

Peiliang Xu

## Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness

Received: 8 August 2004 / Accepted: 16 March 2005 / Published online: 2 June 2005  
© Springer-Verlag 2005

**Abstract** The findings of this paper are summarized as follows: (1) We propose a sign-constrained robust estimation method, which can tolerate 50% of data contamination and meanwhile achieve high, least-squares-comparable efficiency. Since the objective function is identical with least squares, the method may also be called sign-constrained robust least squares. An iterative version of the method has been implemented and shown to be capable of resisting against more than 50% of contamination. As a by-product, a robust estimate of scale parameter can also be obtained. Unlike the least median of squares method and repeated medians, which use a least possible number of data to derive the solution, the sign-constrained robust least squares method attempts to employ a maximum possible number of good data to derive the robust solution, and thus will not be affected by partial near multicollinearity among part of the data or if some of the data are clustered together; (2) although M-estimates have been reported to have a breakdown point of  $1/(t + 1)$ , we have shown that the weights of observations can readily deteriorate such results and bring the breakdown point of M-estimates of Huber's type to zero. The same zero breakdown point of the  $L_1$ -norm method is also derived, again due to the weights of observations; (3) by assuming a prior distribution for the signs of outliers, we have developed the concept of subjective breakdown point, which may be thought of as an extension of stochastic breakdown by Donoho and Huber but can be important in explaining real-life problems in Earth Sciences and image reconstruction; and finally, (4) We have shown that the least median of squares method can still break down with a single outlier, even if no highly concentrated good data nor highly concentrated outliers exist.

**Keywords** Adaptive trimmed mean · Breakdown point ·  $L_1$ -norm · Least median of squares · Outlier · Repeated

median · Robust estimation · Robustness · Sign-constrained robust least squares

### 1 Introduction

Motivated by the challenge of solving practical geodetic/astronomical problems of the eighteenth century, least squares (LS) was invented by Gauss in 1795 (and likely independently by Legendre in 1805) (see e.g. Gauss 1964, 1995; Stigler 1986). The method of least squares had since remained almost unchallenged until Huber published his landmark paper on robust estimation in 1964. Extracting correct or meaningful information on quantities of interest from contaminated data has been substantially investigated in almost every area of engineering and science since then, because least squares is vulnerable to even a single outlier. Identification and removal of contaminated data have been attempted and realized in two different ways, either (1) by first cleaning the data and then applying the classical least squares criterion to the remaining data (see e.g. Anscombe 1960; Baarda 1968; Pope 1976; Belsley et al. 1980; Chatterjee and Hadi 1986; Barnett and Lewis 1994); or (2) by designing robust estimation criteria and applying them directly to contaminated data (see e.g. Huber 1981; Hampel et al. 1986; Jurečková and Sen 1996; Koch 1999). Although the idea of cleaning data is very old, statistical procedures for detecting outliers practically work well only in the case of one single outlier but can often fail in the case of multiple outliers, in particular, if outliers are masked (see e.g. Chatterjee and Hadi 1986; Rocke and Woodruff 1996). If an initial robust estimate of the model parameters can be obtained, correct detection of multiple outliers is indeed possible by using a two-phase approach (see e.g. Xu 1989a; Rocke and Woodruff 1996).

The second line of development, namely robust estimation, has become one of the most intensive research topics in statistics since the publication of Huber (1964) pioneering paper, and has found widest possible applications in science and engineering. Many robust estimation methods have been investigated and proposed since then, of which the most

Peiliang Xu  
Disaster Prevention Research Institute,  
Kyoto University, Uji,  
Kyoto 611-0011, Japan  
E-mail: pxu@rcep.dpri.kyoto-u.ac.jp

important classes include the sample median and the trimmed mean (see e.g. Stigler 1977; Welsh 1987), least absolute deviation or  $L_1$  norm (see e.g. Bloomfield and Steiger 1983; Dodge 1997), robustified maximum likelihood or M-estimates (Huber 1964, 1981; Andrews 1974; Hampel et al. 1986; Jurečková and Sen 1996), order-based or L-estimates (Bickel 1973; Huber 1981; Hampel et al. 1986; Jurečková and Sen 1996), and rank-based or R-estimates (Hodges and Lehmann 1963; Adichie 1967; Jurečková 1971; Jaeckel 1972; Koul 1977). M-estimates have also been extended to correlated observations (see e.g. Gastwirth and Rubin 1975; Portnoy 1977; Boente and Fraiman 1989; Xu 1989b; Yang et al. 2002) and to heterogenous models (Beran 1982; Carroll and Rupert 1982). However, we may note that only the methods proposed by Xu (1989b) and Yang et al. (2002) are generally applicable, since a particular structure of correlation has been presumed in the case of Gastwirth and Rubin (1975), Portnoy (1977) and Boente and Fraiman (1989).

To measure robustness of a robust estimator, Hampel invented two most important concepts: influence curves/functions and breakdown points (see e.g. Hampel 1971, 1974; Hampel et al. 1986). They are designed to measure local and global robustness of a robust estimator, respectively. Since these measures are distribution-dependent, Donoho and Huber (1983) developed the concept of distribution-free breakdown points. Roughly speaking, the breakdown point of a robust estimator is the maximum percentage of contaminated data beyond which the estimator can no longer be able to produce (physically) meaningful solutions, since they can be arbitrarily far away from the true values of the model parameters. The best possible breakdown point that one can expect of a robust estimator has reported to be 0.5 (see e.g. Rousseeuw 1984; Hampel et al. 1986). Since a robust estimator with a high breakdown point has been statistically inefficient up to the present (Stefanski 1991; He 1994), a common practice to achieve both high breakdown point and efficiency is to first use a high breakdown robust estimator in order to remove the outliers and then apply classical techniques or M-estimation in order to achieve high efficiency (see e.g. Yohai 1987). Although the sample median,  $\alpha$ -trimmed mean and M-estimates can reach the maximum breakdown point in the one-dimensional case of location estimation (see e.g. Huber 1984; Donoho and Huber 1983; Hampel et al. 1986), the same is not quite true in the multi-dimensional case or linear models. The exception may be the repeated medians (Siegel 1982) in the well designed setting. Rousseeuw (1984) investigated the least median of squares (LMS) and reported that it can obtain the breakdown point of 0.5 (see also Rousseeuw and Leroy 1987). This may explain why the LMS method has attracted much attention in the last two decades, statistically and numerically. Hettmansperger and Sheather (1992) demonstrated that the LMS method is not necessarily of 0.5 breakdown point, which should have already been implied by Fig. 1 of Stefanski (1991) (see also Dodge 1992). Recently, Wang and Suter (2003) demonstrated that the LMS method cannot resist against clustered outliers, and its breakdown point can be below 0.4. These examples, however, have

a common feature that most or all of outliers, together with some good data, are either highly concentrated or follow a certain particular pattern. Since highly concentrated observations (good or bad) are relatively easy to identify, it might not be appropriate to use them to denounce the high breakdown property of the LMS method. In the case of the least trimmed squares, the breakdown point can even be lower than 0.3 (see also Wang and Suter 2003). In this paper, we will further show in Sect. 2 that a single influential outlier can completely nullify the LMS results, even if no obvious phenomenon of clustering or concentration can be seen.

The paper is organized as follows. First of all, we will use simulations to demonstrate that the LMS method could break down with a single outlier, even if neither good data nor outliers exhibit any phenomenon of clustering or concentration. Then in Sect. 3, by assuming a prior distribution for the signs of outliers, we will develop the concept of subjective breakdown points, which can take the values from zero to one, depending on the prior distribution. This extension in concept is important in explaining some real-life problems. Although the finite sample breakdown point of Donoho and Huber (1983) can also be as large as 1.0, ours is fundamentally different from theirs in two ways: (i) up to the present time, the maximum breakdown point of Donoho and Huber (1983) is only valid either in the case of constant estimates or Bayesian estimates with compact support. These two special cases are practically not meaningful. In the first case, the reason should be obvious, since the estimate has nothing to do with data. In the second case, the Bayesian estimate is always driven to the boundary of the compact support and thus does not make much sense practically. We will investigate the effect of weights of observations on robustness in Sect. 4. Although M-estimates have been reported to have a breakdown point of  $1/(t+1)$ , we will show that the weights of observations can readily invalidate such results and bring the breakdown point of M-estimates to zero. Here  $t$  is the number of unknowns to be estimated. In Sect. 5, we will propose the sign-constrained robust method, which can achieve both the maximum breakdown point of 0.5 and high efficiency. Since the objective function is essentially identical with least squares, the sign-constrained robust method may be also called *sign-constrained robust least squares*. In particular, the iterative version of the method is shown to be capable of resisting against more than 50% of data contamination, which can be of significant importance in practical applications in Earth Sciences and object reconstruction. A number of simulated examples will be given in Sect. 6 to demonstrate the performance of the new robust method.

In the rest of this paper, we assume the following linear model:

$$\mathbf{y} = \mathbf{A}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad D(\mathbf{y}) = \boldsymbol{\Sigma}_0\sigma^2, \quad (1)$$

where  $\mathbf{y}$  is an observation vector of dimension  $n$ ,  $\mathbf{A}$  is a design matrix of full rank,  $\boldsymbol{\beta}$  is the unknown vector of dimension  $t$ ,  $\boldsymbol{\epsilon}$  is the observational error vector,  $\boldsymbol{\Sigma}_0$  is a given positive definite matrix. The inverse of  $\boldsymbol{\Sigma}_0$  is also called weight matrix in adjustment theory and is almost always denoted by  $\mathbf{P}$ ,

namely,  $\mathbf{P} = \Sigma_0^{-1} \cdot \sigma^2$  is an unknown scale parameter or variance of unit weight. If the elements of the first column of  $\mathbf{A}$  are all equal to unity, (1) becomes the standard regression model. In this paper, we will focus only on outliers in the observations  $\mathbf{y}$  and in the design matrix  $\mathbf{A}$ . Outliers will be treated in terms of shifts of observations from their true values. In other words, if an observation, say  $y_i$ , contains an outlier, we will rewrite its corresponding functional model as  $y_i = \mathbf{a}_i\boldsymbol{\beta} + \epsilon_i + \delta y_i$ , where  $\delta y_i$  is the size of the outlier. If  $\delta y_i = 0$ , then the observation  $y_i$  is a good one. In order to identify outliers, one also has to further assume that for each model parameter, there must, at least, exist two good data that contain the information on such a parameter.

## 2 Zero breakdown examples of the LMS method

Previous results (Stefanski 1991; Dodge 1992; Hettmansperger and Sheather 1992; Wang and Suter 2003) suggested that the breakdown point of the LMS method could be lower than the claimed maximum value of 0.5, if outliers, together with some good data, are highly clustered and concentrated. As a consequence, we decide to further investigate quantitatively the breakdown point of the LMS method through simulations by removing the implicit condition of clustering and concentration of outliers from the previous works. To our surprise, the designed experiments have found a great number of examples for which the LMS method breaks down with a single influential outlier, although no obvious phenomenon of clustering and concentration can be seen any more. We will report our experiment design and one of such identified examples in this section.

The linear model is involved with two parameters, namely, 
$$y_i = 10\beta_1 + x_i\beta_2 + \epsilon_i, \quad i = 1, 2, \dots, n. \quad (2)$$

The true values of  $\beta_1$  and  $\beta_2$  are set to  $-2.0$  and  $0.7$  in all our experiments, respectively. For each experiment, we generate one hundred  $x_i$  by using a uniform distribution over the interval of  $[2.0, 10.7]$ , and further one hundred random errors  $\epsilon_i$  by using a normal distribution with mean zero. The standard deviation of  $\epsilon_i$  is equal to  $3.0$ . The true values of one hundred observations  $y_i$  can thus be computed from the true values of  $\beta_1$  and  $\beta_2$  and the generated data of  $x_i$ . Adding the generated random errors to these true values will make 100 observations  $y_i$ . We then simulate one outlier by setting  $x_{101}$  and its corresponding observation  $y_{101}$  to  $1500.0$  and a number around  $-15.6356$ , respectively. If  $y_{101}$  would not be an outlier, then it should be equal to  $1030.0$  plus a random error from the normal distribution.

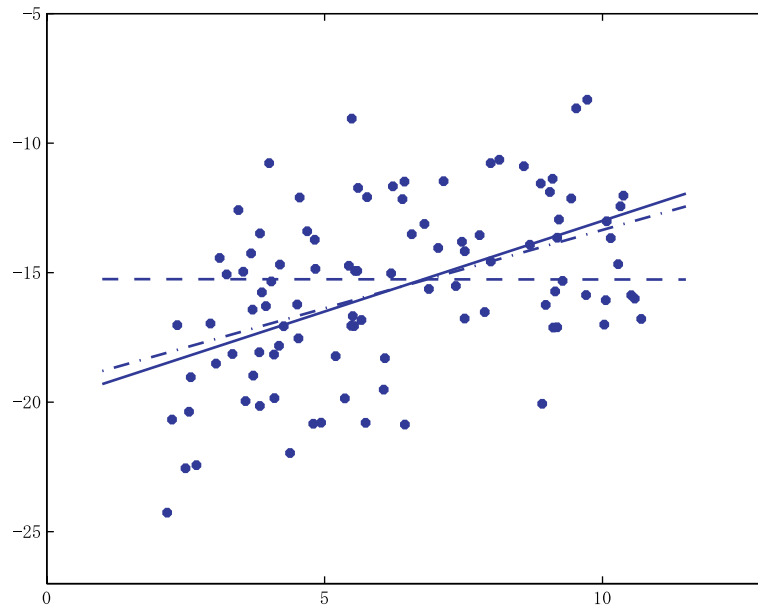
We have used the same experiment setting and conducted a great number of experiments, each of which is designed by generating a new set of  $x_i$  ( $i = 1, 2, \dots, 100$ ) and a new set of random errors  $\epsilon_i$  ( $i = 1, 2, \dots, 100$ ). While  $x_{101}$  remains unchanged, we only slightly modify  $y_{101}$ . The LMS solution in each experiment is obtained by implementing the exact LMS algorithm of Cheney (1966) and Stromberg (1993). In this simulation study, we have found many examples for which the LMS method have broken down.

Shown in Fig.1 are the 100 data points and the true line of regression. Since the outlier point is far away, it is not depicted here. The least squares estimates of the two parameters without the outlier are  $-1.940$  and  $0.605$ , with the standard deviations of  $0.081$  and  $0.121$ , respectively. The LMS estimates of  $\beta_1$  and  $\beta_2$  have been found to be involved with the outlier and are equal to  $-1.525$  and  $-0.001$ , respectively. Obviously, the difference between the LS and LMS estimates of the two parameters cannot reasonably be explained in terms of the standard deviations of the LS estimate. Since the only outlier has been used by the LMS method, this should clearly indicate that the LMS method has a zero breakdown point for this specific example. In fact, we have identified a great many of examples of this kind during the experiments, which are not reported here, however.

## 3 Subjective breakdown point

The concept of breakdown point was first mathematically formulated by Hampel (1971, 1974) as a most important global measure of robustness against outliers. The breakdown point of Hampel (1971) is only asymptotic and is not completely free of the distribution of data. It can be difficult to compute in some cases. As a result, Donoho and Huber (1983) further developed significantly the concept of breakdown point by extending it to the case of finite sample. Since the breakdown point of Donoho and Huber (1983) is valid for finite sample and does not depend on the specified distribution of data, it has been widely adopted in both theoretical and practical literatures of robust statistics.

The maximum breakdown point of a robust procedure has been known to be  $0.5$ , except for two cases of no practical importance as defined and given in Donoho and Huber (1983). This might be interpreted as that no robust procedure could produce meaningful results for a practical problem, if more than  $50\%$  of the data are contaminated by outliers. A widely accepted argument to support this position is that no robust methods are capable of discriminating the minority of good data from the majority of bad data. Should this indicate that problems of this kind are of no practical and/or physical meaning? If we would follow the reasonings as given in the above, our answer to this question is undoubtedly affirmative. In reality, we do have to deal with such kinds of problems. For example, we know that in the determination of stress tensors from earthquake focal mechanisms, one of the (two) nodal planes is the (correct) fault plane (good data) and the other is the auxiliary plane (bad data) (see e.g. Angelier 2002; Xu 2004). In addition, some of many earthquake focal mechanisms may not come from the effect of the same stress tensor, and should be further treated as *truly* erroneous data. In other words, we have to deal with at least  $50\%$  of data contamination in the stress inversion from earthquake focal mechanisms. In image processing, we have also seen a lot of noise other than signal. In this section, we will assume that we have some (rough) prior information about the nature of outliers or bad data. By incorporating the prior information into



**Fig. 1** The simulated data of the example, except for the only outlier which is too far away to be nicely plotted in this figure. Also shown in this figure are the true line of regression (*solid line*), the line by the LMS solution which has broken down due to the only outlier (*dashed line*), and the line by the sign-constrained robust estimator to be presented later in Sect. 5 (*dash-dotted line*). The horizontal and vertical axes show the values of  $x_i$  and  $y_i$ , respectively

robustness, we will naturally develop the concept of *subjective* breakdown point, which might be thought of as a kind of extension or realization of *stochastic breakdown* proposed by Donoho and Huber (1983). This new breakdown point is said to be subjective, since: (1) it is based on certain prior information on the nature of outliers; and (2) such prior information may only reflect the subjective belief of the data analyst on outliers. We will show that a subjective breakdown point can indeed take value beyond the current maximum of 0.5. Consequently, the concept of subjective breakdown point may be meaningfully used to interpret solutions to problems with more than 50% contamination physically.

For simplicity of discussion, assume in the linear model (1) that  $\mathbf{A} = \mathbf{e}$  and  $\Sigma_0 = \mathbf{I}$ . Here  $\mathbf{e}$  is a vector of dimension  $n$  with all its elements equal to unity. In other words, we assume  $n$  independently, identically distributed random samples  $y_1, y_2, \dots, y_n$ . Following the replacement approach of Donoho and Huber (1983), we replace part of these samples with outliers, say  $m$  outliers. Without loss of generality, we assume that the first  $m$  data are outliers, namely,  $(y_i + \delta y_i)(i \leq m)$ .

Although the magnitude of the shift  $\delta y_i$  can take on any large number, we assume that the sign of  $\delta y_i$  has a Bernoulli distribution, namely,

$$f(s_i) = p^{s_i} q^{1-s_i} I_{\{0,1\}}(s_i), \quad (3)$$

where  $I_{\{0,1\}}(s_i)$  is an indicator function, and  $s_i$  is either equal to zero or unity. In other words, we assume that the probabilities of  $\delta y_i$  being positive ( $s_i = 1$ ) and negative ( $s_i = 0$ ) are equal to  $p$  and  $q$ , respectively. For convenience but without loss of generality, we assume that the signs of the other  $(m - 1) \delta y_j (j \neq i)$  are independent and have the same distri-

butions as that of  $\delta y_i$ . Then the joint probability distribution for the signs of the  $m$  outliers has a binomial distribution:

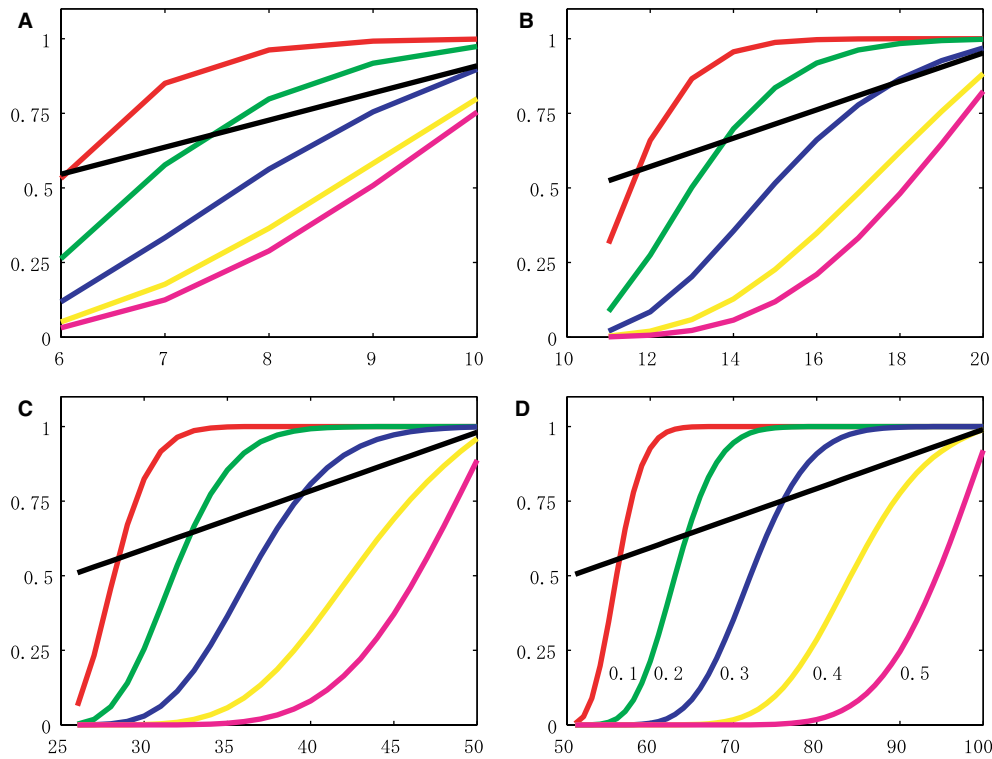
$$f(s) = f(s; m, p) = \binom{m}{s} p^s q^{m-s} I_{\{0,1,\dots,m\}}(s), \quad (4)$$

(see e.g. Mood et al. 1974), where  $f(s)$  is the probability of  $s$  positive  $\delta y_i$  and  $(m - s)$  negative  $\delta y_j (j \neq i)$ .

It is well known that if  $m \geq [n/2] + 1$ , robust procedures will break down, where  $[x]$  stands for the integer around but smaller than  $x$ . The question of interest now is: with what probability will a robust procedure break down? If a robust method breaks down almost surely or with a large probability, we can no longer trust and physically interpret the results from the set of contaminated data. On the contrary, if a robust method would break down only with a very small probability, we know that it hardly breaks down and will have confidence in the computed results from contaminated data either for interpretation or practical use.

Obviously, the arithmetic mean of the samples will *always* break down if  $m \geq 1$ , no matter whether we have the prior information (4) or not. It is nowhere robust. In the rest of this section, we will focus on the sample median and the  $\alpha$ -trimmed mean. As the first example, let us examine the sample median. It is well known that the median does not break down if  $m \leq [n/2]$  ( $n$  odd) or  $m < [n/2]$  ( $n$  even). Thus we will focus on  $[n/2] + 1 \leq m \leq n$ . Given  $n, m (\geq [n/2] + 1)$  and  $p$ , we know the subjective breakdown point of the median is equal to  $m/n (> 0.5)$  and we can compute the probability for this breakdown point as follows:

$$P(\text{breakdown}) = \sum_{i=0}^{m-[n/2]-1} \binom{m}{i} p^i q^{m-i}$$



**Fig. 2** The subjective breakdown points of the sample median and their corresponding probabilities of breakdown. The horizontal and vertical axes show the number of contaminated data and the probability of breakdown, respectively. The six curves in each subplot are the subjective breakdown points of the sample median (black line) and their corresponding probabilities of breakdown with different probabilities for positive sign of shifts  $\delta y$ ; (red line:  $p = 0.1$ ; green line:  $p = 0.2$ ; blue line:  $p = 0.3$ ; yellow line:  $p = 0.4$ ; and purple-red line:  $p = 0.5$ ). The four subplots **A**, **B**, **C** and **D** correspond to the sizes of samples 11, 21, 51 and 101, respectively

$$+ \sum_{i=[n/2]+1}^m \binom{m}{i} p^i q^{m-i}.$$

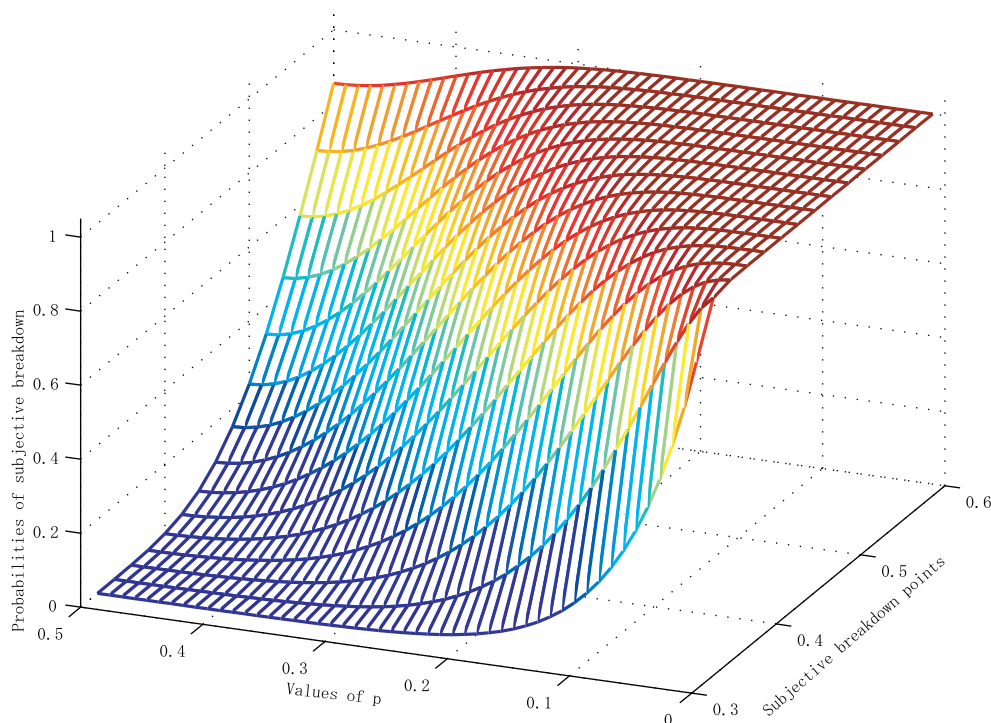
Alternatively, we can also compute the probability that the median does not break down as follows:

$$\begin{aligned} P(\text{not breakdown}) &= 1 - P(\text{breakdown}) \\ &= \sum_{i=m-[n/2]}^{[n/2]} \binom{m}{i} p^i q^{m-i}. \end{aligned} \quad (5a)$$

In particular, if  $p = 1$  (or  $q = 1$ ) and if  $m \geq [n/2] + 1$ , then  $P(\text{breakdown}) = 1$  or  $P(\text{not breakdown}) = 0$ . In this special case, we know the median always breaks down with probability one, and it does not make sense to talk about a subjective breakdown point higher than 0.5. In other words, if  $p = 1$  (or  $q = 1$ ), the median has the maximum breakdown point of 0.5. On the other hand, if  $m \leq [n/2]$ , then we always have  $P(\text{breakdown}) = 0$  or  $P(\text{not breakdown}) = 1$ , which confirms common sense that the sample median will never break down if the contamination of data is less than 50%. If  $p = q = 0.5$ , and let us assume that  $n = 101$  and  $m = 55$ , then the subjective breakdown point is 0.5446 and the probability for the median to break down is  $2.0474 \times 10^{-11}$  – an almost zero! If the number of contaminated data is increased to 75, the subjective breakdown point is 0.7426 and the corresponding probability of breakdown is still as small as 0.0024442. These two examples have

clearly demonstrated that with the prior information (4), the sample median can bear a far more than 50% contamination in the data with a negligible probability to break down. In order to see how the subjective breakdown point and its corresponding probability change with  $p$ ,  $m$  and  $n$ , we have chosen  $p = 0.1, 0.2, 0.3, 0.4, 0.5$  and  $n = 11, 21, 51, 101$ , and shown the results in Fig. 2. Obviously, the probability of a subjective breakdown increases rapidly with the decrease of  $p$  from 0.5 to 0. However, it decreases significantly with the increase of sample size  $n$ . In the ideal situation of  $p = 0.5$ , with the increase of samples, the median can bear more percentage of contamination without worrying to break down (compare the purple-red lines in Fig. 2).

In the similar manner, we can use (4) to investigate the subjective breakdown point of the  $\alpha$ -trimmed mean and its corresponding probability of breakdown. As in the case of the sample median, the subjective breakdown point of the  $\alpha$ -trimmed mean can be twice as large as that in the sense of Hampel (1971) and/or Donoho and Huber (1983). For demonstration purpose, we use the third example in Fig. 2 by setting a breakdown point  $\alpha$  in the sense of Hampel to 0.3. The subjective breakdown points will then be between 0.3 and 0.6. Since the number of contaminated data is not necessarily an integer, we slightly reduce the size of samples from 51 to 50 such that the product of  $50 \times 0.3$  makes an integer. The probabilities of subjective breakdown points are shown in Fig. 3. It can be clearly seen from this figure that



**Fig. 3** Probabilities of subjective breakdown points of the  $\alpha$ -trimmed mean with respect to subjective breakdown points and  $p$ . Here the breakdown point  $\alpha$  is 0.3 (in the sense of Hampel). The two horizontal axes show the subjective breakdown points and the probabilities of positive sign of shifts  $\delta y_i$

if  $p$  is between 0.3 and 0.5, then the  $\alpha$ -trimmed mean has a good subjective breakdown point of up to 0.5, although the original breakdown point is only 0.3 (in the sense of Hampel). In other words, the  $\alpha$ -trimmed mean can bear more contamination in the data if the signs of outliers follow the binomial distribution with  $p$  sufficiently close to 0.5. The subjective breakdown points of other robust estimators and their corresponding probabilities of (subjective) breakdown can be studied in the similar manner but will be omitted here.

#### 4 The effect of weights of observations on robustness

The weights or weight matrix  $\mathbf{P}$  of the observations  $\mathbf{y}$  in the linear model (1) has been one of the indispensable components in the Gauss–Markoff theorem (see e.g. Rao 1973). The property of minimum variance of a linear estimate of the parameters  $\beta$  cannot be obtained without use of  $\mathbf{P}$ . In the literature of robust statistics, the weight matrix  $\mathbf{P}$  has rarely attracted due attention. When weights of observations and correlations are incorporated into robust procedures, emphasis has been solely focused on efficiency, consistency and other asymptotic behaviours of robust estimators (see e.g. Gastwirth and Rubin 1975; Koul 1977; Portnoy 1977; Carroll and Ruppert 1982; Boente and Fraiman 1989; Xu 1989b; Yang et al. 2002).

Although the weights of observations  $\mathbf{P}$ , together with the design matrix  $\mathbf{A}$  and the observations with contaminated outliers, are the three basic components in regression or the

linear model (1), they have surprisingly played almost no role in the development of concepts in robust statistics. Concepts used in regression diagnostics such as influential observations, leverage points and hat matrices are solely based on the design matrix  $\mathbf{A}$  (see e.g. Belsley et al. 1980; Chatterjee and Hadi 1986; Barnett and Lewis 1994). Fortunately, the need of practical application of these concepts has led geodesists to incorporate the weight matrix into the measures related to these concepts (see e.g. Baarda 1968; Schaffrin 1997; Hekimoğlu 1998). In particular, as soon as the concept of breakdown point (in the sense of Hampel) is concerned, it has been always referred either to outliers in the  $x$ -direction (the design matrix  $\mathbf{A}$ ) or outliers in observations (see e.g. Donoho and Huber 1983; Rousseeuw 1984).

It is well known that the least absolute deviation (or  $L_1$ -norm method) and M-estimates are robust if there exist no outliers in the data of the matrix  $\mathbf{A}$  (see e.g. Donoho and Huber 1983; Bloomfield and Steiger 1983; Dodge 1997). In this section, we will show that robust statistical methods can still not be robust against outliers in observations, even if the design matrix  $\mathbf{A}$  is assumed to contain no outliers. More specifically, we will first demonstrate that the (weighted)  $L_1$ -norm method and M-estimates are indeed vulnerable to outliers in observations. Since the least median of squares method has been claimed to have a maximum breakdown point of 0.5 and has been widely applied practically, we will try to further explain the low breakdown phenomenon observed by Wang and Suter (2003) from the point of view of weights, although the simulated examples in Sect. 2 have

already shown that a single outlier can still break the LMS method down, even without clustering and concentration.

As in Sect. 2, we assume  $n$  independently, identically distributed samples  $y_1, y_2, \dots, y_n$  with variance  $\sigma^2$  and mean  $\bar{y}$ . In order to investigate the effect of large weights on robustness of the  $L_1$ -norm method and M-estimates, we replace the  $i$ th sample  $y_i$ . The new sample is assumed to come from the same family of distributions but with a different variance  $\sigma_0^2$  and  $\sigma_0^2 \ll \sigma^2$ . In other words, the relative weight  $p_i = \sigma^2/\sigma_0^2$  is significantly large such that the root of  $p_i$  is much larger than the size of samples  $n$ . We further assume that this new sample  $y_i$  is an outlier. Then the weighted  $L_1$ -norm criterion becomes:

$$\min: \sum_{j=1, j \neq i}^n |y_j - \bar{y}| + \sqrt{p_i} |y_i - \bar{y}|, \quad (6)$$

where  $|x|$  stands for the absolute value of  $x$ . If  $p_i = 1$ , then (6) is the well-known method of least absolute deviation and will result in the median of samples as the estimate of  $\bar{y}$ .

Since  $p_i$  has been assumed to be very large and  $y_i$  is an outlier, the optimization problem (6) must produce the estimate of  $\bar{y}$  that is very close to  $y_i$  itself. In this case, the objective value of (6) is roughly equal to  $(n-1)\Delta$ , where  $\Delta$  is the absolute value of the mean of  $(y_j - y_i)$  for all  $j = 1, 2, \dots, (i-1), (i+1), \dots, n$ . Otherwise, if (6) would have produced something that is close to any of the samples other than the outlier  $y_i$ , then the objective value of (6) would have roughly been equal to  $\sqrt{p_i}\Delta$ . Since  $\sqrt{p_i}$  has been assumed to be much larger than  $n$ ,  $\sqrt{p_i}\Delta$  should be much larger than  $(n-1)\Delta$ . Thus the weighted  $L_1$ -estimate must be close to the outlier  $y_i$  and is obviously not robust at all. The same reasoning is true in the linear model (1).

In the case of M-estimation of location, we have the objective function:

$$\min: \sum_{j=1, j \neq i}^n \rho(y_j - \bar{y}) + \rho(\sqrt{p_i}(y_i - \bar{y})), \quad (7)$$

where  $\rho(\cdot)$  is an appropriately chosen symmetric function with a unique minimum at zero (see e.g. Huber 1981; Donoho and Huber 1983; Rousseeuw and Leroy 1987; Jurečková and Sen 1996). If  $\rho(\cdot)$  is convex and its derivative is bounded, the M-estimate of  $\bar{y}$  has been reported to have a maximum breakdown point of 0.5 (see e.g. Donoho and Huber 1983) in the one dimensional case. Taking the Huber's  $\rho(\cdot)$  function as an example, and by using the same assumptions as in the case of the weighted  $L_1$ -norm method above, we should immediately conclude that the M-estimate of  $\bar{y}$  must be close to the outlier  $y_i$  in order to minimize the objective function (7). In fact, the condition can even be weakened as that  $p_i$  is larger than  $n$ . Once again, we see that an outlier with a significantly large weight can simply nullify the robustness of Huber's M-estimates. Of course, the same argument applies to M-estimates of Huber's type in the linear model (1) but will not be repeated here.

The effect of weights of observations may also be used to explain why clustered outliers are most difficult to detect.

In fact, by rewriting, for example, the second term of (6) as  $\sum_{j=1}^{\sqrt{p_i}} |y_i - \bar{y}|$ , we may re-interpret this second term as an idealized model of clustered outliers which consists of  $\sqrt{p_i}$  independent observations  $y_i$  of unit weight, although, in reality, clustered outliers would (slightly) be different from each other. Thus a sufficient number of clustered outliers may likely fail some of robust estimators. For example, assume that clustered outliers are sufficiently concentrated together and their number is sufficiently large such that together with a minority of good data, they form a majority of data to fit the model. Then, we can expect theoretically that such a data set will almost certainly fail the LMS method; this explains very well the results of low breakdown points for the LMS method reported by Wang and Suter (2003).

## 5 Sign-constrained robust estimation

Although M-estimates, L-estimates and R-estimates are most popular in the robust estimation of the parameters in the linear model (1) and the scale parameter or the unknown variance of unit weight (see e.g. Huber 1981; Hampel et al. 1986; Rousseeuw and Leroy 1987; Jurečková and Sen 1996), they can tolerate only up to about 30% of data contamination (Rousseeuw 1984), except for some special situations, for example, one-dimensional M-estimation of location (Donoho and Huber 1983). The most promising robust methods, namely, the least median of squares and the least trimmed squares (Rousseeuw 1984; Rousseeuw and Leroy 1987), though claimed or proved to have a (maximum) breakdown point of 0.5, have been shown to break down by a less than 40% of contamination or clustered outliers (Wang and Suter 2003) and even worse, by a single outlier without obvious clustering and concentration, as demonstrated in Sect. 2.

The two most robust estimators with a 0.5 breakdown point are the sample median and the repeated median. The former has been well known but is valid only in the one-dimensional case. Since the median is irrelevant to the weights of the observations, its robustness is not affected by the weight matrix  $\mathbf{P}$ . The repeated median was proposed by Siegel (1982). Unlike the sample median, it is applicable, in principle, to problems of any dimension. However, since the repeated median is based on the generalized least squares solutions of all the fundamental subsystems with  $t$  observations, near multi-collinearity in only some of the good data can significantly deteriorate the reliability of the method. With the increase of percentage of contamination, near multi-collinearity among some of good data may cause the method to fail, since a small random error can result in an arbitrarily large error in the solution to such near multi-collinear fundamental subsystems. As a cost of achieving the highest possible breakdown point (in the sense of Hampel), the efficiency of these two robust estimators is not satisfactory, however. The LMS method is also known to have a low efficiency.

The question now is whether one can really construct alternative robust estimators with the maximum breakdown point of 0.5 in the multiple dimensional case or linear model,

in addition to the repeated median. For simplicity of discussion but without loss of generality, we will assume  $\mathbf{P}$  is diagonal with the elements  $p_i$  ( $i = 1, 2, \dots, n$ ), since the idea to be presented can be readily extended to a full positive definite matrix. Although prior information on the signs of outliers was assumed in Sect. 3 to define the concept of subjective breakdown point, we assume no such prior information in this section. As an attempt to provide an affirmative answer to the above question, we propose a new robust estimator, which is defined and formulated as follows:

$$\min: (\mathbf{y} - \mathbf{A}\boldsymbol{\beta})^T \mathbf{P}_R (\mathbf{y} - \mathbf{A}\boldsymbol{\beta}) / r_{P_R}, \quad (8a)$$

subject to the following constraints:

$$\sum_{i=1}^n \operatorname{sgn}(y_i - \mathbf{a}_i \boldsymbol{\beta}) = 0, \quad (8b)$$

where the  $i$ th diagonal element of  $\mathbf{P}_R$  is determined as follows:

$$p_i^R = \begin{cases} p_i, & \text{if } |y_i - \mathbf{a}_i \boldsymbol{\beta}| \leq c \hat{\sigma} / \sqrt{p_i}; \\ 0, & \text{otherwise.} \end{cases} \quad (8c)$$

Here  $c$  is a positive constant. If necessary, one can also use a variable parameter for each observation proposed by Xu (1993).  $\hat{\sigma}$  is an estimate of the standard deviation of unit weight. As a common practice, we suggest that  $\hat{\sigma}$  is set to  $1.483 \operatorname{med}(\sqrt{p_i} |y_i - \mathbf{a}_i \boldsymbol{\beta}|)$ , where  $\operatorname{med}(\sqrt{p_i} |y_i - \mathbf{a}_i \boldsymbol{\beta}|)$  is the median of the absolute values of the residuals  $\sqrt{p_i} (y_i - \mathbf{a}_i \boldsymbol{\beta})$ . Alternatively, one may also use half of the observations with least absolute residuals to estimate  $\hat{\sigma}$ .  $\mathbf{a}_i$  has been given in Sect. 1,  $r_{P_R}$  is the rank or number of non-zero (diagonal) elements of  $\mathbf{P}_R$ , the sign function  $\operatorname{sgn}(x)$  is defined by

$$\operatorname{sgn}(x) = \begin{cases} 1, & \text{if } x > 0; \\ 0, & \text{if } x = 0; \text{ and} \\ -1, & \text{if } x < 0. \end{cases}$$

After the minimization problem (8) is solved, we meanwhile obtain a subset of observations satisfying both (8b) and (8a). We then solve the unconstrained least squares problem by using the identified observations to derive the final robust estimate of the parameters  $\boldsymbol{\beta}$ , which is defined to be the sign-constrained robust solution. This last step is technically trivial but can be rather important in further improving the efficiency of the robust solution, depending on the percentage of data contamination. We say the step is *trivial*, since we do nothing more but simply use the observations from the minimization process to compute the final robust solution. From this point of view, the sign-constrained robust method is fundamentally different from other multi-step robust procedures in the literature (see e.g. Yohai 1987), not to mention that high breakdown robust estimators used as an initial robust estimate may not be of the high breakdown point as claimed. If contamination is sufficiently significant or close to 50 percent, the step will become rather important; otherwise, the improvement in efficiency should be marginal. If the number of observations is even, the estimate of  $\boldsymbol{\beta}$  would have to be limited to

the rounding error of a computer in order to satisfy the second constraint (8b) for some problems. If the problem of the machine error is to be avoided, (8b) is, in this case, numerically equivalent to

$$\left| \sum_{i=1}^n \operatorname{sgn}(y_i - \mathbf{a}_i \boldsymbol{\beta}) \right| = 1,$$

which does not affect the final robust solution, nevertheless. In fact, the above condition would frequently be encountered as an intermediate solution in the search for the global optimal solution of (8), if the number of observations is even.

The robust estimator defined above should be intuitively very interesting for a number of reasons: (1) The robustness of the method is directly implemented as constraints. In fact, the constraint (8b) is essentially a key point of the sample median. The objective function (8a) and the constraint (8c) is then to force the fitted hyperplane to the good data or majority of good data. The method is therefore robust and is expected to have the maximum breakdown point of 0.5. This should be intuitively obvious by itself, since, for example, in the extreme or worst case that half of the data are supposed to be outliers, the constraint (8b) will force the estimate to cut the data into two parts with the same numbers of positive and negative residuals. The minimization formulation (8a), together with the constraint (8c), will force the cutting to clearly separate the good data from the outliers, and will further push it to the side of the good data in order to attain the minimum value of the objective function (8a). If some of good data or outliers are highly concentrated, then extra work will be needed to find a correct estimate, as will be clear in the next section of examples. The sign constraint (8b) is justified, since the event of a residual (a continuous random variable) being equal to zero theoretically has zero probability, unless the corresponding observation is completely occupied by an unknown parameter. In this latter (trivial) case, the residuals of such observations always have zero residuals, which again does not bother the justification of use of sign constraint (8b); (2) The robustness of the method should not be affected by the weight system  $\mathbf{P}$ , since the weights have no role to play in one of the key constraints (8b); (3) Unlike other robust estimators such as the trimmed mean, the sign-constrained robust method is flexible and automatically adapts itself to the data at hand through both constraints (8b) and (8c). For example, if contamination of the data is supposed to be of 10%, while removing the effect of this 10% contaminated data on the estimated parameters, the robust method should be able to retain nearly the rest of 90% of the data, or as many good data as possible, as a direct consequence of (8c); and finally, (4) The method is directly based on the least squares principle (8a). Thus the method may be called *sign-constrained robust least squares*. This confirms the argument by Rousseeuw (1984) that it is not the *squares* that are responsible for the vulnerability of the least squares method to outliers.

On the other hand, we have also clearly seen that the operation *sum* in least squares is innocent as well, as otherwise might be implied in Rousseeuw (1984). Since the LMS method of Rousseeuw (1984), the repeated median of Siegel



(1982) and/or other median-based robust estimators only utilize a least possible number of good data to derive the robust solution, the robustness and solution are easily affected by partial near multi-collinearity among part of the data or if some of the data are clustered together (Rousseeuw 1984; Stefanski 1991; Hettmansperger and Sheather 1992; Sheather et al. 1997; Ellis 1998). However, our robust method will adapt itself to and attempt to utilize all the good data to achieve the robustness for the estimate of  $\beta$  and, as a result, will not be affected by partial multi-collinearity. Therefore, our robust least squares will be able to achieve both high breakdown point and high efficiency simultaneously. This has clearly invalidated the statement on page 357 by Stefanski (1991):

*“Thus it can be stated quite generally that any estimator possessing the exact fit property of any order  $> 1/n$  has arbitrarily low efficiency with respect to least-squares in finite samples.”*

Therefore, our robust method should bring the dilemma of conflict between high breakdown and efficiency (see also He 1994) to a happy end. The variance–covariance matrix of the robust estimator is also easy to compute.

Since the method is obviously nonlinear and nonsmooth, we will have to resort to global optimization methods (see e.g. Hansen 1992; Xu 2002, 2003a) to solve (8) numerically. Although (8) appears to be a global optimization without constraints on  $\beta$ , and in order not to pre-set too small a box and miss the global optimal solution(s), one can actually determine a suboptimal solution to (8), without any difficulty, by slightly adjusting the least squares solution of  $\beta$  to satisfy the constraint (8b). Assuming that the suboptimal solution corresponds to the suboptimal objective value, say  $F_0$ , then one can readily determine from  $F_0$  the initial bounding box for  $\beta$  by using interval analysis, as in the case of determining the bounding box in mixed integer linear models for Voronoi cells (Xu 2003b).

## 6 Examples

In this section, we will simulate a number of typical examples to demonstrate how the sign-constrained robust method works and performs. More specifically, we will design examples with influential observations, with 50% contamination, with scattered and highly concentrated clustering outliers, and with more than 50% contamination. In the rest of this section, we will use a critical value of  $c$  equivalent to 2.5 times standard deviation for data with up to 50% of contamination, and equivalent to 2.0 times standard deviation for data with more than 50% of contamination

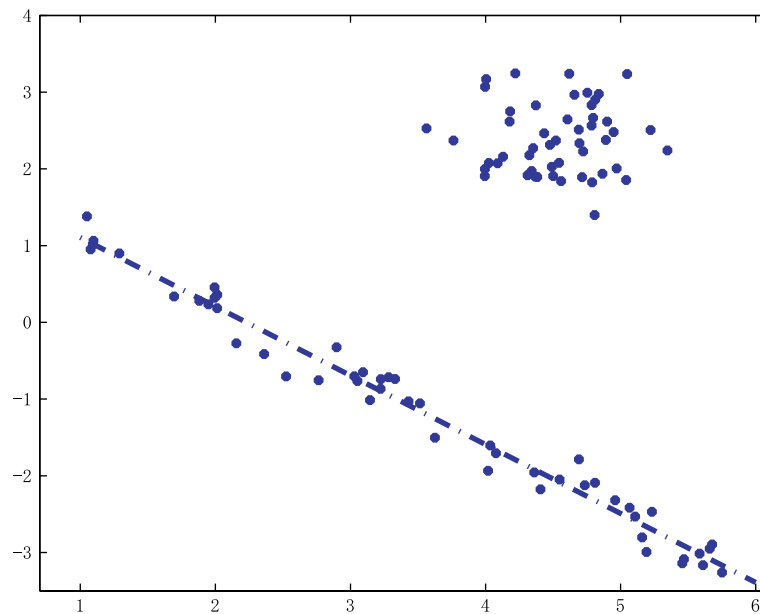
Let us first have a look at examples of one dimension. We simulate five random numbers by using a Gaussian distribution with mean 0 and variance 0.5 and five outliers. The five good data are generated by using MATLAB and are given as follows: 0.0805, 0.7543, 0.0419,  $-0.0676$  and  $-0.5885$ . Five outliers are arbitrarily given as follows: 10.0, 10.1, 20.0,

35.0 and 50.0. Because of the constraint (8b), the solution to the minimization problem (8) must be between 0.7543 and 10.0. If the solution would approach 10.0 from left, then we have the median of 9.9581 for  $\text{med}\{|y_i - 10.0|\}$  and the cost value of 74.5658. On the other hand, if the estimate of the mean approaches 0.7543 from right, we obtain the median of 1.3428 for  $\text{med}\{|y_i - 0.7543|\}$  and the cost value of 0.6880. Thus the solution to the minimization problem is 0.7543. As a last step of our robust method, we leave all the constraints (8b) and (8c) out, and obtain the final robust estimate of the mean by solving the unconstrained least squares problem (8a) with the observations identified by the optimal solution. The final robust solution is equal to 0.0441 – exactly the average of the five good data. We can also see from this example that the last (trivial) step in our robust method does improve the efficiency significantly. Here due to the 50% data contamination, our first step of minimization is focused on achieving robustness by kicking out outliers but at the cost of forcing the solution to the margin of the good data. The unconstrained least squares is simply to adjust among the good data. As a result of the adjustment, the best possible efficiency is obtained. Note, however, that in this special case, even the sample median will fail, since the median in the case of even number of samples is equal to the average of 0.7543 and 10.0 by definition. The same can be said of the repeated median of Siegel (1982), since it also requires at least one more good data than the number of outliers.

Now let us replace some of outliers with good data and see how the sign-constrained robust method would perform. For example, if we replace the last two outliers by  $-0.4892$  and  $0.6067$  from the same distribution as the five good data, then the solution to the minimization problem must be between 0.0805 and  $0.6067$ , again due to the constraint (8b). By repeating the same procedure as in the case before the replacement of outliers with the good data, we have removed the three outliers, retained the seven good data and obtained the robust mean of 0.0483 with variance 0.0714. From this second simple example, we also expect that if the rate of contamination is not sufficiently significant, the least squares adjustment will improve the final robust solution only marginally. With the further decrease of number of outliers, this improvement should be even smaller. Obviously, the mean has adapted itself to the good data. Thus from this point of view, the sign-constrained least squares method may also be called *adaptive trimmed mean*.

The second example is to test the sign-constrained robust method with the example of Sect. 2, which failed the LMS method with a single influential outlier. The sign-constrained robust method has correctly identified the 100 good data. The two parameters have been estimated to be  $-1.9400$  and  $0.6048$  with standard deviations of 0.0811 and 0.1209, respectively, which are essentially the LS solution of the example without the contaminated influential point. The fitted line has been shown in Fig. 1 (the dash-dotted line).

As the third group of examples, we will demonstrate the ability of the sign-constrained robust method against exactly 50% contamination of data in the linear model (1). We know



**Fig. 4** 50 simulated good data, 50 scattered outliers, and the fitted (*dash-dotted*) line obtained by the sign-constrained robust method. The horizontal and vertical axes show the values of  $x_i$  and  $y_i$ , respectively

from Sect. 3 that if all outliers are one-sided, they are most challenging to a robust method. This is exactly the scenario we would like to investigate. As a first step, we simulated the following regression line

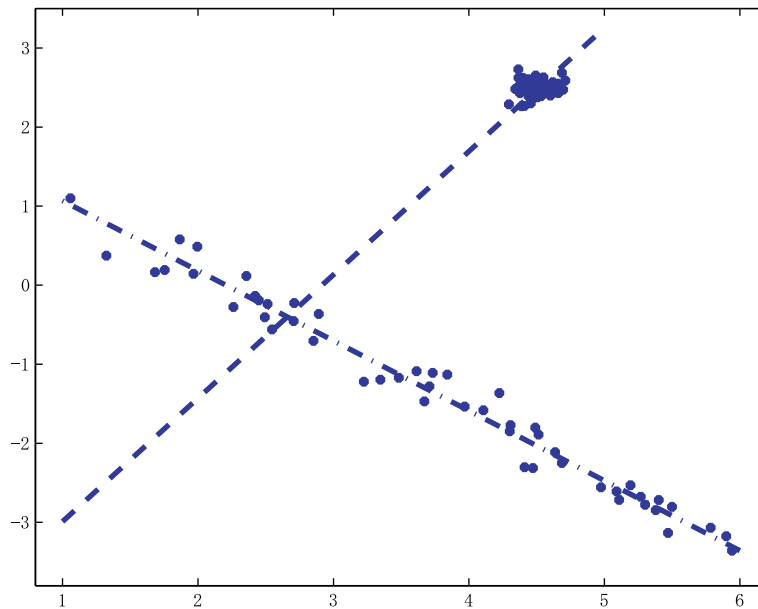
$$y_i = \beta_1 + x_i\beta_2 + \epsilon_i, \quad (9)$$

by using the true parameters  $\beta_1 = 2.0$  and  $\beta_2 = -0.9$  and by generating 50  $x_i$  uniformly distributed over  $[1, 6]$  and 50 Gaussian errors  $\epsilon_i$  with mean 0 and standard deviation 0.2. We then simulated the same number of scattered outliers, influential outliers and highly concentrated outliers. All the simulated outliers are put on one side of the regression line. We generated 50 scattered outliers by using a normal distribution with standard deviation 0.5 for both  $x$  and  $y$  components and centered at the point  $[4.5, 2.5]$ , 50 highly concentrated outliers by using a normal distribution with standard deviation 0.1 – half of the standard deviation of good data, again for both  $x$  and  $y$  components, and centered at the same point as above, and finally 50 influential points by using a uniform distribution with  $x_i$  over  $[1000, 2000]$  and  $y_i$  over  $[0, 1000]$ . Since these 50 influential outliers cannot nicely displayed together with the 50 good data, we will not show them here. For the example with either 50 scattered or 50 influential outliers, the sign-constrained robust method has correctly removed the outliers and kept 50 good data. The final estimation results are  $\hat{\beta}_1 = 2.0050$  and  $\hat{\beta}_2 = -0.8992$ , with standard deviations of 0.0744 and 0.0194, respectively. Of course, they are exactly the LS solution with the 50 good data points. Thus for these two examples, the sign-constrained robust method has successfully removed the effect of outliers and achieved the maximum possible efficiency. The 50 good data and 50 scattered outliers, together with the fitted line of regression, are shown in Fig. 4. For the example with 50 highly concen-

trated outliers, one may readily use Euclidean norm to find them before applying any robust methods.

In order to see whether the sign-constrained robust method can handle such data by separating the good data from the outliers, we first simply apply the method to the original data. As the second step, we apply the method first to the removed data. Then if an original data point is not violated by the new fitting, it is added into this second step. By doing so, we obtained two fitted lines shown in Fig. 5. Obviously, the sign-constrained robust method can correctly separate the good data from the highly concentrated outliers. The final estimates of the two parameters are obtained from 48 good data and are equal to 1.9530 and  $-0.8854$ , with standard deviations of 0.0868 and 0.0216, respectively. We can also expect from this example that if about 50% of good data are extremely concentrated (at a point in the two-dimensional case or on a  $(t - 1)$ -dimensional subspace in the  $t$ -dimensional case), then we will encounter the same problem as in the case of 50% of highly concentrated outliers and have to resort to the above two-step procedure in order to obtain a correct estimate.

As the last set of examples, we will investigate how the sign-constrained robust method can be used to handle more than 50% of data contamination, which may often be encountered in object reconstruction from noisy data. To begin with, we simulated 150 good data by using a normal distribution with mean 0 and standard deviation 0.1. The 150 variables  $x_i$  were generated by using a uniform distribution over  $[1, 100]$ . The true values of two parameters are equal to 1.0 and 2.0101, respectively. We then simulated 400 outliers by using a uniform distribution with  $x$ -component over  $[1, 100]$  and  $y$ -component over  $[1, 200]$ . In order to further complex the problem, we simulated the second example by adding two different sets of scattered outliers at different positions to the



**Fig. 5** 50 simulated good data, 50 highly concentrated outliers and the fitted lines obtained by the sign-constrained robust method. The *dashed line* is obtained without applying the two-step procedure, while the final (*dash-dotted*) line is found after applying the two-step procedure. The horizontal and vertical axes show the values of  $x_i$  and  $y_i$ , respectively

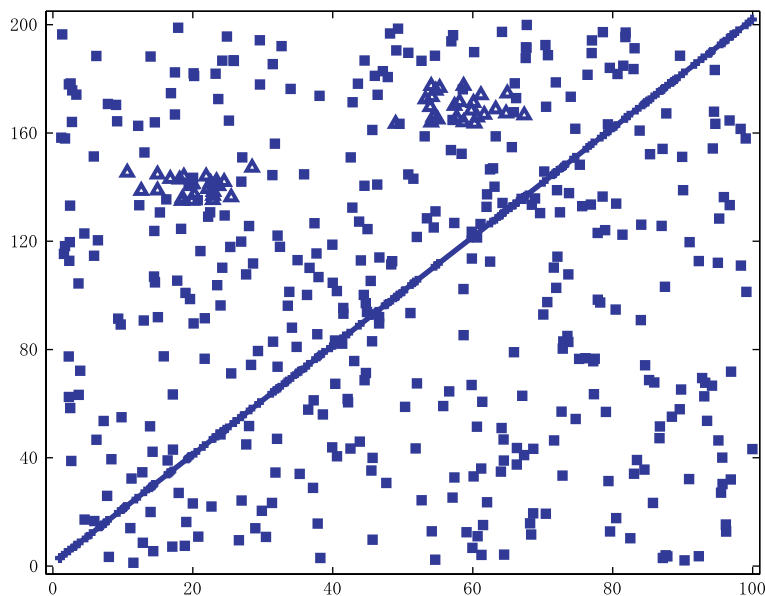
simulated data or the first example, the number of each set of scattered outliers being equal to 25. Thus the two examples contain 72.73% and 75% of data contamination, respectively. More specifically, the two examples have 211 and 261 outliers above the true line, respectively.

Since no robust methods could directly handle more than 50% of contamination, we will implement the iterative version of the sign-constrained robust method and see how it handles such data sets. The iterative version is to repeatedly apply the sign-constrained robust method to the newly identified data until the same set of data is found. In other words, we first apply the method to the original data set. After obtaining the new set of data satisfying the conditions of the sign-constrained robust method, we apply the method to the new data set. We repeat the same procedure until no more data can be removed by the sign-constrained robust method. The iterative version of the method has produced the same results for both data sets and, in total, 128 good data have been identified and used to compute the final robust solution, although the numbers of iteration are different and equal to 8 and 9 for the first and second example, respectively. The estimates of the two parameters are equal to 1.0093 and 2.0097, with standard deviations of 0.0173 and 0.0003, respectively. The 150 good data, 450 uniformly distributed outliers, 50 scattered outliers and the fitted line are all shown in Fig. 6. The iterative version of the sign-constrained robust method has been shown to perform excellently in removing the effect of more than 50% data contamination and meanwhile achieving excellent efficiency. In fact, even if the contamination is not more than 50%, but if almost all of the outliers are one-sided, we still strongly recommend running the iterative version. In such a case, since the scale parameter used in the constraint

(8c) is over-estimated, outliers of intermediate size, say three to ten times of standard deviation, will remain if they do exist in the data. By implementing the iterative version, the issue of over-estimation will be rectified and the effect of such intermediate-sized outliers on the estimated parameters is removed. Of course, if outliers are found to scatter on both sides of the fitted hyperplane by the sign-constrained robust method, then running the iterative version is not necessary.

## 7 Conclusions

Robust estimation has been one of the most intensive research topics in statistics since the publication of Huber (1964) pioneering paper and has found widest possible applications in science and engineering. All robust estimators have been known to have a breakdown point of only up to around 0.3 in linear models with at least two unknown parameters, except for repeated medians (Siegel 1982), the LMS and least trimmed squares methods (Rousseeuw 1984). Although the method of repeated medians is known to have a 0.5 breakdown point, this can only be true for noiseless or well designed data. If data contamination is sufficiently large and if some of the good data are close to each other, the failure of repeated medians should be routinely observed. The LMS and least trimmed squares methods have been found to break down if some of data (good or bad) are concentrated together (see e.g. Stefanski 1991; Dodge 1992; Wang and Suter 2003). In this paper, we have further shown that the LMS can still break down due to a single outlier, even if no clustering phenomenon can be observed in the data. On the other hand, a high breakdown robust method has been known to be least



**Fig. 6** 150 simulated good data (with mark +), 400 uniformly distributed outliers (with square mark), 50 scattered outliers (with triangle mark) around two different points and the fitted (solid) line obtained by the iterative version of the sign-constrained robust method. The horizontal and vertical axes show the values of  $x_i$  and  $y_i$ , respectively

efficient (see e.g. Stefanski 1991; He 1994). To achieve both a high breakdown point and efficiency, one has to combine high breakdown methods with diagnostic tools or M-estimation (see e.g. Yohai 1987). Unfortunately, (prerequisite) high breakdown robust estimators such as repeated medians and the LMS method can break down routinely.

We have proposed the sign-constrained robust method, which is intuitively very appealing and has a maximum breakdown point of 0.5. Unlike other robust estimators with a high breakdown point such as repeated medians (and/or the LMS method), which are based on a least possible number of data to derive the corresponding estimate of the parameters, the sign-constrained robust method automatically adapts itself to the data and attempts to use a maximum possible number of good data to estimate the model parameters. Thus it is also highly efficient. Since the objective function is essentially identical with least squares, the sign-constrained robust method may also be properly called *sign-constrained robust least squares* and is straightforwardly applicable to correlated data/observations; and consequently the estimated results are undoubtedly unbiased and comparable with least squares in terms of efficiency or accuracy. Thus the sign-constrained robust least squares method is the first perfect combination of best robustness and best efficiency.

Unlike the least median of squares method and repeated medians, the sign-constrained robust least squares method attempts to employ a maximum possible number of good data to derive the robust solution and thus will not be affected by partial multi-collinearity among part of the data. Because of the constraints (8b) and (8c), robustness of the method is not affected by the weights of observations either. As a by-product, we can also obtain a robust and efficient estimate of scale parameter or unknown variance of unit weight. In

Earth sciences and image processing, one often has to reconstruct objects from data which may contain more than 50% of contamination, for example, due to the ambiguity of earthquake focal mechanism or due to scanning noise. As a result, the iterative version of the sign-constrained robust method has been implemented and shown to be capable of resisting against more than 50% of data contamination. In fact, if the contamination is sufficiently large (though only up to 50%) and lies on one side of the fitted surface, then the iterative version should be applied in order to further eliminate the effect of outliers of intermediate size (if any) on the estimated parameters.

Although M-estimates have been proved to have a breakdown point of  $1/(t+1)$  (see e.g. Maronna 1976; Donoho and Huber 1983), we have shown that the weights of observations can invalidate this result and bring the breakdown point of M-estimates of Huber's type to zero. We have also shown that the weights of observations can also bring the breakdown point of the  $L_1$  method to zero, even if no influential outliers are contained in the data. In fact, the weighting system is one of the three basic components in the linear model and should be incorporated in the development of robustness measures. By assuming a prior distribution for the signs of outliers, we have developed the concept of subjective breakdown point, which may be thought of as an extension of stochastic breakdown by Donoho and Huber (1983) but can be important in explaining real-life problems in Earth sciences and image reconstruction.

Finally, we would like to note that the sign-constrained robust least squares method is simply a robust version of the least squares method with the ability to resist against outliers. Thus it is applicable, in principle, to any science and engineering problems, for which the least squares method

can be used to solve, not to mention geodetic problems which motivated the invention of least squares more than two centuries ago, as far as outliers are an issue of concern. Since the new method is of the nature of global optimization, a big problem (of high dimension) will need a good computing resource to support.

**Acknowledgements** This work is supported by a Grant-in-Aid for Scientific Research C(2)16540386.

## References

- Adichie JN (1967) Estimates of regression parameters based on rank tests. *Ann Math Stat* 38:894–904
- Andrews DF (1974) A robust method for multiple linear regression. *Technometrics* 16:523–531
- Angelier J (2002) Inversion of earthquake focal mechanisms to obtain the seismotectonic stress IV – a new method free of choice among nodal planes. *Geophys J Int* 150:588–609
- Anscombe FJ (1960) Rejection of outliers. *Technometrics* 2: 123–147
- Baarda W (1968) A testing procedure for use in geodetic networks. Netherlands Geodetic Commission Publications on Geodesy New Series Vol. 2, No.5, Delft
- Barnett V, Lewis T (1994) Outliers in statistical data. 3rd edn. Wiley, Chichester
- Belsley DA, Kuh E, Welsch RE (1980) Regression diagnostics: identifying influential data and sources of collinearity. Wiley, New York
- Beran R (1982) Robust estimation in models for independent non-identically distributed data. *Ann Stat* 10:415–428
- Bickel PJ (1973) On some analogues to linear combinations of order statistics in the linear model. *Ann Stat* 1:597–616
- Bloomfield P, Steiger WL (1983) Least absolute deviations: theory, applications, and algorithms. Birkhäuser, Boston
- Boente G, Fraiman R (1989) Robust nonparametric regression estimation for dependent observations. *Ann Stat* 17:1242–1256
- Carroll RJ, Ruppert D (1982) Robust estimation in heteroscedastic linear models. *Ann Stat* 10:429–441
- Chatterjee S, Hadi AS (1986) Influential observations, high leverage points, and outliers in linear regression (with discussions). *Stat Sci* 1:379–416
- Cheney EW (1966) Introduction to approximation theory. McGraw-Hill, New York
- Dodge Y (1992) Thoughts on real data and statistics. In: Dodge Y (ed),  $L_1$ -statistical analysis and related methods. Elsevier, Amsterdam, pp 3–7
- Dodge Y (1997) LAD regression for detecting outliers in response and explanatory variables. *J Multivariate Anal* 61:144–158
- Donoho DL, Huber PJ (1983) The notion of breakdown point. In: Bickel PJ, Doksum KA, Hodges JL Jr (eds) *A Festschrift for Erich L. Lehmann* Wadsworth, Belmont, pp 157–184
- Ellis SP (1998) Instability of least squares, least absolute deviation and least median of squares linear regression (with discussions). *Stat Sci* 13:337–350
- Gastwirth JL, Rubin H (1975) The behavior of robust estimators on dependent data. *Ann Stat* 3:1070–1100
- Gauss CF (1964) *Abhandlungen zur methode der kleinsten quadrate*, translated into German by A. Börsch & P. Simon, and reproduced by Physica-Verlag, Würzburg
- Gauss CF (1995) Theory of the combination of observations least subject to errors, translated by G.W. Stewart, SIAM, Philadelphia
- Hampel FR (1971) A general qualitative definition of robustness. *Ann Math Stat* 42:1887–1896
- Hampel FR (1974) The influence curve and its role in robust estimation. *J Am Stat Assoc* 69:383–393
- Hampel FR, Ronchetti E, Rousseeuw PJ, Stahel WA (1986) Robust statistics: the approach based on influence functions, Wiley, New York
- Hansen E (1992) Global optimization using interval analysis. Marcel Dekker, New York
- He X (1994) Breakdown versus efficiency – your perspective matters. *Stat Prob Lett* 19:357–360
- Hekimoğlu S (1998) Application of equiredundancy design to M-estimation. *J Surv Eng* 124:103–124
- Hettmansperger TP, Sheather SJ (1992) A cautionary note on the method of least median squares. *Am Stat* 46:79–83
- Hodges Jr JL, Lehmann EL (1963) Estimates of location based on rank tests. *Ann Math Stat* 34:598–611
- Huber PJ (1964) Robust estimation of a location parameter. *Ann Math Stat* 35:73–101
- Huber PJ (1981) Robust statistics. Wiley, New York
- Huber PJ (1984) Finite sample breakdown of M- and P-estimators. *Ann Stat* 12:119–126
- Jaeckel LA (1972) Estimating regression coefficients by minimizing the dispersion of the residuals. *Ann Math Stat* 43:1449–1458
- Jurečková J (1971) Nonparametric estimate of regression coefficients. *Ann Math Stat* 42:1328–1338
- Jurečková J, Sen PK (1996) Robust statistical procedures: asymptotics and interrelations. Wiley Interscience, New York
- Koch K-R (1999) Parameter estimation and hypothesis testing in linear models. 2nd edn. Springer, Berlin Heidelberg New York
- Koul HL (1977) Behavior of robust estimators in the regression model with dependent errors. *Ann Stat* 5:681–699
- Maronna RA (1976) Robust M-estimators of multivariate location and scatter. *Ann Stat* 4:51–67
- Mood AM, Graybill FA, Boes DC (1974) Introduction to the theory of statistics, 3rd edn. McGraw-Hill, Auckland
- Pope J (1976) The statistics of residuals and detection of outliers, NOAA Technical Report Nos 66, NGS 1.
- Portnoy SL (1977) Robust estimation in dependent situations. *Ann Stat* 5:22–43
- Rao CR (1973) Linear statistical inference and its applications, 2nd edn. Wiley, New York
- Rocke DM, Woodruff DL (1996) Identification of outliers in multivariate data. *J Am Stat Assoc* 91:1047–1061
- Rousseeuw PJ (1984) Least median of squares regression. *J Am Stat Assoc* 79:871–880
- Rousseeuw PJ, Leroy AM (1987) Robust regression and outlier detection. Wiley, New York
- Schaffrin B (1997) Reliability measures for correlated observations. *J Surv Eng* 123:126–137
- Sheather SJ, McKean JW, Hettmansperger TP (1997) Finite sample stability properties of the least median of squares estimator. *J Stat Comput Simul* 58:371–383
- Siegel AF (1982) Robust regression using repeated medians. *Biometrika* 69:242–244
- Stefanski LA (1991) A note on high-breakdown estimators. *Stat Prob Lett* 11:353–358
- Stigler SM (1977) Do robust estimators work with real data? *Ann Stat* 5:1055–1098 (with discussions)
- Stigler SM (1986) The history of statistics, Harvard University Press, Cambridge
- Stromberg AJ (1993) Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression. *SIAM J Sci Comput* 14:1289–1299
- Wang H, Suter D (2003) Using symmetry in robust model fitting. *Patt Recog Lett* 24:2953–2966
- Welsh AH (1987) The trimmed mean in the linear model. *Ann Stat* 15:20–36
- Xu PL (1989a) Statistical criteria for robust methods. *ITC J No.1*:37–40
- Xu PL (1989b) On robust estimation with correlated observations. *Bull Géod* 63:237–252
- Xu PL (1993) Consequences of constant parameters and confidence intervals of robust estimation. *Boll Geod Sc Aff* 52:231–249

- 
- Xu PL (2002) A hybrid global optimization method: the one-dimensional case. *J Comput Appl Math* 147:301–314
- Xu PL (2003a) A hybrid global optimization method: the multi-dimensional case. *J Comput Appl Math* 155:423–446
- Xu PL (2003b) Voronoi cells, probabilistic bounds and hypothesis testings in mixed integer linear models, paper presented at the International Union of Geodesy and Geophysics 2003 Sapporo Assembly, June 30–July 11, 2003, Sapporo, Japan
- Xu PL (2004) Determination of regional stress tensors from fault-slip data. *Geophys J Int* 157:1316–1330
- Yang Y, Song L, Xu T (2002) Robust estimator for correlated observations based on bifactor equivalent weights. *J Geod* 76:353–358
- Yohai VJ (1987) High breakdown-point and high efficiency robust estimates for regression. *Ann Stat* 15:642–656