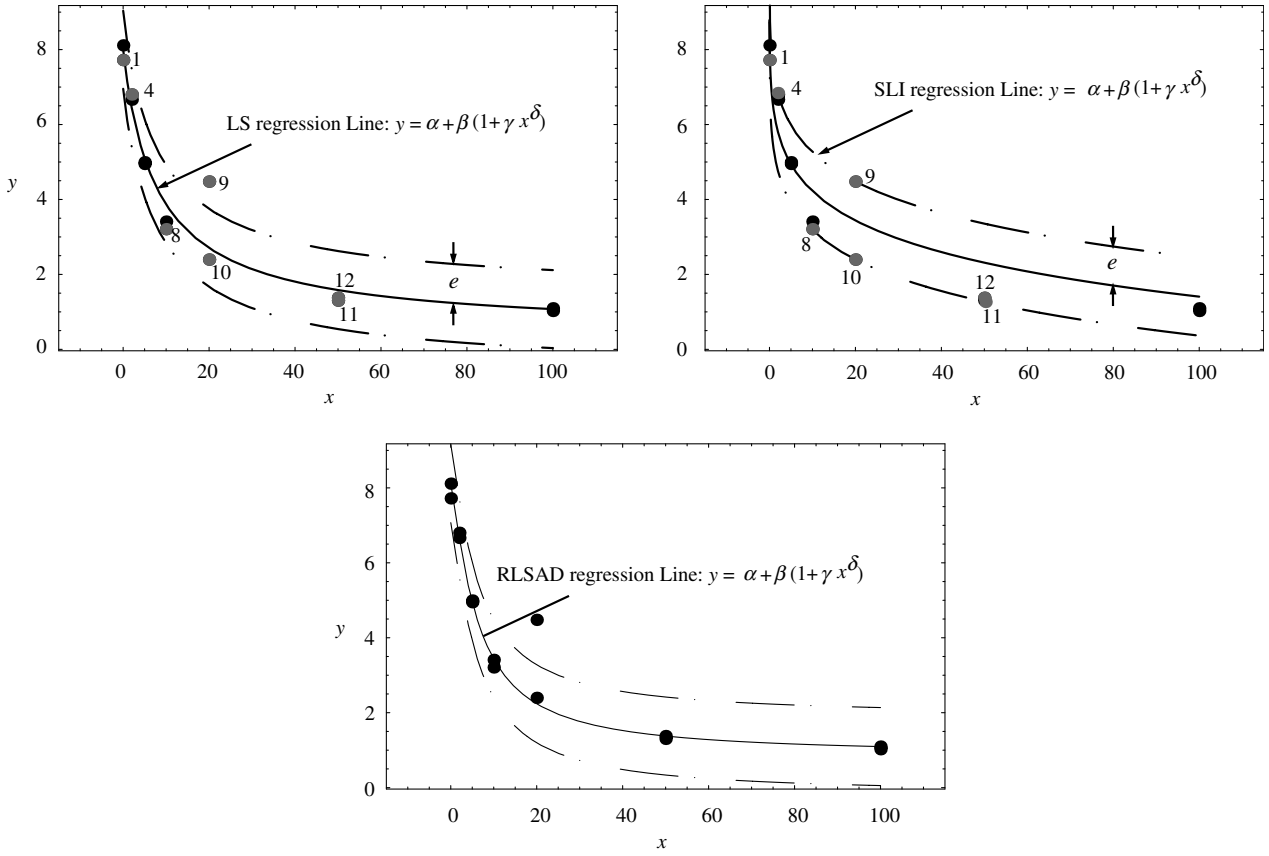

Fig. 5. Scatter plots of y_i versus x_i, for the Calibration data set in Table 3 and the ℓ_2 (upper left plot), the SLI fitted models (upper right plot) and the RLSAD fitted models (lower plot). Lower and upper bands (dashed lines) at a distance $\varepsilon = \pm 1.041$, i.e., the ℓ_∞ error, for the ℓ_2, SLI and RLSAD models are shown.

the previous iteration ones, the SLI reaches the solution at the third iteration, which proves that the ℓ_∞ estimators do not have a unique solution.

The following estimates of the calibration data are obtained:

$$\ell_2: \quad Z_{\ell_2} = 4.31, \quad \hat{\alpha} = 0.444, \quad \hat{\beta} = 7.551, \quad \hat{\gamma} = 0.133, \quad \hat{\delta} = 0.958,$$
$$\text{SLI}: \quad Z_{\text{SLI}} = 1.041, \quad \alpha = -12.545, \quad \hat{\beta} = 22.578, \quad \hat{\gamma} = 0.194, \quad \hat{\delta} = 0.251.$$

The scatter plot of y versus x for the calibration data set in Table 3 and the ℓ_2 (left plot) and SLI (right plot) fitted models are shown in Fig. 5. In addition to the regression lines, the lower and the upper bands (dashed lines) at a distance $\varepsilon = \pm 1.041$, i.e., the ℓ_∞ error, for the ℓ_2 and SLI estimates, are shown.

Note that data point 9 is outside the upper band for the least squares model and that the ℓ_2 gets a better fit with respect to the ℓ_∞ fit of the data points on the right, while the SLI gets a better fit of the data points on the left.

The RLSAD estimates are

$$Z_{\ell_1} = 3.36785, \quad Z_{\ell_2} = 5.2842, \quad \hat{\alpha} = 0.8731, \quad \hat{\beta} = 7.2425, \quad \hat{\gamma} = 0.1064, \quad \hat{\delta} = 1.2356,$$

which coincide with the ℓ_1 estimates. The RLSAD model is shown in Fig. 5 (lower plot).

6.3. The Weibull data

Castillo (2004) studies the dependence of the shape parameter β of a Weibull model when using data from the SUPERTANK project (see Kraus and McKee Smith, 1994) and fits the regression model

Table 4
Weibull shape parameter data

i	x_i	y_i	i	x_i	y_i	i	x_i	y_i	i	x_i	y_i
1	0.5440	−0.2218	21	0.9013	−0.3452	41	1.0879	−0.4686	61	1.2358	−0.5205
2	0.6702	−0.2539	22	0.9374	−0.4130	42	1.1322	−0.4678	62	1.2697	−0.5310
3	0.7776	−0.2907	23	0.9866	−0.4348	43	1.1579	−0.4617	63	1.3362	−0.4959
4	0.8570	−0.3144	24	1.0336	−0.4220	44	1.1997	−0.4530	64	1.5044	−0.5697
5	0.9026	−0.3490	25	1.0840	−0.4751	45	1.2355	−0.4791	65	0.5320	−0.1677
6	0.9340	−0.3989	26	1.1308	−0.4663	46	1.2710	−0.5104	66	0.5536	−0.1764
7	0.9874	−0.3941	27	1.1560	−0.5123	47	1.3382	−0.5091	67	0.7142	−0.3020
8	1.0327	−0.4191	28	1.1994	−0.4955	48	1.5044	−0.5722	68	0.8543	−0.3245
9	1.0854	−0.4594	29	1.2369	−0.4732	49	0.5849	−0.1769	69	0.9073	−0.3518
10	1.1303	−0.4662	30	1.2694	−0.5294	50	0.5312	−0.2573	70	0.9456	−0.3839
11	1.1573	−0.4890	31	1.3393	−0.4970	51	0.7399	−0.3065	71	0.9869	−0.3915
12	1.1993	−0.5158	32	1.5044	−0.5246	52	0.8558	−0.3197	72	1.0428	−0.4258
13	1.2356	−0.4852	33	0.5736	−0.1663	53	0.9020	−0.3499	73	1.0955	−0.4726
14	1.2698	−0.4855	34	0.6064	−0.3004	54	0.9342	−0.3960	74	1.1336	−0.4585
15	1.3388	−0.5281	35	0.7672	−0.3165	55	0.9827	−0.3964	75	1.1663	−0.4520
16	1.5044	−0.5377	36	0.8568	−0.3304	56	1.0336	−0.4293	76	1.2043	−0.4620
17	0.4974	−0.1901	37	0.9054	−0.3477	57	1.0889	−0.4530	77	1.2383	−0.5140
18	0.6451	−0.2674	38	0.9261	−0.3854	58	1.1295	−0.4468	78	1.2717	−0.4937
19	0.7722	−0.3128	39	0.9817	−0.3959	59	1.1602	−0.4578	79	1.3385	−0.4963
20	0.8574	−0.3234	40	1.0338	−0.4199	60	1.1993	−0.4736	80	1.5044	−0.5809

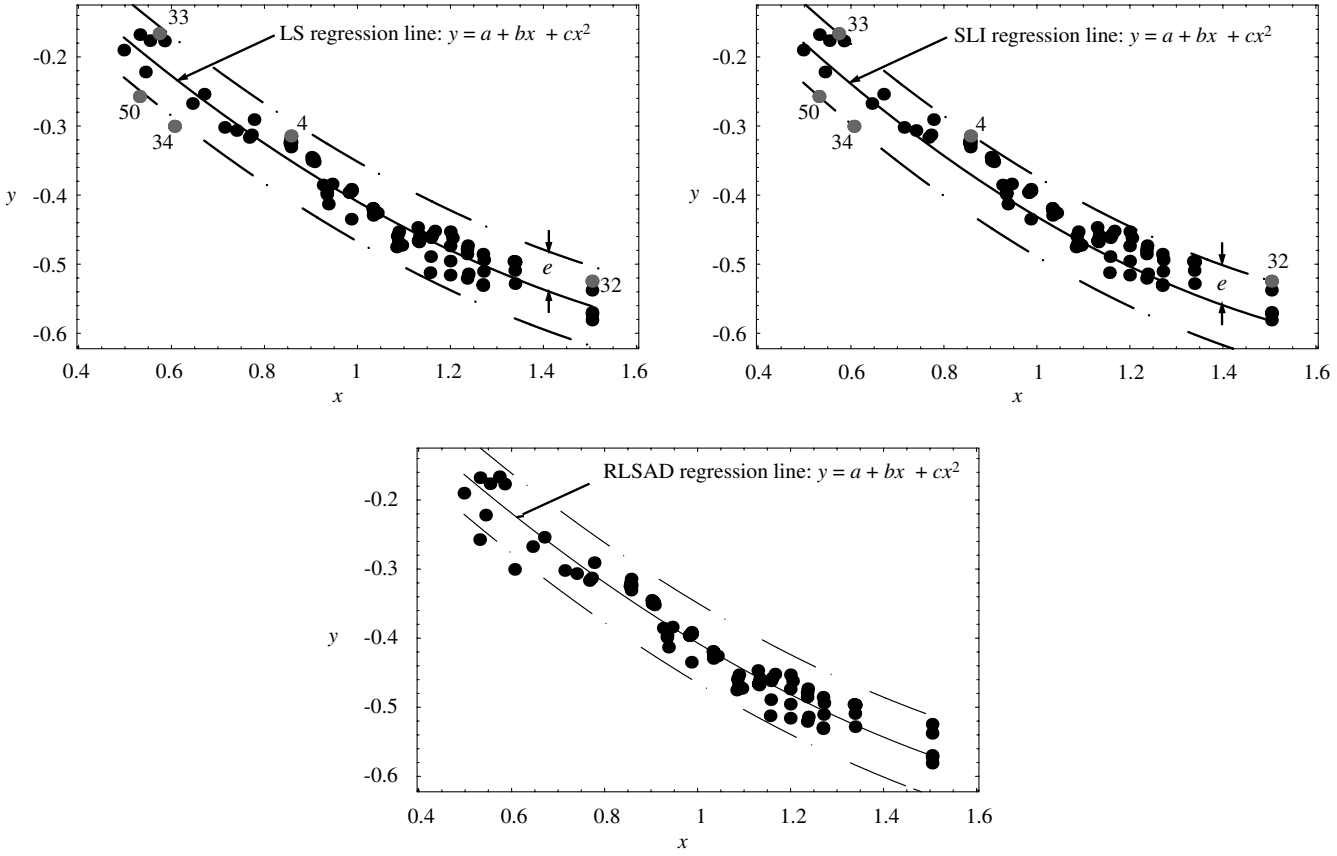

Fig. 6. Scatter plots of y_i versus x_i, for the Weibull data set in Table 4 and the ℓ_2 (upper left plot), SLI fitted models (upper right plot) and RLSAD fitted models (lower plot). Lower and upper bands (dashed lines) at a distance $\varepsilon = \pm 0.0578$, i.e., the ℓ_∞ error, for the ℓ_2, SLI and RLSAD models are shown.

$$\beta_i = a + bx_i + cx_i^2 + \varepsilon_i.$$

In this case, the SLI method stops after the second iteration, because the ℓ_∞ regression model for these data has a unique solution and the following parameter estimates (variances in parenthesis) are obtained:

$$\ell_2: \quad Z_{\ell_2} = 0.044, \quad \hat{a} = 0.149, \quad \hat{b} = -0.730, \quad \hat{c} = 0.171.$$
$$\text{SLI}: \quad Z_{\text{SLI}} = 0.0578, \quad \hat{a} = 0.170(0.0066), \quad \hat{b} = -0.801(0.0398), \quad \hat{c} = 0.200(0.0115).$$

The variances have been calculated by the bootstrap method by resampling and using 1000 simulations.

The scatter plot of y versus x for the Weibull data set in Table 4 and the ℓ_2 (left plot) and SLI (right plot) fitted models are shown in Fig. 6. In addition to the regression lines, the lower and the upper bands (dashed lines) at a distance $\varepsilon = \pm 0.0578$, i.e., the ℓ_∞ error, for the ℓ_2 and SLI estimates are shown.

Note that data points 34 and 50 are outside the lower band for the least squares model and that the ℓ_2 gets a better fit with respect to the ℓ_∞ fit of the data points on the right, while the SLI gets a better fit of the data points on the left.

The RLSAD estimates are

$$Z_{\ell_1} = 1.515, \quad Z_{\ell_2} = 0.046, \quad \hat{a} = 0.159, \quad \hat{b} = -0.727, \quad \hat{c} = 0.161,$$

which coincide with the ℓ_1 estimates.

7. Small sample and asymptotic performance of the SLI method

It is clear that, being the SLI and the RLSAD methods defined as the solution of a sequence of linear or non-linear programming problems (see Bazaraa et al. (1993)), their statistical properties for small and large samples seem practically intractable in an analytical way at a first look. Fortunately, some results are possible, both theoretically and using simulations.

7.1. Simulation experiments and results

Initially, the statistical properties of the estimates are tested using simulations. It is shown that their performances are good for the two regression models.

7.1.1. Linear regression model

First, the following regression model is considered

$$y_i = a + bx_i + \varepsilon_i; \quad i = 1, 2, \ldots, n.$$

Next, the SLI and RLSAD methods performance for three sampling designs are analyzed. The case of $a = 0$, $b = 1$ and $\epsilon_i \sim N(0, 0.1)$, and the case of Random replicated data with Random and equally spaced X_i; $i = 1, 2, \ldots, k$ with $k = 5$ and $k = 10$ levels are considered. Fig. 7 illustrates one example for $n = 50$ and 5 X-levels, and shows the true model.

Table 5 shows the results of 10 000 simulations with the averages over simulation replicate values and mean squared errors (MSE) of the parameter estimates for different sample sizes and sampling designs for this model. The expected numbers of iterations required for the SLI method and the percentage of simulations for which the ℓ_∞ has a unique solution appear in the last two columns. Note that one iteration means that the SLI method coincides with the standard ℓ_∞ method. They show that increasing the number of levels decreases the expected number of iterations and increases the uniqueness frequency. However, unexpectedly, the sample size has a reverse effect. This confirms that the uniqueness problem is not due to small sample sizes. Note that the Random selection of the X values leads to practically almost sure uniqueness. This was explained in Section 5 based on Proposition 4.

Table 6 is the corresponding table for the RLSAD method, where the last column shows the percentage of samples for which the ℓ_1 estimates are unique. Note that in this case the Random allocation of the X values does not lead to sure uniqueness and that this becomes worse with increasing sample size. However, the fre-

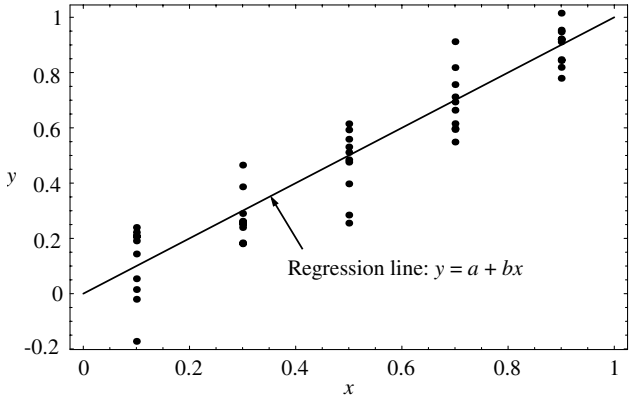

Fig. 7. Illustration of a typical simulated sample of size $n = 50$ showing the $k = 5$ X equally spaced levels and the true model: $y_i = x_i + \varepsilon_i$.

Table 5
Results of 10000 simulations showing the averages over simulation replicate values and mean squared errors (MSE) of the parameter estimates for different sample sizes and sampling designs for the model $y_i = a + bx_i + \varepsilon_i$ with $a = 0; b = 1$ and $\varepsilon \sim N(0, 0.1)$ together with the expected number of iterations required for the SLI method and the percentage of simulations for which the ℓ_∞ has a unique solution

Levels	n	$\bar{E}[\hat{a}]$	MSE$[\hat{a}]$	$\bar{E}[\hat{b}]$	MSE$[\hat{b}]$	E[iter]	%
5	20	0.00058	0.00483	0.99944	0.01391	2.429	63
5	40	0.00053	0.00374	0.99976	0.01062	2.488	60
5	100	0.00019	0.00290	2.00034	0.00834	2.530	58
5	200	0.00035	0.00244	0.99979	0.00682	2.529	59
5	500	0.00117	0.00199	0.99853	0.00557	2.539	58
10	20	0.00014	0.00526	1.00081	0.01516	2.127	87
10	40	0.00047	0.00409	1.00003	0.01172	2.188	82
10	100	0.00041	0.00317	0.99982	0.00881	2.220	80
10	200	0.00092	0.00264	0.99869	0.00733	2.228	79
10	500	0.00108	0.00218	0.99862	0.00611	2.238	79
Random	20	0.00082	0.00514	0.99857	0.02547	2.000	100
Random	40	0.00091	0.00408	0.99866	0.01429	2.000	100
Random	100	0.00021	0.00331	1.00018	0.01063	2.000	100
Random	200	0.00002	0.00270	1.00023	0.00818	2.000	100
Random	500	0.00047	0.00230	1.00012	0.00642	2.000	100

quency of non-uniqueness is much less than the one for grouped designs. This was explained in Section 5, based on Proposition 4.

All these tables show that uniqueness problems for the ℓ_1 and ℓ_∞ are not rare but a frequent event, that does not disappear with increasing the sample size.

The fact that Random selection of the predicting variable X leads to much more cases in which the estimate is unique has a theoretical explanation as it will shown later.

7.1.2. Quadratic regression model

Next, we consider the regression model

$$y_i = a + bx_i + cx_i^2 + \varepsilon_i; \quad i = 1, 2, \ldots, n, \tag{39}$$

and compare how the SLI method performs for various sampling designs. We consider the case of $a = 0$, $b = 1$, $c = -2$ and $\epsilon_i \sim N(0, 0.1)$, and the case of random replicated data equally spaced X_i; $i = 1, 2, \ldots, n$ with $k = 5$ and 10 levels. Fig. 8 illustrates one example for $n = 50$ and 5 X-levels, and shows the true model.

Table 7 shows the results of 10000 simulations with the expected values and mean squared errors (MSE) of the parameter estimates for different sample sizes and sampling designs for this model. The expected numbers

Table 6
Results of 10000 simulations showing the averages over simulation replicate values and mean squared errors (MSE) of the parameter estimates for different sample sizes and sampling designs for the model $y_i = a + bx_i + \varepsilon_i$ with $a = 0; b = 1$ and $\varepsilon \sim N(0, 0.1)$ together with the percentage of samples for which the ℓ_1 estimates are unique

Levels	n	$\bar{E}[\hat{a}]$	MSE$[\hat{a}]$	$\bar{E}[\hat{b}]$	MSE$[\hat{b}]$	%
5	20	0.00323	0.00329	0.99956	0.00990	66
5	40	0.00226	0.00159	1.00005	0.00482	67
5	100	0.00049	0.00064	1.00084	0.00194	69
5	200	0.00035	0.00031	1.00028	0.00096	69
5	500	0.00017	0.00013	0.99999	0.00038	70
10	20	0.00171	0.00318	0.99969	0.00951	83
10	40	0.00122	0.00161	1.00006	0.00479	84
10	100	0.00025	0.00063	1.00052	0.00189	84
10	200	0.00031	0.00031	1.00007	0.00094	84
10	500	0.00002	0.00013	1.00007	0.00038	83
Random	20	0.00088	0.00306	0.99831	0.00957	99
Random	40	0.00025	0.00164	1.00047	0.00562	99
Random	100	0.00021	0.00056	0.99998	0.00214	98
Random	200	0.00013	0.00027	0.99999	0.00104	96
Random	500	−0.00001	0.00011	1.00005	0.00038	93

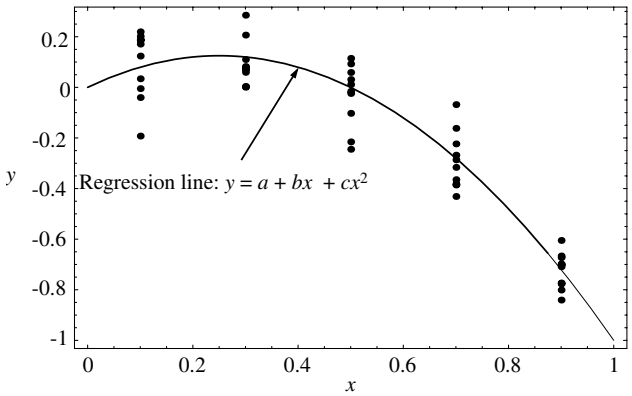

Fig. 8. Illustration of the simulation examples for $n = 50$ showing the five X-levels and the true model: $y_i = x_i - 2x_i^2 + \varepsilon_i$.

of iterations required for the SLI method and the percentage of simulations for which the ℓ_∞ has a unique solution appear in the last two columns. Similar conclusions as those for Table 5 can be obtained.

Table 8 is the corresponding table for the RLSAD method, where the last column shows the percentage of samples for which the ℓ_1 estimates are unique.

The analysis of both tables shows that:

1. The expected values of the estimates are very close to the true values, suggesting an unbiased or a very low bias estimates.
2. The mean squared errors (MSE) of the estimates decrease with sample size, suggesting that they are consistent.
3. The number of X-levels has a small influence on the quality of the parameter estimates, but a larger effect on the number of iterations required for the SLI method.

7.1.3. All possible solutions for the ℓ_∞ method

To illustrate the case of several solutions, we have selected one of the simulated samples, which is shown in Table 9.

For this data set, the set of all possible solutions of the problem (9)–(11) for the quadratic model (39) can be obtained using the techniques described in Section 2, i.e., solving the system of inequalities (10) and (11) and

Table 7
Results of 10000 simulations showing the averages over simulation replicate values and mean squared errors (MSE) of the parameter estimates for different sample sizes and sampling designs for the model $y_i = a + bx_i + cx_i^2 + \varepsilon_i$ with $a = 0;\ b = 1;\ c = -2$ and $\varepsilon \sim N(0, 0.1)$ together with the expected number of iterations required for the SLI method and the percentage of simulations for which the ℓ_∞ has a unique solution

Levels	n	$\bar{E}[\hat{a}]$	MSE$[\hat{a}]$	$\bar{E}[\hat{b}]$	MSE$[\hat{b}]$	$\bar{E}[\hat{c}]$	MSE$[\hat{c}]$	E[iter]	%
5	20	−0.00079	0.01109	1.00253	0.22684	−2.00154	0.20989	3.050	38.82
5	40	0.00006	0.00826	0.99953	0.17090	−1.99864	0.15861	3.265	34.58
5	100	0.00024	0.00612	0.99954	0.12865	−1.99942	0.11972	3.382	32.90
5	200	−0.00042	0.00492	0.99917	0.10480	−1.99837	0.09806	3.412	32.00
5	500	−0.00009	0.00407	1.00150	0.08697	−2.00120	0.08068	3.459	30.60
10	20	−0.00129	0.01150	1.00663	0.23162	−2.00534	0.21165	2.245	78.03
10	40	−0.00064	0.00884	1.00449	0.17371	−2.00323	0.15660	2.377	69.96
10	100	0.00058	0.00663	0.99766	0.13644	−1.99703	0.12493	2.449	66.76
10	200	0.00042	0.00542	1.00013	0.11184	−2.00096	0.10213	2.481	64.57
10	500	0.00130	0.00436	0.99734	0.08873	−1.99915	0.08047	2.492	64.49
Random	20	−0.00067	0.00694	1.00100	0.14948	−1.99924	0.16214	2.000	100.00
Random	40	0.00093	0.01005	0.99790	0.22872	−1.99775	0.22139	2.000	100.00
Random	100	0.00028	0.00656	0.99966	0.15245	−1.99981	0.14014	2.000	100.00
Random	200	−0.00049	0.00546	1.00394	0.11322	−2.00409	0.10432	2.000	100.00
Random	500	0.00058	0.00425	1.00201	0.09173	−2.00355	0.08798	2.000	100.00

Table 8
Results of 10000 simulations showing the expected values and mean squared errors (MSE) of the parameter estimates for different sample sizes and sampling designs for the model $y_i = a + bx_i + cx_i^2 + \varepsilon_i$ with $a = 0; b = 1; c = -2$ and $\varepsilon \sim N(0, 0.1)$ together with the percentage of samples for which the ℓ_1 estimates are unique

X-levels	n	$E[\hat{a}]$	MSE$[\hat{a}]$	$E[\hat{b}]$	MSE$[\hat{b}]$	$E[\hat{c}]$	MSE$[\hat{c}]$	LSAD (%)
5	20	0.00215	0.00709	0.99100	0.15558	−1.99105	0.14663	46
5	40	0.00582	0.00410	0.97735	0.09005	−1.97692	0.08516	54
5	100	0.00224	0.00163	0.98989	0.03621	−1.98908	0.03453	59
5	200	0.00131	0.00081	0.99448	0.01802	−1.99429	0.01705	60
5	500	0.00044	0.00034	0.99802	0.00756	−1.99791	0.00718	62
10	20	−0.00016	0.00685	1.00181	0.14759	−2.00250	0.13857	81
10	40	0.00031	0.00368	0.99999	0.07766	−1.99962	0.07232	87
10	100	−0.00063	0.00147	1.00335	0.03125	−2.00265	0.02919	88
10	200	−0.00009	0.00074	1.00103	0.01604	−2.00087	0.01504	87
10	500	−0.00001	0.00029	0.99989	0.00639	−1.99983	0.00606	86
Random	20	−0.00033	0.00879	1.00348	0.17717	−2.00440	0.14784	99
Random	40	0.00201	0.00501	0.99192	0.10494	−1.99218	0.09450	99
Random	100	−0.00020	0.00134	1.00290	0.03253	−2.00318	0.03287	97
Random	200	−0.00018	0.00058	1.00181	0.01442	−2.00191	0.01486	95
Random	500	−0.00020	0.00025	1.00127	0.00565	−2.00126	0.00554	90

replacing this solution into (9), removing its cone part and identifying its polytope components (those with the minimum ε component). Then, one obtains the general solution as

$$\begin{pmatrix} a \\ b \\ c \\ \varepsilon \end{pmatrix} = \begin{pmatrix} 0.03331 & 0.05563 & 0.03725 & 0.06700 \\ 0.97333 & 0.91667 & 0.92833 & 0.78667 \\ -2.01458 & -1.97917 & -1.95833 & -1.81667 \\ 0.19050 & 0.19050 & 0.19050 & 0.19050 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{pmatrix}; \quad \sum_{i=1}^{4} \lambda_i = 1;\ \lambda_i \geqslant 0\ \forall i. \tag{40}$$

Note that all the generators of the cone part of the polyhedral solution of the system (10) and (11) have disappeared, as it occurred in the illustrative example in Section 1, and the polytope components with associated values larger than 0.19050 have also disappeared.

Fig. 9 shows the selected data together with the four generating solutions in (40). In this case, due to the almost coincidence of the four solutions, the uniqueness problem is not important from a practical point of view.

Table 9
One example of the simulated data

i	x_i	y_i	i	x_i	y_i	i	x_i	y_i	i	x_i	y_i	i	x_i	y_i
1	0.100	0.049	11	0.100	−0.011	21	0.100	−0.035	31	0.100	0.301	41	0.100	0.180
2	0.300	0.153	12	0.300	0.154	22	0.300	0.191	32	0.300	0.058	42	0.300	0.045
3	0.500	0.046	13	0.500	−0.066	23	0.500	0.102	33	0.500	−0.022	43	0.500	0.076
4	0.700	−0.463	14	0.700	−0.329	24	0.700	−0.352	34	0.700	−0.376	44	0.700	−0.160
5	0.900	−0.793	15	0.900	−0.532	25	0.900	−0.820	35	0.900	−0.731	45	0.900	−0.636
6	0.100	−0.017	16	0.100	0.008	26	0.100	0.081	36	0.100	0.054	46	0.100	−0.063
7	0.300	0.081	17	0.300	−0.010	27	0.300	0.021	37	0.300	0.067	47	0.300	0.330
8	0.500	0.093	18	0.500	−0.138	28	0.500	−0.091	38	0.500	−0.104	48	0.500	0.067
9	0.700	−0.356	19	0.700	−0.248	29	0.700	−0.214	39	0.700	−0.082	49	0.700	−0.263
10	0.900	−0.720	20	0.900	−0.601	30	0.900	−0.563	40	0.900	−0.637	50	0.900	−0.555

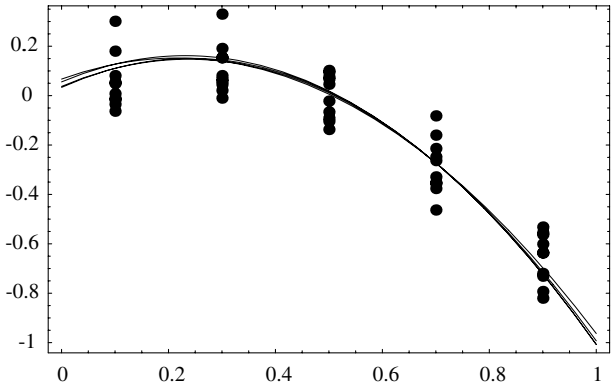

Fig. 9. Scatter plot of y_i versus x_i, for the illustrative example and the resulting four solutions generating the polytope for the ℓ_∞.

7.2. Some asymptotic results for the RLSAD

Many multiple linear regression estimators $\hat{\boldsymbol{\beta}}$ satisfy

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(0, V(\hat{\boldsymbol{\beta}}, F)\mathbf{W}), \tag{41}$$

where

$$\frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{n} \to \mathbf{W}^{-1}$$

and when the residuals ε_i are iid with a cdf F and a unimodal pdf f that is symmetric with a unique maximum at 0. If the variance $V(\varepsilon_i)$ exists (see Koenker and Bassett, 1978; Bassett and Koenker, 1978),

$$V(\ell_2, F) = V(\varepsilon_i) = \sigma^2, \tag{42}$$

$$V(\ell_1, F) = \frac{1}{4[f(0)]^2}. \tag{43}$$

Broffitt (1974) compares ℓ_2, ℓ_1, and ℓ_∞ in the location model and shows that the rate of convergence of the Chebyshev estimator is often very poor.

The ℓ_1 asymptotic theory holds whether the solution is unique or not, and this result could suggest that the ℓ_1 solutions are asymptotically equivalent and that the asymptotic theory for the RLSAD estimator should be the same as that for the ℓ_1 estimator. Though later some results in favor of this statement for some designs are obtained, the simulation results indicate that this is not always the case. Note that the percentage of uniqueness of solution decreases with sample size for some experimental designs.

If the RLSAD and ℓ_1 estimators were asymptotically equivalent, then the RLSAD estimator will not be used for large samples since the ℓ_1 estimator is much faster to compute (see for example, Portnoy and Koenker, 1997). However, the simulations in Section 6 indicate that this is not always the case.

The ℓ_∞ estimator behaves very erratically even for the location model (see Broffitt, 1974). The authors are not aware of asymptotic theory (of the form equation (41)) for the ℓ_∞ estimator in the regression setting. If there is such theory, then again the ℓ_∞ estimator and the SLI estimator are not asymptotically equivalent as the simulations in Section 6 indicate because the percentage of uniqueness of solution decreases with the sample size.

As a practical check of formulas (41)–(43) for the linear regression model the results in Table 6 for $k = 5$ with those resulting from (41)–(43) are compared:

$$\left(\frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{n}\right)^{-1} = \begin{pmatrix} 4.125 & -6.25 \\ -6.25 & 12.5 \end{pmatrix}$$
$$V(\ell_1, F) = 0.015708,$$

and then

$$\mathrm{Var}[(\hat{a}, \hat{b})] = \frac{1}{n}\begin{pmatrix} 0.06480 & -0.09817 \\ -0.09817 & 0.19635 \end{pmatrix},$$

that leads to

$$\mathrm{Var}(\hat{a}) \equiv 0.0648/n; \quad \mathrm{Var}(\hat{b}) \equiv 0.196375/n.$$

Table 10 shows that these approximations are nearly identical to the MSEs for RLSAD reported in Table 6.

Similarly, for the quadratic regression model one has

$$\left(\frac{\mathbf{X}^{\mathrm{T}}\mathbf{X}}{n}\right)^{-1} = \begin{pmatrix} 10.576 & -44.196 & 37.946 \\ -44.196 & 235.714 & -223.214 \\ 37.946 & -223.214 & 223.214 \end{pmatrix},$$
$$V(\ell_1, F) = 0.015708,$$

and then

$$\mathrm{Var}[(\hat{a}, \hat{b}, \hat{c})] = \frac{1}{n}\begin{pmatrix} 0.166 & -0.694 & 0.596 \\ -0.694 & 3.703 & -3.506 \\ 0.596 & -3.506 & 3.506 \end{pmatrix},$$

that leads to

$$\mathrm{Var}(\hat{a}) \equiv 0.166/n; \quad \mathrm{Var}(\hat{b}) \equiv 3.703/n; \quad \mathrm{Var}(\hat{c}) \equiv 3.506/n.$$

Table 11 shows that these approximations are nearly identical to the MSEs for RLSAD reported in Table 8.

Table 10
Linear regression case

n	$0.0648/n$	$0.196375/n$
20	0.00324	0.0098
40	0.00162	0.0049
100	0.00065	0.00196
200	0.00032	0.00098
500	0.00013	0.00039

Asymptotic variances of the parameter estimates for the RLSAD.

Table 11
Quadratic regression case

n	$0.166/n$	$3.703/n$	$3.506/n$
20	0.008306	0.185130	0.175312
40	0.004153	0.092565	0.087656
100	0.001661	0.037026	0.035062
200	0.000831	0.018513	0.017531
500	0.000332	0.007405	0.007012

Asymptotic variances of the parameter estimates for the RLSAD.

8. Conclusions

The main conclusions of this paper are

1. The set of all solutions of the ℓ_1 and ℓ_∞ linear regression have a polytope structure, i.e., a linear convex combination of some extreme solutions.
2. This set can be found by obtaining the general polyhedral solution of the system of linear inequalities, removing its cone parts, and identifying its polytope components as those with minimum ε for the ℓ_∞ and $\sum_{j=1}^{n}\varepsilon_j$ for the ℓ_1.
3. The previous method has been shown to be feasible for sample sizes of moderate size, but can be clearly infeasible for very large sample sizes. Fortunately, the critical sample size value can be large because each constraint of the ℓ_1 and ℓ_∞ in (10), (11) and (6), (7) contains only $p+1$ non-zero terms each.
4. The SLI and the RLSAD regression models, which give the same optimal values for the optimization function as the standard ℓ_∞ regression and ℓ_1 methods, respectively, and lead to a unique solution that has other important properties, have been introduced as alternatives for the ℓ_1 and ℓ_∞ methods.
5. The propositions in Giloni and Padberg are crucial to check the uniqueness of the ℓ_1 and ℓ_∞ methods and to guarantee if and when the RSLAD method leads to uniqueness.
6. Application of these methods to several examples in the existing literature and the performed simulations have shown to have a good behavior for small and moderately large sample sizes.
7. The proposed methods allow one to determine when the ℓ_1 and ℓ_∞ regression estimators have no unique solution for a given set of data, information that is not given by standard existing regression methods.
8. Simulation experiments indicate that increasing the number of levels leads to a less frequent uniqueness problem, which can completely disappear when the number of levels increases or the sample is chosen Randomly for X.
9. Increasing the sample size decreases the uniqueness frequency in the ℓ_∞ method, but increases that frequency for the ℓ_1, with the exception of the Random level samples.
10. The uniqueness problems of the ℓ_1 and ℓ_∞ methods are more related to the experimental design than to the sample size.
11. The proposed methods apart from leading to uniqueness of the modified ℓ_1 and ℓ_∞ methods, in addition to the minimization of the ℓ_1 and ℓ_∞ errors, enforce extra desirable conditions and lead to efficient solution methods, which make them competitive with alternative uniqueness proposals.

Acknowledgements

The authors are indebted to the Spanish Ministry of Science and Technology (Project BIA2005-07802-C02-01) for partial support of this work. The authors also thank the editor and the referees, whose suggestions and references allowed a substantial improvement of the paper.

References

Bassett, G.W., Koenker, R.W., 1978. Asymptotic theory of least absolute error regression. Journal of the American Statistical Association 73, 618–622.

Bazaraa, M.S., Sherali, H.D., Shetty, C.M., 1993. Nonlinear Programming, second ed. John Wiley & Sons, New York.

Belsley, D.A., Kuh, E., Welsch, R.E., 1980. Regression Diagnostics: Identifying Influential Data and Sources of Multicollinearity. John Wiley & Sons, New York.
Bloomfield, P., Steiger, W., 1980. Least absolute deviations curve-fitting. SIAM Journal on Scientific and Statistical Computing 1 (2), 290–301.
Broffitt, J.D., 1974. An example of the large sample behavior of the midrange. The American Statistician 28, 69–70.
Castillo, M.C., 2004. Reliability and Optimization in Maritime Engineering. Monograph Series. Ph.D. Thesis, 2004. University of Granada, Spain (in Spanish).
Castillo, E., Jubete, F., 2004. The Γ-algorithm and some applications. International Journal of Mathematical Education in Science and Technology, 369–389.
Castillo, E., Cobo, A., Jubete, F., Pruneda, R.E., 1999. Orthogonal Sets and Polar Methods in Linear Algebra: Applications to Matrix Calculations, Systems of Equations and Inequalities, and Linear Programming. John Wiley and Sons, New York.
Castillo, E., Gutiérrez, J.M., Cobo, A., Castillo, C., 2000a. A minimax method for learning functional networks. Neural Processing Letters 11 (1), 39–49.
Castillo, E., Cobo, A., Jubete, F., Pruneda, R.E., Castill, C., 2000b. An orthogonally based pivoting transformation of matrices and some applications. SIAM Journal on Matrix Analysis and Applications 22 (3), 666–681.
Castillo, E., Conejo, A., Pedregal, P., García, R., Alguacil, N., 2001. Building and Solving Mathematical Programming Models in Engineering and Science. John Wiley & Sons, New York.
Castillo, E., Jubete, F., Pruneda, R.E., Solares, C., 2002. Obtaining simultaneous solutions of linear subsystems of equations and inequalities. Linear Algebra and its Applications 346, 131–154.
Castillo, E., Hadi, A.S., Conejo, A., Fernández-Canteli, A., 2004. A general method for local sensitivity analysis with application to regression models and other optimization problems. Technometrics 46 (4), 430–444.
Chatterjee, S., Hadi, A.S., 1988. Sensitivity Analysis in Linear Regression. John Wiley & Sons, New York.
Chatterjee, S., Hadi, A.S., Price, B., 2000. Regression Analysis by Example, third ed. John Wiley & Sons, New York.
Dodge, Y., 1987. Statistical Data Analysis Based on the L1-Norm and Related Methods. Elsevier Science Publishers (North-Holland), Amsterdam, The Netherlands.
Dodge, Y., 1992. L_1-Statistical Analysis and Related Methods. North-Holland Publishing, Amsterdam, The Netherlands.
Dodge, Y., 1997. L_1-Statistical Procedures and Related Topics. Institute of Mathematical Statistics, Hayward, Calif, USA.
Dodge, Y., Falconer, W., 2002. Statistical Data Analysis Based on the L_1-Norm and Related Methods. Barika Photography & Productions, New Bedford, Mass, USA.
Edgeworth, F.Y., 1887. On observations relating to several quantities. Hermathena 6, 279–285.
Edgeworth, F.Y., 1888. On a new method of reducing observations relating to several quantities. Philisophical Magazine 25, 184–191.
Eisenhart, C., 1961. Boscovitch and the combination of observations. In: Boscovitch, Roger Joseph, Whyte, L.L. (Eds.). Fordham University Press, New York.
Euler, L., 1749. Inégalités du Movement the Saturn te de Jupiter. Martin, Coignard et Guerin, Paris.
Giloni, A., Padberg, M., 2002. Alternative methods of linear regression. Mathematical and Computer Modelling 35, 361–374.
Hawley, R.W., Gallagher Jr., N.C., 1994. On Edgeworth's method for minimum absolute error linear regression. Transaction on Signal Processing 42 (8), 2045–2054.
Koenker, R.W., Bassett, G., 1978. Regression quantiles. Econometrica 46, 33–50.
Kraus, N.C., McKee Smith, J., 1994, Supertank Laboratory Data Collection Project, vol. I: Main text. Tech. Report, CERC-94-3, US Army Eng. Wat. Exp., St. Vicks, Mss., 274 pp.
Laplace, P.S., 1789. Sur quelques points du système du monde. Mémoires de l'Académie des Sciences de Paris (reprinted in Oevres Complétes 11, 475–558, Gauthier-Villars, Paris).
Luenberger, D.G., 1989. Linear and Nonlinear Programming, second ed. Addison-Wesley, Reading, MA.
Meketon, M.S., 1986, Least absolute value regression. Technical Report, Bell Labs, Holmdel, N.J.
Mosteller, F., Rulon, P.J., Harris, T.E., 1950. Regression using minimum absolute deviations. The American Statistician 4 (1), 14–15.
Padberg, M., 1995. Linear Optimization and Extensions. Springer, Berlin.
Portnoy, S., Koenker, R., 1997. The Gaussian hare and the Laplacian tortoise: Computability of squared error versus absolute-error estimators. Statistical Science 12, 279–300.
Rao, C.R., Toutenburg, H., 1999. Linear Models: Least Squares and Alternatives. Springer.
Seber, G.A.F., Wild, C.J., 1989. Nonlinear Regression. John Wiley & Sons, New York.
Sielken, R.L., Hartley, H.O., 1973. Two linear programming algorithms for unbiased estimation of linear models. JASA 68, 639–641.
Stigler, S.M., 1984. Boscovich, Simpson and a 1760 manuscript note on fitting a linear relation. Biometrika 71, 615–620.
Stigler, S.M., 1986. The history of statistics: Measurements of uncertainty before 1900. Harvard University Press.
Tiede, J.J., Pagano, M., 1979. The application of robust calibration to radioimmunoassay. Biometrics 35, 567–574.
Wesolowsky, G.O., 1981. A new descent algorithm for the least absolute value regression problem. Communications in Statistics, Simulation and Computation B10 (5), 479–491.
Yinbo, L., Arce, G.R., 2004. A maximum likelihood approach to least absolute deviation regression. EURASIP Journal on Applied Signal Processing 12, 1762–1769.