Variation characteristics of MDBs in robust estimation

J. Guo, J. Ou

The least-squares (LS) method is very susceptive to outliers. The effects of any perturbation of the variance components and weight elements on the LS method are analyzed and tested. A local sensitivity-based robust estimation is developed. Theoretical analyses and numerical results indicate that the MDB (minimal detectable bias) measures become larger and larger with progress in the iterative process, thus confirming the robustness of the proposed method.

1 Introduction

The least-squares (LS) method is very sensitive to outliers [1, 6, 7]. Outliers in observations invalidate the outputs of LS adjustment and any conclusion based on them. Therefore, it is important to have ways to control the influence of outliers.

One way is to apply outlier detection test procedures [1, 9, 12]. However, the conventional outlier detection test procedures suffer from a common weakness in that their abilities to correctly localize gross errors are seriously limited in the case of more than one gross error [4, 11]. In this situation, it is desirable to use robust estimation methods to get reliable estimates [14, 15].

In the robust estimation literature, the influence of outliers is addressed under the assumption of a variance-inflation model [5, 6]. As pointed out by HAMPEL et al. [5], the robustness theories are the sensitivity analysis and perturbation theory of statistical procedures. A key role in LS adjustment is played by the weighted sum of the squares of the LS residuals [8]. Therefore, potentially outlying observations can be identified reasonably well based on the sensitivity of this value to any perturbation of the variance components. The minimal detectable biases (MDBs), as introduced by Baarda [1], are an important diagnostic tool for inferring the strength of the model validation. However, the variation characteristics of the MDBs have received little attention in the robust estimation literature.

The paper is organized as follows. Firstly, the stability of the LS method is investigated in detail. A local sensitivity-based robust estimation is developed. The variation characteristics of the MDBs during the iterative process are addressed to evaluate the robustness of the proposed method. Finally, a numerical example is used to verify the method.

2 Model description

Consider the Gauss-Markov model [7]

$$\boldsymbol{AX} = E(\boldsymbol{L}) \quad \text{with} \quad D(\boldsymbol{L}) = \sigma_0^2 \boldsymbol{P}^{-1} \tag{1}$$

where $\boldsymbol{A}$ is the $n \times u$ design matrix with full column rank, $\boldsymbol{X}$ the $u \times 1$ vector of unknowns and $\boldsymbol{L}$ the $n \times 1$ vector of observations, $D(\boldsymbol{L})$ is the covariance matrix of $\boldsymbol{L}$, $\boldsymbol{P}$ the associated weight matrix, and σ_0^2 the a-priori variance factor of the unit weight. Whenever necessary, the observations are assumed to be normally distributed.

Under the assumption of uncorrelated observations, the covariance matrix $D(\boldsymbol{L})$ is a diagonal matrix

$$D(\boldsymbol{L}) = diag(\sigma_{ii}), \ \sigma_{ii} > 0 \tag{2}$$

The weight matrix $\boldsymbol{P}$ then follows with

$$\boldsymbol{P} = diag(p_i), \quad p_i > 0 \tag{3}$$

where

$$p_i = \sigma_0^2/\sigma_{ii}, \quad i = 1, 2, \therefore, n. \tag{4}$$

The weighted LS estimator of the unknowns in Eq. (1) is [7]

$$\hat{\boldsymbol{X}} = (\boldsymbol{A}^T\boldsymbol{PA})^{-1} \boldsymbol{A}^T\boldsymbol{PL} \tag{5}$$

The corresponding (weighted) residual vector is obtained as

$$\boldsymbol{V} = \boldsymbol{A}\hat{\boldsymbol{X}} - \boldsymbol{L} = -\boldsymbol{RL} \tag{6}$$

with $\boldsymbol{R} = \boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^T\boldsymbol{PA})^{-1}\boldsymbol{A}^T\boldsymbol{P}$. One can easily verify that the matrix $\boldsymbol{R}$, which contains extremely useful information [4, 6], is idempotent and has the following useful property: .

$$\boldsymbol{R}^T\boldsymbol{PR} = \boldsymbol{PR}. \tag{7}$$

3 Sensitivity of the least squares method

The basic principle of robust estimation is to deflate the influence of outliers through a refinement of the stochastic model. As previously mentioned, a key role in the LS adjustment is played by the weighted sum of squares of the LS residuals $\Omega = \boldsymbol{V}^T\boldsymbol{PV}$ [8]. Consequently, the potentially outlying observations can be identified reasonably well based on the sensitivity of the Ω to any perturbations of the variance components and/or weight elements.

If we rewrite Ω as

$$\Omega = \boldsymbol{L}^T\boldsymbol{PRL} \tag{8}$$

and differentiate with respect to p_i $(i = 1, 2, \therefore, n)$, we obtain

$$\frac{\partial \Omega}{\partial p_i} = \boldsymbol{L}^T \left(\frac{\partial}{\partial p_i} [\boldsymbol{PR}] \right) \boldsymbol{L} \tag{9}$$

Using Theorem A.96 of reference [10] yields

$$\frac{\partial \boldsymbol{S}}{\partial p_i} = -\boldsymbol{S} \cdot \frac{\partial}{\partial p_i} [\boldsymbol{A}^T \boldsymbol{PA}] \cdot \boldsymbol{S} = -\boldsymbol{SA}^T \frac{\partial \boldsymbol{P}}{\partial p_i} \boldsymbol{AS} \tag{10}$$

where $\boldsymbol{S} = (\boldsymbol{A}^T \boldsymbol{PA})^{-1}$. With Eq. (10) and the definition of $\boldsymbol{R}$ it follows that

$$\frac{\partial \boldsymbol{R}}{\partial p_i} = -\boldsymbol{A} \frac{\partial \boldsymbol{S}}{\partial p_i} \boldsymbol{A}^T \boldsymbol{P} - \boldsymbol{ASA}^T \frac{\partial \boldsymbol{P}}{\partial p_i} = -\boldsymbol{ASA}^T \frac{\partial \boldsymbol{P}}{\partial p_i} \boldsymbol{R} \tag{11}$$

from which we get

$$\frac{\partial}{\partial p_i} [\boldsymbol{PR}] = \frac{\partial \boldsymbol{P}}{\partial p_i} \boldsymbol{R} + \boldsymbol{P} \frac{\partial \boldsymbol{R}}{\partial p_i} = \boldsymbol{R}^T \frac{\partial \boldsymbol{P}}{\partial p_i} \boldsymbol{R} \tag{12}$$

Substituting Eq. (12) into Eq. (9) gives

$$\partial \Omega / \partial p_i = v_i^2 \tag{13}$$

which leads to

$$\frac{\partial \Omega}{\partial \sigma_{ii}} = \frac{\partial \Omega}{\partial p_i} \cdot \frac{\partial p_i}{\partial_{ii}} = -\frac{\left(\boldsymbol{c}_i^T \boldsymbol{PV} \right)^2}{\sigma_0^2} \tag{14}$$

with

$$\frac{\partial p_i}{\partial \sigma_{ii}} = -\frac{\sigma_0^2}{\sigma_{ii}^2} = -\frac{p_i^2}{\sigma_0^2} \tag{15}$$

where $\boldsymbol{c}_i$ is a unit vector with 1 as its i-th entry. The local sensitivities defined by Eq. (14) are very important because they indicate the direction to be followed to get the greatest decrease in the LS objective function $\Omega = \boldsymbol{V}^T \boldsymbol{PV}$ [2].

4 Local sensitivity-based robust estimation

Under the assumption of a variance-inflation model, the possible outlying observations are still normally distributed with expectation 0, but with an increased variance [14]. Consequently, the local sensitivities of Ω to any perturbation of the variance components are important information for the design of the robust estimation.
The i-th standardized local sensitivity indicator

$$T_i = \frac{\left(\boldsymbol{c}_i^T \boldsymbol{PV} \right)^2}{\sigma_0^2 \boldsymbol{c}_i^T \boldsymbol{PRc}_i} \tag{16}$$

has a Chi-squared distribution. One can conclude that Ω is sensitive to a perturbation in the i-th variance component, provided that the following condition

$$T_i > \chi^2_{1-\alpha;1} \tag{17}$$

is fulfilled, where α is a given significance level and $\chi^2_{1-\alpha;1}$ is the upper α-percentage point of the χ^2-distribution with 1 degree of freedom [7]. In such a situation, it is appropriate to multiply σ_{ii} by an inflation factor $\gamma_{ii}(> 1)$ to reduce its corrupting effects. For example, γ_{ii} can be selected as

$$\gamma_{ii} = \begin{cases} 1 & T_i \le \chi^2_{1-\alpha;1} \\ T_i / X^2_{1-\alpha;1} & T_i > \chi^2_{1-\alpha;1} \end{cases} \tag{18}$$

Motivated by the famous Huber's M-estimation principle [6], the iterative robust estimator of unknowns can be obtained under the following condition:

$$\boldsymbol{V}^T [\bar{D}(\boldsymbol{L})]^{-1} \boldsymbol{V} = \min \tag{19}$$

or equivalently, under

$$\boldsymbol{V}^T \bar{\boldsymbol{P}} \boldsymbol{V} = \min \tag{20}$$

with

$$\bar{D}(\boldsymbol{L}) = diag(\gamma_{ii}\sigma_{ii}), \ \bar{\boldsymbol{P}} = \sigma_0^2 [\bar{D}(\boldsymbol{L})]^{-1} = diag(f_i p_i)$$

where $f_i = 1/\gamma_{ii}$ $(i = 1, 2, ..., n)$ are the so-called damping coefficients [3]. Thus, one can obtain an estimate of unknowns in model (1) as follows

$$\hat{\boldsymbol{X}} = (\boldsymbol{A}^T \bar{\boldsymbol{P}} \boldsymbol{A})^{-1} \boldsymbol{A}^T \bar{\boldsymbol{P}} \boldsymbol{L} \tag{21}$$

The above iterative process continues until a stop criterion is satisfied or a given number of iterations has been accomplished.

5 Variation characteristics of MDB measure

The ability of an estimation of parameters to detect outliers is called, according to Baarda [1], the reliability. The internal reliability as represented by the MDB describes the smallest size of the model errors which can be detected with the uniformly most powerful (UMP) test statistics [12, 13]. Interestingly enough, the UMP test statistic, which was derived under the assumption of a mean-shifted model in [12, 13], is simply the standardised local sensitivity indicator defined by Eq. (16).
The MDB measure related to the test statistic T_i is readily obtained as [12, 13]

$$\text{MDB}_i = \sigma_0 \sqrt{\frac{\lambda_0}{c_i^T \boldsymbol{PRc}_i}} \tag{22}$$

where λ_0 is the non-centrality parameter. A level of significance (the probability of rejecting H_{0i}, although it is true) together with a detection power (1 minus the probability of accepting H_{0i} when H_{ai} is true) will give a non-centrality parameter. A typical choice for λ_0 is $\lambda_0 = 4.13^2 \approx 17$ [13].
With Eq. (12), we get

$$\frac{\partial}{\partial p_j} \left[\boldsymbol{c}_i^T \boldsymbol{PRc}_i \right] = \left(\boldsymbol{c}_j^T \boldsymbol{Rc}_i \right)^2 \tag{23}$$

It follows immediately that

$$\begin{aligned} \boldsymbol{c}_i^T \bar{\boldsymbol{P}} \bar{\boldsymbol{R}} \boldsymbol{c}_i - \boldsymbol{c}_i^T \boldsymbol{PRc}_i &= \sum_{j \in J} \frac{\partial}{\partial p_j} \left[\boldsymbol{c}_i^T \boldsymbol{PRc}_i \right] \cdot \left(f_j p_j - p_j \right) \\ &= \sum_{j \in J} \left(\boldsymbol{c}_j^T \boldsymbol{Rc}_i \right)^2 \left(f_j - 1 \right) p_j \end{aligned} \tag{24}$$

where $\bar{\boldsymbol{R}} = \boldsymbol{I} - \boldsymbol{A} (\boldsymbol{A}^T \bar{\boldsymbol{P}} \boldsymbol{A})^{-1} \boldsymbol{A}^T \bar{\boldsymbol{P}}$ and $J = \{ j : f_j < 1 \}$ denotes the set of indices of the observations which are flagged as potential outliers using the UMP test statistics. As a result, we get

$$\sigma_0\sqrt{\frac{\lambda_0}{\boldsymbol{c}_i^T\boldsymbol{PRc}_i}} \le \sigma_0\sqrt{\frac{\lambda_0}{\boldsymbol{c}_i^T\bar{\boldsymbol{P}}\bar{\boldsymbol{R}}\boldsymbol{c}_i}},\quad i = 1,2,\ldots,n \tag{25}$$

which shows that the adjustment outputs, including the estimate of the unknown parameters, become more and more insensitive to outliers during the iterative process. Consequently, the M-estimator given by Eq. (21) is robust.

6 Discussions

As an alternative to T_i, its squared root

$$w_i = \frac{\boldsymbol{c}_i^T\boldsymbol{PV}}{\sigma_0\sqrt{\boldsymbol{c}_i^T\boldsymbol{PRc}_i}} \tag{26}$$

has a standard normal distribution. Accordingly, the damping coefficients can be determined as follows

$$f_i = \begin{cases} 1 & |w_i| \le c \\ (c/|w_i|)^2 & |w_i| > c \end{cases},\ (i = 1,2,\ldots,n) \tag{27}$$

where the constant c serves as a threshold value and usually can be taken from the interval [1.5, 2.0]. Unfortunately, the variance factor of unit weight is unknown in many practical applications. In such a case, the parameter σ_0 used in Eq. (26) may be replaced by its normalized median absolute deviation (MAD) estimator [11].

In the modified M-estimation of Huber [6, 7], the damping functions are as follows:

$$f_i = \begin{cases} 1 & |w_i| \le c \\ c/|w_i| & |w_i| > c \end{cases},\ (i = 1,2,\ldots,n) \tag{28}$$

Comparing Eqs. (28) and (27), we see that the outlying observations receive a smaller weight in the proposed method than in Huber's method.

It is easy to verify that, the test statistic w_i coincides with the i-th standardized LS residual for uncorrelated observations. However, the situation is completely different in the case of correlated observations.

7 Numerical example

A simulated leveling network shown in Fig. 1 was used as a test example. The observations and their weights are listed in Table 1. (The elevations of A and B are 100.000 m and 110.000 m, respectively).

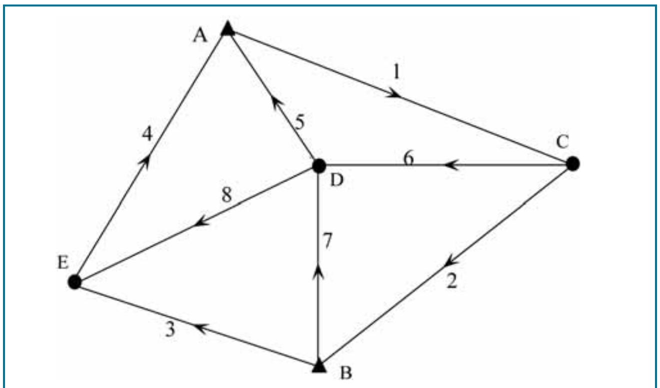

Fig. 1: Leveling network

Table 1: Simulated leveling network

Line	Observed height difference (m)	Relative Weight
1	6.148	2
2	3.860	1
3	– 4.518	1
4	– 5.489	2
5	– 4.187	5
6	– 1.966	1
7	– 5.807	2
8	1.311	1

Table 2: Computation results

Scheme	Estimated elevations for C, D, and E (m)			RMS (m)
1	106.1472	104.1867	105.4894	0.0052
2	106.2224	104.1876	105.5122	0.0499
3	106.1500	104.1882	105.4872	0.0057

For illustrative purposes, two artificial outliers, + 0.15 and + 0.09 m are introduced to the first and the eighth observation, respectively. The following three schemes are performed:

Scheme 1: LS estimation without any additional outliers;

Scheme 2: LS estimation with two simulated outliers in the data; and

Scheme 3: local sensitivity-based robust estimation with two simulated outliers in the data.

In the Scheme 3, the iteration process stops when the difference between the estimated unknowns at two consecutive iteration steps is less than 0.01 m. The tuning constant c in Eq. (27) was set to be 1.5. Convergence is already obtained after three iterations. The estimated elevations and their RMS errors by using these three schemes are presented in Table 2.

It can be seen from Table 2 that the LS estimator is sensitive to outliers. Compared to the LS method, the local sensitivity-based robust estimation provides more reliable

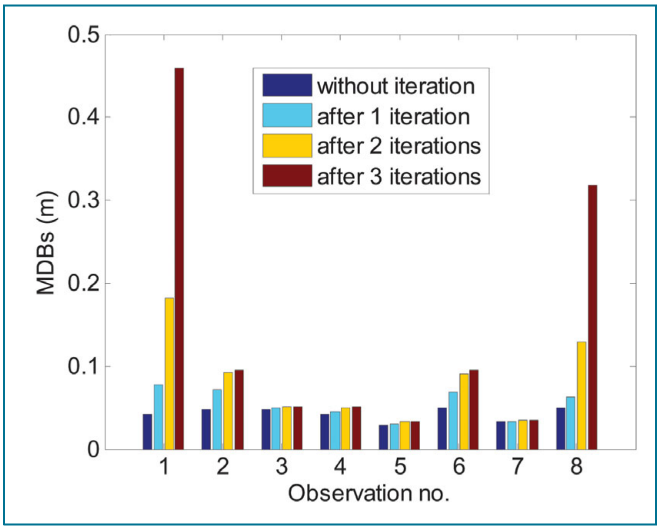

Fig. 2: Changes of the MDBs during the iteration process

and accurate results when the observations are contaminated by gross errors. The estimated elevations using the Scheme 3 are much closer to the LS estimates without outliers (Scheme 1). Fig. 2 depicts the variation characteristics of the MDBs in the robust estimation. Clearly, for every observation, the associated MDB measure becomes larger and larger with progress in the iterative process. This shows that the outputs of the adjustment get more and more insensitive to outliers in the observations.

8 Conclusions

Outliers in observations invalidate the outputs of LS adjustments and, thus, any conclusion based on them. The effects of any perturbation of the variance components and/or weight elements on the LS method are analyzed and tested. A local sensitivity-based robust estimation is developed. Theoretical analyses and numerical results indicate that the outputs of the adjustment become more and more insensitive to outliers in the process of the iterations, thus confirming the robustness of the proposed method.

9 Acknowledgements

This project is sponsored by the National Natural Science Foundation of China (Grant Nos. 40874007 and 40674012).

References

[1] Baarda, W., 1968. A testing procedure for use in geodetic networks. Netherlands Geod. Comm. Publ. on Geodesy, New Series, 2(5)

[2] Castillo, E.; Hadi, A. S.; Conejo, A.; Fernández-Canteli, A., 2004. A general method for local sensitivity analysis with application to regression models and other optimization problems. Technometrics, 46(4): 430–444

[3] Gargula, T.; Krupinski, W., 2007. The use of conic equation as a damping function in robust estimation. Allgemeine Vermessungs-Nachrichten, 114(10): 337–340

[4] Guo, J.; Ou, J.; Wang, H., 2007. Quasi-accurate detection of outliers for correlated observations. Journal of Surveying Engineering, 133(3): 129–133

[5] Hampel, F. R.; Ronchetti, E. M.; Rousseeuw, P. J.; Stahel, W. A., 1986; Robust Statistics: The Approach Based on Influence Functions. John Wiley and Sons, New York

[6] Huber, P. J., 1981. Robust Statistics. John Wiley and Sons, New York

[7] Koch, K. R., 1999. Parameter Estimation and Hypothesis Testing in Linear Models, 2nd Edn. Springer, New York

[8] Leick, A., 2004. GPS Satellite Surveying, 3rd Edn. John Wiley and Sons, New York

[9] Pope, A. J., 1976. The statistics of residuals and the detection of outliers. NOAA Tech. Rep. NOS 65 NGS 1, Rockville, Maryland, USA

[10] Rao, C. R.; Toutenburg, H., 1995. Linear Models: Least Squares and Alternatives. Springer, New York

[11] Rousseeuw, P. J.; Leroy, A. M., 1987. Robust Regression and Outlier Detection. John Wiley and Sons, New York

[12] Teunissen, P. J. G., 1985. Quality control in geodetic networks. In: Optimization and design of geodetic networks. Eds. E. W. Grafarend and F. Sanso, Springer, Berlin, 526–547

[13] Teunissen, P. J. G., 1998. Minimal detectable biases of GPS data. Journal of Geodesy, 72(4): 236–244

[14] Wieser, A.; Brunner, F. K., 2002. Short static GPS sessions: robust estimation results. GPS Solutions, 5(3): 70–79

[15] Xu, P., 2005. Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness. Journal of Geodesy, 79(1–3): 146–159

Adresses of the author:
Jianteng Guo, Information Engineering Univ., P.O. Box 22 01-1 60, Zhengzhou 450001, China E-Mail: jianfeng.guo@gmail.com
Jikun Ou, Institute of Geodesy and Geophysics, Chinese Academy of Sciences, Wuhan 430077, China