Marek Banaś*

A Review of Robust Estimation Methods Applied in Surveying

1. Introduction

Despite the continuous improvement of survey methods and advances made in survey equipment technology, the elimination of outliers still remains an issue today. When performing an adjustment one often assumes a very simple probability distribution of errors, such as a normal distribution. In classical statistics the correctness of the results relies on the assumption, that the chosen errors distribution model is strictly true. This is, in fact, often not the case, as the large errors occur considerably more often than the normal distribution would suggest. Even the high-quality samples analysed in astronomical research, containing several thousands of measurements each, do not follow the normal probability distribution. Deviations from the model may occur due to e.g. blunders in measuring, incorrect point numbering, errors made during data copying etc. [12].

Although there exists a wide range of literature concerned with gross errors detection and elimination, this surveying problem is still being discussed. There are many so-called methods robust against the influence of gross errors, which can generally be divided into two groups.

The first group includes methods based on the criteria of so-called robust estimation. These methods minimise the influence of the outlying observations on the final result of the computations by modifying of the observation weights.

The second of them consists of methods where results, obtained by the least squares adjustment are analysed with the use of statistical tests. In these methods an identified outlier is removed from the dataset. If multiple outliers occur, the iterative process of least squares adjustment is conducted and followed by tests. The observations suspected of gross errors are discarded from the dataset [1]. A few commonly used methods of these groups are presented below.

---

* The Bronisław Markiewicz State Higher School of Technology and Economics in Jarosław, Institute of Technical Engineering, Poland

2.   Robust Estimation Methods

2.1. Active Methods

Introduction

In recent years, the methods based on the *M*-estimation, introduced by Huber, have become very popular. The *M*-estimation is a wide class of estimators, which, among others, includes least squares estimators and maximum likelihood estimators. It uses the optimisation criterion provided below [13]:

$$\min_{X}\left\{\xi(X)=\sum_{i=1}^{n}\rho(v_{i})\right\}=\xi(\hat{X})=\sum_{i=1}^{n}\rho(\hat{v}_{i})$$ (1)

where:

$\xi(X)$ – the objective function for an adjustment task,

$\rho(v_{i})$ – a component of the objective function and n is a number of observations.

A detailed analysis of *M*-estimation class methods is carried out with use of the following characteristic functions:
  – influence function

$$\psi(v)=\frac{d}{dv}\rho(v)$$ (2)

  – weight function

$$\psi(v)=\frac{d}{d(v^{2})}\rho(v)$$ (3)

  – discipline function

$$\psi(v)=\frac{d^{2}}{dv^{2}}\rho(v)$$ (4)

The weight function is particularly important in determining the characteristics of individual robust estimation methods. One way of implementing the *M*-estimation is the use of nonlinear algorithms, the other way is the use of a modified least squares method. The second one uses the least squares method (LS) with iteratively modified observation weights.

The Huber Method

This method was proposed by Peter J. Huber in the paper [6] and repeated in the paper [7]. The objective function in this method has the following form:

$$\rho(v)=\begin{cases}\dfrac{1}{2}v^{2} & \text{for}\quad |v|\le f \\[2mm] f|v|-\dfrac{1}{2}f^{2} & \text{for}\quad |v|>f\end{cases}$$ (5)

where:

f – a parameter which defines interval of the admissible correction values,

v – a correction of the observation.

After introducing a weight p into the equation (5), the objective function takes the following form:

$$\rho(v) = \begin{cases} \dfrac{1}{2}pv^2 & \text{for} \quad |v| \le f \\ p\left(f|v| - \dfrac{1}{2}f^2\right) & \text{for} \quad |v| > f \end{cases} \tag{6}$$

The weight function in this method has the form:

$$w(v) = \begin{cases} \dfrac{1}{2}p & \text{for} \quad |v| \le f \\ \dfrac{1}{2}\dfrac{pf}{|v|} & \text{for} \quad |v| > f \end{cases} \tag{7}$$

The f parameter defines the limit of the admissible correction values v. Usually its value is determined in empirical way and depends on the quality level of the measurement data and the character of the adjustment problem [15].

The Hampel Method

The Hampel function was proposed in paper [5]. This method is characterised by use of a damping function of lower damping capacity in comparison to the Huber Method. Two additional ranges of admissible correction values are defined, the first one to the left and the second one to the right of the admissible correction values interval. After introducing an observation weight the components of the objective function take the following form [10]:

$$\rho(v) = \begin{cases} \dfrac{1}{2}pv^2 & \text{for} \quad |v| \le f \\ p\left(f|v| - \dfrac{1}{2}f^2\right) & \text{for} \quad f < |v| \le g \\ \dfrac{p \cdot f\left(\dfrac{1}{2}v^2 - h \cdot |v|\right)}{g - h} \\ \quad + p\left(f\left(g - \dfrac{f}{2}\right) + \dfrac{f \cdot g}{h - g}\left(\dfrac{g}{2} - h\right)\right) & \text{for} \quad g < |v| \le h \\ 0 & \text{for} \quad |v| > h \end{cases} \tag{8}$$

The weight function corresponding to the objective function (8) takes the form:

$$w(v) = \begin{cases} \dfrac{1}{2}p & \text{for} \quad |v| \leq f \\[2mm] \dfrac{1}{2}\dfrac{pf}{|v|} & \text{for} \quad f < |v| \leq g \\[2mm] \dfrac{\dfrac{1}{2}p \cdot f \cdot \dfrac{h-|v|}{h-g}}{|v|} & \text{for} \quad f < |v| \leq g \\[2mm] 0 & \text{for} \quad |v| > h \end{cases} \tag{9}$$

Similarly as in the Huber Method the values of parameters which define the boundary values of the correction intervals v are determined in an empirical way. Within these correction intervals there is a different way of weight modification for the observations. In his paper [5] Hampel suggests setting uniform boundary values, which are a multiple of robust estimator of standard deviation and amount to $f = 2.0$, $g = 4.0$, $h = 8.0$.

The Danish Method

Torben Krakup [11] – an eminent Danish mathematician, physicist and geodesist – is the author of this low-cost and effective method, which was used for the outliers detection by the Danish Geodetic Institute for over 15 years. The method consists in performing the least squares adjustments, one after another, with an iterative change of the weight matrix for observations according to the damping function presented below [14]:

$$t(v_i) = \exp\left(-d\left(\frac{|v_i|}{f}\right)^k\right) \tag{10}$$

where d and k are empirical parameters of the damping function.

In paper [2] another form of the damping function and the equivalent weight matrix is presented. The observations weights are modified in the following way:

$$(P_{ii})_{k+1} = \begin{cases} (P_{ii})_k & |(v_i)_k| \leq \dfrac{c \cdot \sigma_0}{\sqrt{p_{ii}}} \\[3mm] (P_{ii})_k \cdot \exp\left(-\dfrac{|(v_i)_k| \cdot \sqrt{p_{ii}}}{c \cdot \sigma_0}\right) & |(v_i)_k| > \dfrac{c \cdot \sigma_0}{\sqrt{p_{ii}}} \end{cases} \tag{11}$$

where:

k – a number of iterations,

c – a constant parameter in interval,

σ_0 – root-mean-square error (RMSE) of unit weight.

The iteration process is repeated until convergence is achieved. The outlying observations get zero weight and the applied correction values depend on the observations value. There are two ways of proceeding after the process is performed. Either the estimated parameters are taken as the ultimate ones or the outlying observations are removed and the least squares method, with the use of the initial weights, is applied.

The Gaździcki Method

This method, which was proposed by Jerzy Gaździcki in his paper [4], is a development of the Danish method. The damping function in this method takes the following form:

$$t(v_i) = \left(1 + \frac{v_i - f}{f(v_i)} \cdot \frac{2(1 - P^G)}{P^G (g - f)^2} \cdot \int_f^g f(v) dv \right)^{-1} \tag{12}$$

where:

$$f(v_i) = \frac{1}{\sqrt{2\Pi}} \exp\left(-\frac{v_i^2}{2} \right) \quad - \quad \text{a density function of a standardized normal distribution,}$$

$$P^G \quad - \quad \text{a probability that the correction value } v_i \text{ is not influenced by a gross error occurring in an other observation,}$$

$$\int_f^g f(v) dv \quad - \quad \text{a probability for the correction value } v_i \text{ to take a value from the interval } \langle f, g \rangle.$$

The calculations are performed by the least squares method with the use of the weights matrix presented below. The matrix is called equivalent due to the fact of being modified with the use of the damping function (12) after each iteration.

$$\hat{P} = \begin{cases} P_{i,i} & \text{for} \quad |v_i| < f \\ P_{i,i} \cdot t(v_i) & \text{for} \quad f \le |v_i| < g \\ 0 & \text{for} \quad |v_i| \ge g \end{cases} \tag{13}$$

The Linear Method

The method was introduced in the textbook [16]. In this method the a priori assigned mean errors of the observations undergo iterative modification. In the first step the calculations are performed by applying the classic least squares method. Inverses of squares of the mean observations are elements on the diagonal of the weights matrix **P**. The standard deviation of the corrections, which is an output of the first iteration, is used for the modification of errors according to the formula (14):

$$m_i^{(j+1)} = \begin{cases} m_i & \text{if} \quad \left|v_i^{(j)}\right| \le k_i^{(j)} \cdot m_i \\ m_i + \left|v_i^{(j)}\right| & \text{if} \quad \left|v_i^{(j)}\right| > k_i^{(j)} \cdot m_i \end{cases} \tag{14}$$

The coefficient k is calculated with the following formula:

$$k_i^{(j)} = \frac{\sigma^{(j)}}{m_i} \tag{15}$$

If the correction value of a particular observation exceeds the predetermined allowable range, this excess is added to the mean error of the observation. In this way the outliers get smaller weights in the following iteration. Having obtained the stabilisation of equation, we introduce the constant value of the coefficient $k = 2$ and perform final iterations.

The Method of Least Absolute Deviation

This method was proposed by Edgeworth in his work [3] and is probably the oldest method which incorporates the features of robust estimation [15]. The method, in its classic, unmodified form, consists in the minimisation of the sum of the absolute correction values for the observations.

$$\min \sum_{i=1}^{n} \left|\frac{v_i}{\sqrt{Q_{ii}}}\right| \tag{16}$$

where $Q_{ii} - i^{\text{th}}$ diagonal element of the variance-covariance matrix for observations.

The problem can be solved with use of the linear programming method. Another approach was proposed by R. Kadaj [8]. He replaced the function $y = |x|$ with function $y = \sqrt{x^2 + c^2}$, which enables multiple differentiation. The proposed function has similar properties provided that the c value is small enough:

$$\lim_{c \to 0} \sqrt{x^2 + c^2} = |x| \tag{17}$$

The component of the objective function is expressed as follows:

$$\rho(v) = p\sqrt{v^2 + c^2} \tag{18}$$

where:
        v – a correction of observation,
        p – an initial weight of observation.

The influence function takes the following form:

$$\psi(v) = p\frac{v}{\sqrt{v^2 + c^2}} \tag{19}$$

The weight function is defined by the following formula:

$$w\left(v\right) = p\frac{1}{2\sqrt{v^2 + c^2}} \tag{20}$$

This method is called the Alternative Method of Least Absolute Deviations.

A Choice Rule of Alternative

A method presented by R. Kadaj [9], i.e. "A Choice Rule of Alternative", is an important representative of robust methods. The method is distinguished from other approaches by the fact that its weight function is not a piecewise function. Acceptable intervals are not defined for this function.

The characteristic functions of this method take the following forms:

– component of the objective function

$$\rho\left(v\right) = -e^{-\frac{v^2}{2\sigma^2}} \tag{21}$$

where is the variance of observation, being defined before adjustment;

– influence function

$$\psi\left(v\right) = p \cdot v \cdot e^{-p\frac{v^2}{2}} \tag{22}$$

$$p = \frac{1}{\sigma^2} \tag{23}$$

– weight function

$$w\left(v\right) = \frac{1}{2}p \cdot e^{-p\frac{v^2}{2}} \tag{24}$$

2.2. Passive Methods

The idea of passive methods consists in performing statistical tests on the observations, being preliminary adjusted by the classic least squares method.

Iterative Data Snooping (IDS)

The procedure of data snooping proposed by W. Baarda [1] is a very popular method, which begins with performing a global test of the form:

$$\frac{\hat{v}^T P \hat{v}}{\sigma_0^2} \leq \chi_{\alpha,b}^2 \tag{25}$$

In the formula (25):

$\hat{v}$ – an estimator of the correction vector,

σ_0 – an *a priori* coefficient of variation,

$\chi^2_{\alpha,b}$ – a critical value read out from the table of the Chi-squared probability distribution (α – level of significance, b – the number of redundant observations).

Not satisfying the relation (25) by a set of observations means that it is contaminated by a gross error. In such a case a detailed search is conducted. It begins with the normalisation of the corrections conducted in the following way:

$$u_i = \frac{|v_i|}{\sigma_0 \sqrt{q_{v_i v_i}}} \qquad (26)$$

where:

v_i – a correction of the ith observation,

$q_{v_i v_i}$ – the ith diagonal element of a matrix, which is expressed as follows:

$$\mathbf{P}^{-1} - \mathbf{A}\left(\mathbf{A}^T \mathbf{P} \mathbf{A}\right)^{-1} \mathbf{A}^T .$$

Then the following test is performed:

$$u_i \leq k_\alpha \qquad (27)$$

where k_α is the critical value of the normal distribution $N(0,1)$ at a defined significance level α.

From the set of observations, which do not satisfy the criterion (25), the observation with the highest u is removed. Then the least squares adjustment is carried out once again (on the set not containing the before mentioned observation) and tests are performed. In case there are many outliers, a special iterative procedure is applied in order to specify the group of observations, which are indeed contaminated by gross errors [26].

τ-Test

The use of τ-test is a similar approach to the one proposed by Baarda. Both methods differ in the way the coefficient σ_0^2 is determined. In the case of data snooping procedure the coefficient is assigned a priori, whereas in τ-test it is determined with use of the least squares estimator of the coefficient of variation. The test value u_i^τ is expressed by the following formula:

$$u_i^\tau = \frac{|v_i|}{\hat{\sigma}_{v_i}} = \frac{|v_i|}{\hat{\sigma}_0 \sqrt{q_{v_i v_i}}} \qquad (28)$$

Since v_i and $\hat{\sigma}_0$ are dependent variables, we cannot use the Student's t-distribution. The test value u_i^τ is subject to τ distribution, which can be related to the Student's t-distribution as follows:

$$\tau_f = \frac{t_{\alpha,f}\sqrt{f}}{\sqrt{f-1+t_{\alpha,f}^2}} \tag{29}$$

where:

$t_{\alpha,f}$ – critical value of Student's t-distribution,
a – significance level,
f – number of degrees of freedom.

The observation whose u_i^τ value exceeds the critical value is removed from the observations set in the same way as it was done in the data snooping procedure. In case there is more than one outlier in the observations set, use of both presented procedures is possible. This might prove problematic though. The weak point of both methods is that they may cause elimination of the observation, which would be necessary to obtain a solution.

3. Conclusions

Wherever there is a need of handling measurement data, there exists also the problem of outliers. Robust estimation methods are there to help us to deal with it. This branch of mathematical statistics is used in various fields and is therefore a constant object of scientific research.

As far as geodesy and cartography are concerned, the robust estimation methods are applied in [15]:

- the adjustment of horizontal and vertical control networks,
- the processing of measurements obtained with the use of satellite techniques,
- the transformation of coordinates,
- the determination of a geoid model,
- the deformation studies of building objects,
- the determination of geometric distortion of building objects,
- the identification of intermediate points in the process of automatic aerial triangulation etc.

Such a wide range of applications implies a constant improvement of the existing robust estimation methods as well as the development of new ones.

References

[1] Baarda W.: *A testing procedure for use In geodetic networks.* Netherlands Geodetic Commission, Publications on Geodety, New Series, vol. 2, no. 5, Delft 1968.

[2] Caspary W.F.: *Concepts of Network and Deformation Analysis.* School of Geomatic Engineering, Monograph 11, The University of New South Wales, Sydney 1987.

[3] Edgeworth F.Y.: *On discordant observations.* Philosophical Magazine (5th series), 23, 1887, pp. 364–375.

[4] Gaździcki J.: *Least squares adjustment with a weight function.* Proceedings of the 7th International Sympodium of Geodetic Computations, Kraków 1985, pp. 299–310.

[5] Hampel F.: *The Influence Curve and Its Role in Robust Estimation.* Journal of American Statistical Association, vol. 69, no. 346, 1974, pp. 383–393.

[6] Huber P.J.: *Robust Estimation of a Location Parameter.* The Annals of Mathematical Statistics, vol. 35, no. 1, 1964, pp. 73–101.

[7] Huber P.J.: *The 1972 Wald Lecture Robust Statistics: A Review.* The Annals of Mathematicals Statistics, vol. 43, no. 4, 1972, pp. 1041–1067

[8] Kadaj R.: *Eine verallgemeinerte Klasse von Schätzverfahren mit praktichen Anwendungen.* Zeitschrift für Vermessungswesen, vol. 113, no. 4, 1988, pp. 157–166.

[9] Kadaj R.: *Die Methode der besten Alternative: ein Ausgleichungsprinzip für Beobachtungssysteme.* Zeitschrift für Vermessungswesen, vol. 109, no. 6, 1984, pp. 301–308.

[10] Kamiński W., Wiśniewski Z.: *Analiza wybranych, odpornych na błędy grube metod wyrównania obserwacji geodezyjnych. Część II. Analiza.* Geodezja i Kartografia, t. XLI, z. 3–4, 1992, pp. 183–195.

[11] Krarup T., Kubik K., Juhl J.: *Götterdämmerung Over Least Squares.* Proceedings of International Society for Photogrammetry 14th Congress, Hamburg 1980, pp. 370–378.

[12] Kubik K.K., Weng W., Frederiksen P.: *Oh, Grosserrors.* The Australian Journal of Geodesy, Photogrammetry and Surveying, no. 42, 1985, pp. 1–18.

[13] Kwaśniak M.: *Effectiveness of chosen robust estimation methods compared to the level of network reliability.* Geodesy and Cartography, vol. 60, no. 1, 2011, pp. 3–19.

[14] Muszyński Z.: *Zastosowanie metod estymacji odpornej do geodezyjnego opisu deformacji obiektu budowlanego.* Acta Scientiarum Polonarum. Geodesia et Descriptio Terrarum, 7(4), 2008, pp. 3–14.

[15] Muszyński Z.: *Zastosowanie metod estymacji odpornej do identyfikacji obiektów budowlanych na podstawie pomiarów geodezyjnych.* Politechnika Wrocławska, 2007 (Ph.D. thesis).

[16] Osada E.: *Geodezja: podręcznik elektroniczny w Mathcadzie 2001i.* Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2002.

[17] Prószyński W., Kwaśniak M.: *Niezawodność sieci geodezyjnych.* Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2002.