Changing residuals by reweighting one or more observations

Şerif Hekimoğlu and Uğur Şanlı

Summary

M-Estimation usually uses iteratively reweighted Least Squares Estimation (LSE). At each stage, the weights of some of the observations are reduced with respect to the absolute values of the related residuals from LSE. Thus, it is expected that the absolute values of these residuals increase. Is this really always the case? Further, what happens if we increase the new weights of some of the observations? In this paper, we have investigated how absolute values of residuals of one or more observations from LSE change after changing the weights of the corresponding observations. We have proved that weight reduction only causes an increase in the absolute values of the LSE residuals with a certain probability. However, reducing the weights of some of the randomly chosen observations may create the opportunity to detect small outliers.

Zusammenfassung

In der M-Schätzung wird im Allgemeinen die Methode der kleinsten Quadrate mit iterativer Neugewichtung benutzt. Dabei werden die neuen Gewichte einiger Beobachtungen entsprechend der absoluten Werte ihrer Verbesserungen in jedem Schritt verkleinert. Es wird erwartet, dass die absoluten Werte dieser Verbesserungen sich dabei vergrößern. Dabei stellt sich die Frage, ob das immer der Fall ist. Was geschieht, falls die Gewichte einiger Beobachtungen sich vergrößern? Wir untersuchen in diesem Beitrag, wie die absoluten Werte der Verbesserungen bei einer Beobachtung oder bei mehreren Beobachtungen variieren, nachdem ihre Gewichte geändert wurden. Schließlich weisen wir nach, dass die absoluten Werte der Verbesserungen sich nicht immer vergrößern, wenn ihre Gewichte verkleinert werden. Die Verkleinerung der Gewichte bei einigen zufällig ausgewählten Beobachtungen sollte die Möglichkeit geben, kleine Ausreißer zu entdecken.

1 Introduction

Robust methods, such as M-Estimations, Generalised M-Estimations, Least Median of Squares, L_1-Norm, etc., have been extensively used in geodesy in the last few decades (Kampmann 1989, Xu 1989, Caspary and Haen 1990, Carasio 1995, Koch 1996, etc.). Our concentration is here on M-Estimations. It is known that normal equations from M-Estimations are non-linear. Iteratively reweighted least squares method is usually used to solve the normal equations (Huber 1981, Hampel et al. 1986, Staudte and Sheather 1990, Dollinger and Staudte 1991). The new weights of some of the observations are obtained with respect to the absolute values of residuals from the preceding iteration. Huber (1981) calls them modified weights. The greater the absolute value of the residual of an observation from the preceding iteration, the smaller its new weight. At the final stage, bad observations (outliers) may receive small or even zero weights (Hoaglin et al. 1983).

It is also known that the partial redundancy r_i changes if the weight p_i of an observation is changed (Hekimoğlu 1998). For instance, if the weight p_i of the i^{th} observation o_i goes to zero, the related partial redundancy r_i goes to 1 (Chatterjee and Hadi 1988, p. 19), and the related residual v_i from LSE becomes large. Thus, the outlier can easily be identified by a robust method. Consequently, the relation between the weight of an observation and its residual from LSE is very important when using M-Estimation. Usually, it is expected that the absolute values of the residuals of some of the observations from LSE will be increased when decreasing the weights of these observations. Is this always the case? However, usually we do not exactly know how the residuals of some observations from LSE change when their weights are changed. In this paper, we try to answer this question in general.

In addition, identifying outliers by using M-Estimation becomes very difficult when the observations include multiple outliers with small magnitudes (Hekimoğlu and Koch 1999). To detect outliers with small magnitudes by robust methods reliably, the absolute values of the corresponding residuals must be increased. For this purpose, it is important to know how the residuals of some of the observations from LSE are increased when decreasing the weights of corresponding observations.

2 Changing the weight of one observation

A linear model, called Gauss-Markov model, is given by Koch (1999):

$$\bar{\mathbf{o}} + \bar{\mathbf{v}} = \bar{\mathbf{A}}\,\mathbf{x} \quad \text{with} \quad \mathbf{C}_{\bar{o}} = \sigma^2\,\bar{\mathrm{P}}^{-1}, \tag{1}$$

where $\bar{\mathbf{A}}$ is the $n \times u$ design matrix with rank u, $\bar{\mathbf{o}}$ the $n \times 1$ vector of observations and $\bar{\mathbf{v}}$ the error vector. $\mathbf{x}$ is the $u \times 1$ vector of unknowns, $C_{\bar{o}}$ the $n \times n$ a priori covariance matrix of the observations, $\bar{\mathbf{P}}$ the $n \times n$ diagonal weight matrix of the observations, σ^2 the a priori variance factor, n the number of observations, and u the number of unknowns ($u < n$). Suppose that the random errors are independent and normally distributed.

By transformation of the observation equations $\bar{\mathbf{o}}+\bar{\mathbf{v}}=\bar{\mathbf{A}}\ \mathbf{x}$ into $\mathbf{F}^T(\bar{\mathbf{o}}+\bar{\mathbf{v}})=\mathbf{F}^T\ \bar{\mathbf{A}}\ \mathbf{x}$,

$$\mathbf{o}+\mathbf{v}=\mathbf{A}\ \mathbf{x} \quad \text{with} \quad C_{\mathbf{o}}=\sigma^2\ I \tag{2}$$

is obtained, where

$$\mathbf{F}=\text{diag}\ \left(\sqrt{p_1},\ \sqrt{p_2},\ldots,\ \sqrt{p_n}\right). \tag{3}$$

This transformation is called homogenization in LSE.

First, a linear model given by (1) is homogenized as in (2) and a simple model is obtained. Let the matrix $\mathbf{H}$ be the hat matrix of n observations $\{\mathbf{H}$ is defined further in (8)$\}$. It is the transformation matrix that, when applied to $\mathbf{o}$, produces ($\hat{\mathbf{o}}=\mathbf{o}+\mathbf{v}$), i. e., $\hat{\mathbf{o}}=\mathbf{H}\ \mathbf{o}$ (Chatterjee and Hadi 1988, p. 9). Then, only the weights of m randomly chosen observations are changed by adding the different amounts of Δp where ($m\le(n-1)$). In this case, let $\mathbf{P}^c$, $\mathbf{H}^c$, $\mathbf{v}^c$ be the new weight matrix, the new hat matrix and the vector of residuals respectively.

After changing the weights of m observations, the new weight matrix $\mathbf{P}^c$ can be written as follows:

$$\mathbf{P}^c=\mathbf{P}+\mathbf{P}_1=\mathbf{I}+\mathbf{P}_1 \tag{4}$$

with

$$p_i^c=p_i+p_{1\,i}=1+p_{1\,i}\ ;\ i\in(1,2,3,\ldots,n). \tag{5}$$

For example, let us change the weights of the third, the sixth, and the ith observation. In this case, the matrix reads

$$\mathbf{P}_1=\text{diag}\left(0,0,\Delta p_3,0,0,\Delta p_6,0,0,0,\ldots,\Delta p_i,\ldots,0,0\right), \tag{6}$$

where

$$(\mathbf{P}_1)_{33}=\Delta p_3,\ (\mathbf{P}_1)_{66}=\Delta p_6,\ \ldots,\ (\mathbf{P}_1)_{ii}=\Delta p_i\ . \tag{7}$$

The hat matrices are given by

$$\mathbf{H}=\mathbf{A}(\mathbf{A}^T\mathbf{P}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{P}=\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \ \text{and} \tag{8}$$

$$\begin{aligned}\mathbf{H}^c&=\mathbf{A}(\mathbf{A}^T\mathbf{P}^c\mathbf{A})^{-1}\mathbf{A}^T\mathbf{P}^c\\&=\mathbf{A}(\mathbf{A}^T\mathbf{P}^c\mathbf{A})^{-1}\mathbf{A}^T\mathbf{P}+\mathbf{A}(\mathbf{A}^T\mathbf{P}^c\mathbf{A})^{-1}\mathbf{A}^T\mathbf{P}_1\ .\end{aligned} \tag{9}$$

As a result, we obtain

$$\mathbf{A}^T\mathbf{P}^c\mathbf{A}=\mathbf{A}^T\mathbf{P}\mathbf{A}+\mathbf{A}^T\mathbf{P}_1\mathbf{A}=\mathbf{A}^T\mathbf{A}+\mathbf{A}^T\mathbf{P}_1\mathbf{A}\ . \tag{10}$$

To derive a relation between $\mathbf{H}^c$ and $\mathbf{H}$, Gauss's matrix identity is used:

$$(\mathbf{D}+\mathbf{G}^T\mathbf{C})^{-1}=\mathbf{D}^{-1}-\mathbf{D}^{-1}\mathbf{G}^T(\mathbf{I}+\mathbf{C}\ \mathbf{D}^{-1}\mathbf{G}^T)^{-1}\mathbf{C}\ \mathbf{D}^{-1}, \tag{11}$$

where the dimension of the symmetrical and regular matrix $\mathbf{D}$ is $u\times u$, and the dimension of $\mathbf{G}$ and $\mathbf{C}$ is $q\times u$, with the same ranks of q (Staudte and Sheather 1990). If the identity (11) is applied on (10) by considering the equations

$$\mathbf{D}=\mathbf{A}^T\mathbf{A},\ \ \mathbf{G}^T=\mathbf{A}^T,\ \ \mathbf{C}=\mathbf{P}_1\mathbf{A}\ , \tag{12}$$

the following equation is obtained:

$$\begin{aligned}(\mathbf{A}^T\mathbf{P}^c\mathbf{A})^{-1}&=(\mathbf{A}^T\mathbf{A})^{-1}\\&-(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\ \{\mathbf{I}+\mathbf{P}_1\ \mathbf{A}\ (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\}^{-1}\mathbf{P}_1\mathbf{A}\ (\mathbf{A}^T\mathbf{A})^{-1}\ .\end{aligned} \tag{13}$$

A relation between $\mathbf{H}^c$ and $\mathbf{H}$ can be organized by substituting (13) into (9):

$$\mathbf{H}^c=\mathbf{H}+\mathbf{H}\ \mathbf{P}_1-\mathbf{H}\ (\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\mathbf{P}_1\mathbf{H}\ \mathbf{P}^c\ . \tag{14}$$

After changing the weights of m randomly chosen observations by adding the different amounts of Δp, the new residual vector $\mathbf{v}^c$ is obtained considering (14):

$$\mathbf{v}^c=(\mathbf{H}^c-\mathbf{I}\ )\ \mathbf{o}=\mathbf{v}+\Delta\mathbf{v}\ , \tag{15}$$

where

$$\Delta\mathbf{v}=\mathbf{H}\ \left\{\mathbf{P}_1-(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\mathbf{P}_1\mathbf{H}\ \mathbf{P}^c\right\}\ \mathbf{o} \tag{16}$$

and

$$\mathbf{v}=(\mathbf{H}-\mathbf{I})\ \mathbf{o}\ . \tag{17}$$

Theorem 1: If the weight (p_i) of an observation (o_i) is changed with an amount of Δp, the absolute values of all the new residuals from LSE are modified. If $\Delta p<0$, the absolute value of the corresponding residual v_i^c always becomes larger $\left(|\,v_i^c\,|>|\,v_i\,|\right)$, and if $\Delta p>0$, it always becomes smaller $\left(|\,v_i^c\,|<|\,v_i\,|\right)$.

Proof of theorem 1: If only the weight p_i of the ith observation is changed by adding the amount of Δp, i. e., $p_i^c=p_i+\Delta p=1+\Delta p$, the residual v_i changes to $v_i^c=v_i+\Delta v_i$. Thus, Δv_i is obtained from (16). Considering the following property of the matrix $\mathbf{P}_1$:

$$\begin{aligned}&(\mathbf{P}_1)_{ii}=\Delta\text{p},\\&(\mathbf{P}_1)_{ij}=0,\ \ j\ne i,\ \ i=1,2,\ldots,n,\ \ j=1,2,\ldots,n\ ,\end{aligned} \tag{18}$$

we can write

$$\begin{aligned}&\mathrm{Row}_i(\mathbf{P}_1\mathbf{H}\,\mathbf{P}^c)\\&=[\Delta p\,h_{i1},\ \Delta p\,h_{i2},\ldots,\Delta p\,h_{ii}(1+\Delta p),\ldots,\ \Delta p\,h_{in}]\,,\end{aligned} \tag{19a}$$

$$(\mathbf{P}_1\mathbf{H}\mathbf{P}^c)_{kj}=0,\quad k\neq i,\quad k=1,2,\ldots,n,\quad j=1,2,\ldots,n\,, \tag{19b}$$

and

$$\begin{aligned}&\left\{\mathrm{Row}_i\left\{(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\right\}\right.\\&=\left[\frac{-\Delta p\,h_{i1}}{1+\Delta p\,h_{ii}},\ \frac{-\Delta p\,h_{i2}}{1+\Delta p\,h_{ii}},\ldots,\frac{1}{1+\Delta p\,h_{ii}},\ldots,\frac{-\Delta p\,h_{in}}{1+\Delta p\,h_{ii}}\right];\end{aligned} \tag{20a}$$

$$\left\{(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\right\}_{kj}=0\,,\ k\neq i\ \text{ and }\ k\neq j\,, \tag{20b}$$

$$\left\{(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\right\}_{kj}=1\,,\ k\neq i\ \text{ and }\ k=j\,. \tag{20c}$$

Also

$$\begin{aligned}&\mathrm{Row}_i\left\{\mathbf{P}_1-(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\mathbf{P}_1\ \mathbf{H}\mathbf{P}^c\right\}\\&=\left[\frac{-\Delta p\,h_{i1}}{1+\ \Delta p h_{ii}},\ \frac{-\Delta p\,h_{i2}}{1+\Delta p\,h_{ii}},\ldots,\Delta p-\frac{\Delta p\,h_{ii}(1+\Delta p)}{1+\Delta p\,h_{ii}},\ldots,\frac{-\Delta p\,h_{in}}{1+\Delta p\,h_{ii}}\right]\end{aligned} \tag{21a}$$

and

$$\left\{\mathbf{P}_1-(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\mathbf{P}_1\mathbf{H}\mathbf{P}^c\right\}_{kj}=0\,,\ k\neq i\,. \tag{21b}$$

If we substitute the equation (21a) and (21b) in (16), Δv_i and v_i^c can be found as:

$$\Delta v_i=\frac{-\Delta p h_{ii}}{1+\Delta p h_{ii}}v_i \tag{22}$$

and

$$v_i^c=\frac{v_i}{1+\Delta p h_{ii}}\,. \tag{23}$$

If $\Delta p<0$, then the sign of Δv_i is the same as the sign of v_i. Therefore, the following inequality can be written:

$$|\,v_i^c\,|>|\,v_i\,|,\quad \text{if}\quad -1<\Delta p<0\,, \tag{24a}$$

$$|\,v_i^c\,|<|\,v_i\,|\quad \text{if}\quad \Delta p>0\,. \tag{24b}$$

Thus, theorem 1 has been proved.

To see how the residual of the j^{th} observation changes when changing the weight of the i^{th} observation, the following equations can be obtained from (16) similar to (22) and (23):

$$\Delta v_j=\frac{-\Delta p\,h_{ji}}{1+\Delta p\,h_{ii}}v_i \tag{25a}$$

and

$$v_j^c=v_j-\frac{\Delta p\,h_{ji}v_i}{1+\Delta p\,h_{ii}}\,. \tag{25b}$$

The residual v_j^c is changed depending on the signs of v_i, v_j, h_{ji} and Δp.

3 Changing the weights of some of the observations

Theorem 2: If the weights of m randomly chosen observations $\{$i.e., $2<m\leq(n-1)\}$ are decreased in different amounts separately (for example, $m=3$ and $p_i^c=p_i+\Delta p_i$, $p_j^c=p_j+\Delta p_j$, $p_k^c=p_k+\Delta p_k$), the absolute values of the new corresponding residuals from LSE become larger with certain probability, and if the weights are increased separately, they become smaller with the same certain probability. The probabilities are 87.5 % for $m=2$, 81.25 % for $m=3$, 78.125 % for $m=4$, etc. The greater the number m, the smaller the probability.

Proof of theorem 2: Let $m=2$, i.e., only the weights of the i^{th} and the j^{th} observations are changed by adding Δp_i and Δp_j respectively. In this case, Δv_i and Δv_j can be calculated again from (16). Taking into account

$$\begin{aligned}&(\mathbf{P}_1)_{ii}=\Delta p_i\,,\ \ (\mathbf{P}_1)_{jj}=\Delta p_j\,,\\&(\mathbf{P}_1)_{ij}=0,\quad j\neq i,\quad i=1,2,\ldots,n,\quad j=1,2,\ldots,n\,,\end{aligned} \tag{26}$$

we can write

$$\begin{aligned}&\mathrm{Row}_i(\mathbf{P}_1\mathbf{H}\mathbf{P}^c)\\&=\Delta p_i\,[h_{i1},\,h_{i2},\ldots,(1+\Delta p_i)h_{ii},\ldots,\,(1+\Delta p_i)h_{ij}\,,\ldots,\,h_{in}]\,,\\&\mathrm{Row}_j(\mathbf{P}_1\mathbf{H}\mathbf{P}^c)\\&=\Delta p_j[h_{j1},\,h_{j2},\ldots,\,(1+\Delta p_j)h_{ji},\ldots,\,(1+\Delta p_j)h_{jj}\,,\ldots,\,h_{jn}]\,,\end{aligned} \tag{27a}$$

$$(\mathbf{P}_1\mathbf{H}\mathbf{P}^c)_{kl}=0,\ \ k\neq i\,,\ k\neq j\,,\ k=1,2,\ldots,n\,,\ l=1,2,\ldots,n\,, \tag{27b}$$

and

$$\mathrm{Row}_i\left\{(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\right\}=\left[\omega_1,\ \omega_2,\ldots,\omega_i,\ldots,\omega_j,\ldots,\omega_n\right], \quad (28a)$$

$$\mathrm{Row}_j\left\{(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\right\}=\left[\eta_1,\ \eta_2,\ldots,\eta_i,\ldots,\eta_j,\ldots,\eta_n\right], \quad (28b)$$

$$\left\{(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\right\}_{kl}=0,\ k\neq i,\ k\neq j \text{ and } k\neq l, \quad (28c)$$

$$\left\{(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\right\}_{kl}=1,\ k\neq i,\ k\neq j \text{ and } k=l. \quad (28d)$$

Also

$$\begin{aligned}&\mathrm{Row}_i\left\{\mathbf{P}_1-(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\mathbf{P}_1\mathbf{H}\mathbf{P}^c\right\}\\&=[-\Delta p_i\ h_{i1}\omega_i-\Delta p_j\ h_{j1}\omega_j,\ldots,\Delta p_i-\Delta p_i(1+\Delta p_i)h_{ii}\omega_i\\&\quad-\Delta p_j(1+\Delta p_i)h_{ji}\omega_j,\ldots,-\Delta p_i(1+\Delta p_j)h_{ij}\omega_i\\&\quad-\Delta p_j(1+\Delta p_j)h_{jj}\omega_j,\ldots,-\Delta p_i h_{in}\omega_i-\Delta p_j h_{jn}\omega_j], \quad (29a)\end{aligned}$$

$$\begin{aligned}&\mathrm{Row}_j\left\{\mathbf{P}_1-(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\mathbf{P}_1\mathbf{H}\mathbf{P}^c\right\}\\&=[-\Delta p_i\ h_{i1}\eta_i-\Delta p_j\ h_{j1}\eta_j,\ldots,-\Delta p_i(1+\Delta p_i)h_{ii}\eta_i\\&\quad-\Delta p_j(1+\Delta p_i)h_{ji}\eta_j,\ldots,\Delta p_j-\Delta p_i(1+\Delta p_j)h_{ij}\eta_i\\&\quad-\Delta p_j(1+\Delta p_j)h_{jj}\eta_j,\ldots,-\Delta p_i h_{in}\eta_i-\Delta p_j h_{jn}\eta_j] \quad (29b)\end{aligned}$$

$$\{\mathbf{P}_1-(\mathbf{I}+\mathbf{P}_1\mathbf{H})^{-1}\mathbf{P}_1\mathbf{H}\mathbf{P}^c\}_{kl}=0,\ k\neq i,\ k\neq j \quad (29c)$$

where

$$\omega_i=\frac{1+\Delta p_j h_{jj}}{D}\quad,\quad \omega_j=-\frac{\Delta p_i h_{ij}}{D}, \quad (30a)$$

$$\eta_i=-\frac{\Delta p_j h_{ji}}{D},\quad \eta_j=\frac{1+\Delta p_i h_{ii}}{D},\text{ and} \quad (30b)$$

$$D=(1+\Delta p_i h_{ii})\ (1+\Delta p_j h_{jj})-\Delta p_i \Delta p_j h_{ji} h_{ij}\ . \quad (30c)$$

Substituting equations (29a, b and c), and (30a, b and c) in (16), Δv_i and Δv_j can be rewritten as

$$\Delta v_i=-\frac{\Delta p_i}{D}\left\{\ h_{ii}(1+\Delta p_j h_{jj})-\Delta p_j h_{ij} h_{ji}\right\}\ v_i-\frac{\Delta p_j h_{ij}}{D}v_j \quad (31)$$

and

$$\Delta v_j=-\frac{\Delta p_j}{D}\left\{\ h_{jj}(1+\Delta p_i h_{ii})-\Delta p_i h_{ji} h_{ij}\right\}\ v_j-\frac{\Delta p_i h_{ji}}{D}v_i\ . \quad (32)$$

In addition, any k^{th} residual can be expressed as

$$\begin{aligned}\Delta v_k=&-\frac{\Delta p_i}{D}\left\{h_{ki}(1+\Delta p_j h_{jj})-\Delta p_j h_{ji} h_{kj}\right\}v_i\\&-\frac{\Delta p_j}{D}\left\{h_{kj}(1+\Delta p_i h_{ii})-\Delta p_i h_{ij} h_{ki}\right\}v_j\ , \quad (33)\end{aligned}$$

where $k\neq i$, $k\neq j$, $k=1,\ 2,\ldots,\ n$.

Considering $h_{ii}h_{jj}-h_{ij}^2\geq 0$ (Chatterjee and Hadi 1988, p. 19) and $h_{ij}=h_{ji}$ ($\mathbf{H}$ is symmetric because $\mathbf{P}=\mathbf{I}$), we can write

$$h_{ii}(1+\Delta p_j h_{jj})-\Delta p_j h_{ij} h_{ji}>0 \quad (34)$$

and

$$h_{jj}(1+\Delta p_i h_{ii})-\Delta p_i h_{ji} h_{ij}>0\ . \quad (35)$$

We now consider increasing or decreasing the absolute values of v_i^c and v_j^c altogether. First we take into consideration the increase of the absolute value of v_i^c. Taking into account (34), the absolute value of the residual Δv_i increases or decreases depending on the signs and the magnitudes of v_i, v_j, h_{ij}, Δp_i, and Δp_j as seen in (31). If the sign of Δp_i is different from the sign of Δp_j, the first and second terms of (31) cancel. If the signs of Δp_i and Δp_j are negative, the absolute value of the residual Δv_i changes depending on the signs and the magnitudes of v_i, v_j, and h_{ij}. In this case, the absolute value of v_i^c ($v_i^c=v_i+\Delta v_i$) is increased mostly because often the first term in (31) is dominant. If the signs of Δp_i and Δp_j are positive, the absolute value of v_i^c decreases mostly. These considerations are also valid for increasing the absolute values of v_j^c.

Let the weights of the i^{th} and the j^{th} observations be decreased (i.e., $-1<\Delta p_i<0$ and $-1<\Delta p_j<0$). The absolute value of the residual v_i^c increases or decreases depending on the signs and the magnitudes of v_i, v_j, h_{ij} and Δv_i. Both v_i and v_j are elementary random events. Δv_i and Δv_j are also random events combined with v_i and v_j. To investigate the case of increasing the absolute value of v_i^c, all-possible 12 random events combined with the random events (v_i, v_j, h_{ij} and Δv_i) are presented in Tab. 1. The probability of each residual being positive or negative is 0.5 assuming random errors come from a normal distribution. But, the probability of h_{ij} (being positive or negative) may be computed from the hat matrix $\mathbf{H}$. Let the number of the positive h_{ij} be a. Then the number of all-possible h_{ij} is C_n^2. Consequently, the number of the negative h_{ij} must be C_n^2-a. Thus, the probability of $h_{ij}>0$ is a/C_n^2. For example, for a simple regression ($n=10$, $u=2$) $a=37$ and $C_n^2=45$. For a multiple regression ($n=13$, $u=5$) $a=55$ and $C_n^2=78$. From these experiments, it is accepted that the probability of $h_{ij}>0$ is greater than the probability of $h_{ij}<0$. But, the probability of Δv_i (being either positive or negative) depends on the signs and the magnitudes of v_i, v_j and h_{ij}.

The 12 random events given in Tab. 1 may be divided into two groups in view of the probability of Δv_i being positive or negative. In the first group, there are the random events of Δv_i with probabilities approximately equal to 1. They are given in the rows 1, 2, 3 and 4 in Tab. 1. In the second group, the random events of Δv_i with probabilities approximately equal to 0.5 are presented. They are given in Tab. 1, in the rows 5, 6, 7, 8, 9, 10, 11, and 12.

Tab. 1: All-possible random events of $|v_i^c|$

Groups	Nr	v_i	v_j	h_{ij}	Δv_i	Comparing $\|v_i^c\|$ with $\|v_i\|$
I	1	+	+	+	+	$\|v_i^c\|>\|v_i\|$
	2	+	–	–	+	$\|v_i^c\|>\|v_i\|$
	3	–	+	–	–	$\|v_i^c\|>\|v_i\|$
	4	–	–	+	–	$\|v_i^c\|>\|v_i\|$
II	5	+	+	–	+	$\|v_i^c\|>\|v_i\|$
	6	+	+	–	–	$\|v_i^c\|>\|v_i\|$ or $\|v_i^c\|<\|v_i\|$
	7	+	–	+	+	$\|v_i^c\|>\|v_i\|$
	8	+	–	+	–	$\|v_i^c\|>\|v_i\|$ or $\|v_i^c\|<\|v_i\|$
	9	–	+	+	+	$\|v_i^c\|>\|v_i\|$ or $\|v_i^c\|<\|v_i\|$
	10	–	+	+	–	$\|v_i^c\|>\|v_i\|$
	11	–	–	–	+	$\|v_i^c\|>\|v_i\|$ or $\|v_i^c\|<\|v_i\|$
	12	–	–	–	–	$\|v_i^c\|>\|v_i\|$

The probability of $|v_i^c|>|v_i|$ can be computed from Tab. 1 as follows:

$$\begin{aligned}
&\mathrm{P}\left(|v_i^c|>|v_i|\right)\cong 0.5\times0.5\times\left(a/C_n^2\right)\times1\\
&+0.5\times0.5\times\left[\left(C_n^2-a\right)/C_n^2\right]\times1\\
&+0.5\times0.5\times\left[\left(C_n^2-a\right)/C_n^2\right]\times1\\
&+0.5\times0.5\times\left(a/C_n^2\right)\times1\\
&+0.5\times0.5\times\left[\left(C_n^2-a\right)/C_n^2\right]\times0.5\\
&+0.5\times0.5\times\left[\left(C_n^2-a\right)/C_n^2\right]\times0.5\times0.5\\
&+0.5\times0.5\times\left(a/C_n^2\right)\times0.5\\
&+0.5\times0.5\times\left(a/C_n^2\right)\times0.5\times0.5\\
&+0.5\times0.5\times\left(a/C_n^2\right)\times0.5\times0.5\\
&+0.5\times0.5\times\left(a/C_n^2\right)\times0.5\\
&+0.5\times0.5\times\left[\left(C_n^2-a\right)/C_n^2\right]\times0.5\times0.5\\
&+0.5\times0.5\times\left[\left(C_n^2-a\right)/C_n^2\right]\times0.5\cong0.875\,.
\end{aligned}$$

Also $P\left(|v_j^c|>|v_j|\right)\cong0.875$. Consequently, we can write $P\left(|v_i^c|<|v_i|\right)=P\left(|v_j^c|<|v_j|\right)\cong1-0.875\cong0.125$.

All-possible random events combined with both v_i^c and v_j^c residuals can be given as:

$$\left\{v_i^c(+)v_j^c(+),\quad v_i^c(+)v_j^c(-),\quad v_i^c(-)v_j^c(+),\quad v_i^c(-)v_j^c(-)\right\}$$

where »+« denotes the increase in the absolute value of the relevant residual whereas »–« indicates the decrease in the absolute value of the relevant residual. The probability of the first event is $0.875\times0.875\cong0.766$, the second one is $0.875\times0.125\cong0.109$, and finally the third one is $0.125\times0.875\cong0.109$. The event $v_i^c(-)v_j^c(-)$ may arise seldom because of its very small probability, i. e., $0.125\times0.125\cong0.016$.

This proof is also valid when $\Delta p_i>0$ and $\Delta p_j>0$. It means that $P\left(|v_i^c|<|v_i|\right)=P\left(|v_j^c|<|v_j|\right)\cong0.875$.

Proof of theorem 2 for $m=3$: Let only the weights of the ith, the jth and the kth observations be changed by adding Δp_i, Δp_j and Δp_k (i. e., $p_i^c=p_i+\Delta p_i$, $p_j^c=p_j+\Delta p_j$, $p_k^c=p_k+\Delta p_k$) respectively. In this case, Δv_i, Δv_j and Δv_k can be obtained again from (16) as done for $m=2$. As a result, Δv_i, Δv_j and Δv_k can be rewritten as follows:

$$\begin{aligned}
\Delta v_i=-\frac{\Delta p_i}{D}&\left[h_{ii}\{(1+\Delta p_j h_{jj})(1+\Delta p_k h_{kk})-\Delta p_j\Delta p_k h_{jk}h_{kj}\}\right.\\
&+h_{ij}(\Delta p_j\Delta p_k h_{jk}h_{ki}-\Delta p_j h_{ji}-\Delta p_j\Delta p_k h_{ji}h_{kk})\\
&\left.+h_{ik}(\Delta p_j\Delta p_k h_{ji}h_{kj}-\Delta p_k h_{ki}-\Delta p_j\Delta p_k h_{ki}h_{jj})\right]v_i\\
&-\frac{\Delta p_j}{D}\left(h_{ij}+\Delta p_k h_{ij}h_{kk}-\Delta p_k h_{ik}h_{kj}\right)v_j\\
&-\frac{\Delta p_k}{D}\left(h_{ik}+\Delta p_j h_{ik}h_{jj}-\Delta p_j h_{ij}h_{jk}\right)v_k\,,
\end{aligned}\tag{36}$$

$$\begin{aligned}
\Delta v_j=-\frac{\Delta p_j}{D}&\left[h_{jj}\{(1+\Delta p_i h_{ii})(1+\Delta p_k h_{kk})-\Delta p_i\Delta p_k h_{ik}h_{ki}\}\right.\\
&+h_{ji}(\Delta p_i\Delta p_k h_{ik}h_{kj}-\Delta p_i h_{ij}-\Delta p_i\Delta p_k h_{ij}h_{kk})\\
&\left.+h_{jk}(\Delta p_i\Delta p_k h_{ij}h_{ki}-\Delta p_k h_{kj}-\Delta p_i\Delta p_k h_{kj}h_{ii})\right]v_j\\
&-\frac{\Delta p_i}{D}\left(h_{ji}+\Delta p_k h_{ji}h_{kk}-\Delta p_k h_{jk}h_{ki}\right)v_i\\
&-\frac{\Delta p_k}{D}\left(h_{jk}+\Delta p_i h_{jk}h_{ii}-\Delta p_i h_{ji}h_{ik}\right)v_k\,,
\end{aligned}\tag{37}$$

and

$$\begin{aligned}
\Delta v_k=-\frac{\Delta p_k}{D}&\left[h_{kk}\{(1+\Delta p_j h_{jj})(1+\Delta p_i h_{ii})-\Delta p_i\Delta p_j h_{ji}h_{ij}\}\right.\\
&+h_{kj}(\Delta p_i\Delta p_j h_{ji}h_{ik}-\Delta p_j h_{jk}-\Delta p_i\Delta p_j h_{jk}h_{ii})\\
&\left.+h_{ki}(\Delta p_i\Delta p_j h_{jk}h_{ij}-\Delta p_i h_{ik}-\Delta p_i\Delta p_j h_{ik}h_{jj})\right]v_k\\
&-\frac{\Delta p_j}{D}\left(h_{kj}+\Delta p_i h_{kj}h_{ii}-\Delta p_i h_{ki}h_{ij}\right)v_j\\
&-\frac{\Delta p_i}{D}\left(h_{ki}+\Delta p_j h_{ki}h_{jj}-\Delta p_j h_{kj}h_{ji}\right)v_i\,,
\end{aligned}\tag{38}$$

where

$$D=(1+\Delta p_i h_{ii})\{(1+\Delta p_j h_{jj})(1+\Delta p_k h_{kk})-\Delta p_j \Delta p_k h_{jk} h_{kj}\}$$
$$+\Delta p_i h_{ij}\{\Delta p_j \Delta p_k h_{jk} h_{ki}-\Delta p_j h_{ji}(1+\Delta p_k h_{kk})\}$$
$$+\Delta p_i h_{ik}\{\Delta p_j \Delta p_k h_{ji} h_{kj}-\Delta p_k h_{ki}(1+\Delta p_j h_{jj})\}\,. \qquad (39)$$

We have numerically investigated that the coefficient of v_i in (36), the coefficient of v_j in (37) and the coefficient of v_k in (38) are always positive. Now, we consider only Δv_i. If $-1<\Delta p_i<0$, $-1<\Delta p_j<0$ and $-1<\Delta p_k<0$, Δv_i changes depending on v_i, B, v_j, C, and v_k since **A** is always positive:

$$\Delta v_i=-\frac{\Delta p_i}{D}A\,v_i-\frac{\Delta p_j}{D}B\,v_j-\frac{\Delta p_k}{D}C\,v_k\,, \qquad (40)$$

where

$$A=h_{ii}\{(1+\Delta p_j h_{jj})(1+\Delta p_k h_{kk})-\Delta p_j \Delta p_k h_{jk} h_{kj}\}$$
$$+h_{ij}(\Delta p_j \Delta p_k h_{jk} h_{ki}-\Delta p_j h_{ji}-\Delta p_j \Delta p_k h_{ji} h_{kk})$$
$$+h_{ik}(\Delta p_j \Delta p_k h_{ji} h_{kj}-\Delta p_k h_{ki}-\Delta p_j \Delta p_k h_{ki} h_{jj}),$$
$$B=h_{ij}+\Delta p_k h_{ij} h_{kk}-\Delta p_k h_{ik} h_{kj}\,,\ \text{and} \qquad (41)$$
$$C=h_{ik}+\Delta p_j h_{ik} h_{jj}-\Delta p_j h_{ij} h_{jk}\,.$$

The probability of Δv_i (being either positive or negative) depends on the signs and the magnitudes of v_i, v_j, v_k and the related elements of the hat matrix as seen in (40) and (41). It can approximately be taken 0.5 or 1 as done for the case of $m=2$. Let the number of $B<0$ be a and $C<0$ be b. Also the number of all the possibilities is C_n^3, the number of positive B is C_n^3-a, and the number of positive C is C_n^3-b. In this case, 56 different independent random events arise. For instance,

$$\{v_i(+)\ B(+)\ v_j(+)\ C(+)\ v_k(+)\ \Delta v_i(+),$$
$$v_i(-)\ B(-)\ v_j(-)\ C(-)\ v_k(+)\ \Delta v_i(-),\ \ldots\}\,.$$

The probability of $|v_i^c|>|v_i|$ can be computed with the variable $g{=}C_n^3$ as follows:

$$P\left(|v_i^c|>|v_i|\right)\cong$$
$$\{0.5\times(a/g)\times 0.5\times(b/g)\times 0.5\times 1\}\times 2$$
$$+\,[0.5\times\{(g{-}a)/g\}\times 0.5\times(b/g)\times 0.5\times 1]\times 2$$
$$+\,[0.5\times(a/g)\times 0.5\times\{(g{-}b)/g\}\times 0.5\times 1]\times 2$$
$$+\,[0.5\times\{(g{-}a)/g\}\times 0.5\times\{(g{-}b)/g\}\times 0.5\times 1]\times 2$$
$$+\,\{0.5\times(a/g)\times 0.5\times(b/g)\times 0.5\times 0.5\}\times 6$$
$$+\,[0.5\times(a/g)\times 0.5\times\{(g{-}b)/g\}\times 0.5\times 0.5]\times 6$$
$$+\,[0.5\times\{(g{-}a)/g\}\times 0.5\times(b/g)\times 0.5\times 0.5]\times 6$$
$$+\,[0.5\times\{(g{-}a)/g\}\times 0.5\times\{(g{-}b)/g\}\times 0.5\times 0.5]\times 6$$
$$+\,\{0.5\times(a/g)\times 0.5\times(b/g)\times 0.5\times 0.5\times 0.5\}\times 6$$
$$+\,[0.5\times(a/g)\times 0.5\times\{(g{-}b)/g\}\times 0.5\times 0.5\times 0.5]\times 6$$
$$+\,[0.5\times\{(g{-}a)/g\}\times 0.5\times(b/g)\times 0.5\times 0.5\times 0.5]\times 6$$
$$+\,[0.5\times\{(g{-}a)/g\}\times 0.5\times\{(g{-}b)/g\}\times 0.5\times 0.5\times 0.5]\times 6$$
$$\cong 0.8125.$$

Also $P(|v_j^c|>|v_j|)=P(|v_k^c|>|v_k|)\cong 0.8125$.

This means that the absolute value of the residual v_i^c (or v_j^c, or v_k^c) increases with a probability of 81.25 % if the corresponding weights p_i^c, p_j^c and p_k^c are decreased to the different amounts of Δp. Consequently, $P(|v_i^c|<|v_i|)=P(|v_j^c|<|v_j|)=P(v_k^c|<|v_k|)$ $=1-0.8125\cong 0.1875$.

This proof is also valid if $\Delta p_i>0$, $\Delta p_j>0$ and $\Delta p_k>0$.

Proof of theorem 2 for $m=4$ has been derived, but it is not given here because it is very long. The probability of $|v_i^c|>|v_i|$ is obtained as 78.125 %. The proof of the theorem 2 for $m=5,\ldots,n-1$ has not been derived. But looking at Tab. 2, one observes a decrease in the magnitude of the probabilities with a rate of 50 % as m grows higher. Assuming this is valid for all cases of m, the probabilities for $m=5,\ldots,n-1$ may be predicted in a sequential form as follows:
$P(m=5)\cong 76.5625\,\%$, $P(m=6)\cong 75.78125\,\%$,
$P(m=7)\cong 75.390625\,\%$, $P(m=8)\cong 75.1953125\%$, etc.

The computed and the predicted probabilities mentioned above are called the theoretical probabilities.

Tab. 2: Theoretical probabilities of $|v_i^c|>|v_i|$ for different number of m

m	probability of $\lvert v_i^c\rvert>\lvert v_i\rvert$	difference of the sequential probabilities
1	100 %	
2	87.50 %	12.50 %
3	81.25 %	6.25 %
4	78.125 %	3.125 %

To verify the theoretical probabilities, simulations have been performed. The simulations take into consideration all possible situations where a simple straight line and two multiple linear regression models (i. e. u = 5 and u = 10 respectively) are used.

One thousand sets for the observations were generated by using normally distributed random errors. This data is used to change the weights of m randomly chosen observations. First, the absolute values of the corresponding residuals are obtained from LSE. Then, the absolute values of the new residuals of m randomly chosen observations, whose weights are reduced, are obtained. The absolute values of both types of residuals are computed 1000 times. The absolute values of both types of residuals are compared with each other. The cases where the absolute value of the residual of an observation (whose weight is reduced) is greater than the residual of the observation (the weight of which is not changed) are counted for each one of m observations. Then, the mean value is computed by dividing the counted number to $1000 \times m$. It is called the estimated mean probability. It changes depending on the number m, the number of observations, the number of unknowns, and the different values of Δp.

As a result, the estimated mean probabilities approach to the theoretical probabilities sufficiently. Mostly, the greater the number of observations and unknowns, the closer the estimated mean probabilities represent the theoretical ones.

4 Conclusion

We have proved that if the weight of an observation is decreased, the absolute value of the new corresponding residual from LSE always becomes larger; and if the weight is increased, the absolute value of the residual always becomes smaller. In general, if the weights of m randomly chosen observations $\{1 < m \leq (n-1)\}$ are decreased altogether to the randomly chosen amounts of Δp or to the same amount of Δp, the absolute values of the new corresponding residuals from LSE become larger with a particular probability. In other words, they do not increase always. If the weights of m randomly chosen observations are increased, the absolute values of the new corresponding residuals will be smaller with the same amount of probability.

In this study, we do not attempt to prove the theorems for $m > 4$ since the proof gets too complicated. However, the probabilities are obtained up to the case of $m = 4$, such as 100 % for $m = 1$, 87.5 % for $m = 2$, 81.25 % for $m = 3$, 78.125 % for $m = 4$, etc. Based on our results, we make the assumption that the probabilities could be predicted for $m > 4$ in such that they should be decreased when the number m grows higher. These assumptions have been confirmed by performing simulations. The greater the number of unknowns and observations, the closer the estimated mean probabilities to the theoretical ones.

One major outcome of this study is that »the ability of robust methods could be increased for detecting small outliers«. Robust methods alone may fail to identify some outliers having small magnitudes. However using the approach outlined here, if the randomly chosen set included a bad observation (or more), the absolute value of its residual would be increased. Thus, we would expect that robust methods would have much success for detecting them.

Acknowledgements
I would like to thank Prof. K. Rudolf Koch for his critics. I would also like to thank Research Assistants Mustafa Berber and Cuneyt Aydin for proof reading the manuscript.

References

Carosio, A.: Ausgleichung geodätischer Netze mit Verfahren der robusten Statistik. Verm., Photog., Kulturtech. 93: 188–191, 1995.

Caspary, W., W. Haen: Simultaneous estimation of location and scale parameters in the context of robust M-estimation. Manuscripta Geodaetica 15: 273–282, 1990.

Chatterjee, S., A. S. Hadi: Sensitivity analysis in linear regression. John Wiley & Sons, Inc., New York, 1988.

Dollinger, M., R. G. Staudte: Influence functions of iteratively reweighted least squares estimators. J. Am. Stat. Assoc. 86: 709–716, 1991.

Hampel, F., E. Ronchetti, P. Rousseeuw, W. Stahel: Robust statistics: the approach based on influence functions. John Wiley & Sons, New York, N. Y., 1986.

Hekimoğlu, Ş.: Change of the diagonal elements of the hat matrix under changing weight and changing position of an observation. Z. f. Verm.wesen 123: 266–271, 1998.

Hekimoğlu, Ş., K. R. Koch: How can reliability of the robust methods be measured? Third Turkish-German Joint Geodetic Days, 1–4 June 1999, Istanbul, Vol. 1: 179–196, 1999.

Hoaglin, D. C., F. Mosteller, S. W. Tukey: Understanding robust and exploratory data analysis. John Wiley & Sons, Inc., New York, 1983.

Huber, P. J.: Robust statistics. John Wiley & Sons, Inc., New York, 1981.

Kampmann, G.: Zur Ausgleichung freier Netze mit der L1-Norm-Methode. Allg. Verm. Nachr. 96: 110–118, 1989.

Koch, K. R.: Robuste Parameterschätzung. Allg. Verm. Nachr. 103: 1–18, 1996.

Koch, K. R.: Parameter estimation and hypothesis testing in linear models. 2nd Ed. Springer-Verlag, Berlin, Heidelberg, New York, 1999.

Staudte, R. G., S. J. Sheather: Robust estimation and testing. A Wiley-Interscience Publication, John Wiley & Sons, Inc, New York, 1990.

Xu, P.: Statistical criteria for robust methods. ITC Journal 1: 37–40, 1989.

Authors' address:
Prof. Dr. Şerif Hekimoğlu
Yildiz Teknik Üniversitesi Insaat Fakültesi
Jeodezi ve Fotogrametri Müh. Böl.
Yildiz - Istanbul, 34349 Türkiye
Fax No: +90-212-2610767
hekim@yildiz.edu.tr

Dr. Uğur Şanlı
Yildiz Teknik Üniversitesi Insaat Fakültesi
Jeodezi ve Fotogrametri Müh. Böl.
Yildiz - Istanbul, 34349 Türkiye
Fax No: +90-212-2610767
usanli@yildiz.edu.tr