GEODEZJA I KARTOGRAFIA
GEODESY AND CARTOGRAPHY
Vol. 55, No 2, 2006, pp. 77-94

The problem of solving systems of linear equations by means of neural networks

Jozef Gil

Department of Land and Environment Engineering
Institute of Building Engineering
University of Zielona Gora
2 Szafrana St., 65-516 Zielona Gora, Poland
e-mail: j.gil@ib.uz.zgora.pl

Received: 2 November 2005/Accepted: 5 May 2006

Abstract: The paper addresses the problem of solving overdetermined systems of linear equations by means of methods of robust estimations, which eliminate the effect of outliers on the estimation results. The process of estimating a vector of parameters was accomplished by means of circular in structure neural networks. Formulating the problem in the aspect of a method for estimating parameters requires formulating an energy function (objective function) whose form was modified by means of a determined weighting function.

In the final part of the paper the effectiveness of the methods described was evaluated in terms of controlling and diagnosing a geodetic observation system. The article is merely an introduction to a broadly understood problem of geodetic uses of robust estimators.

Keywords: Linear equations, robust estimation, neural networks

1. Introduction

Solving a system of linear equations is one of the basic problems in science and technology. One of the most frequent basic and practical tasks is estimating components of the vector of parameters of overdetermined systems of linear equations

$$\mathbf{A}\mathbf{x} = \mathbf{l} \tag{1}$$

where $\mathbf{A} = [a_{ij}] \in \mathrm{R}^{m \times n}$ $(m > n)$ is the design matrix, $\mathbf{l} \in \mathrm{R}^{m}$ is the vector of observations, and $\mathbf{x} = [x_1, x_2, ..., x_n]^{\mathrm{T}} \in \mathrm{R}^{n}$ is the vector of parameters of the linearised functional model being estimated. Particular role in that respect play methods for estimating parameters of the overdetermined system of observation equations. The method most frequently applied for estimating parameters of linear models (Gauss-Markov models) with a specified redundancy is the least squares method with the assumption of Gauss distribution of observation errors.

Although the least squares solution that is optimum in the sense of the l_2 norm frequently considerably approximates solutions obtained using other norms (Dahlquist and Björck, 1983), research on robust statistics proves (Andrews, 1974; Hampel et al., 1986) that for a distribution of errors subject to the Cauchy distribution (Fig. 1) the l_1 norm is the optimum minimization criterion in terms of the minimum variance of estimators; that norm is also used when the vector of observations $\mathbf{l}$ includes outliers, or when the distribution of errors of the vector $\mathbf{l}$ is not sufficiently known.

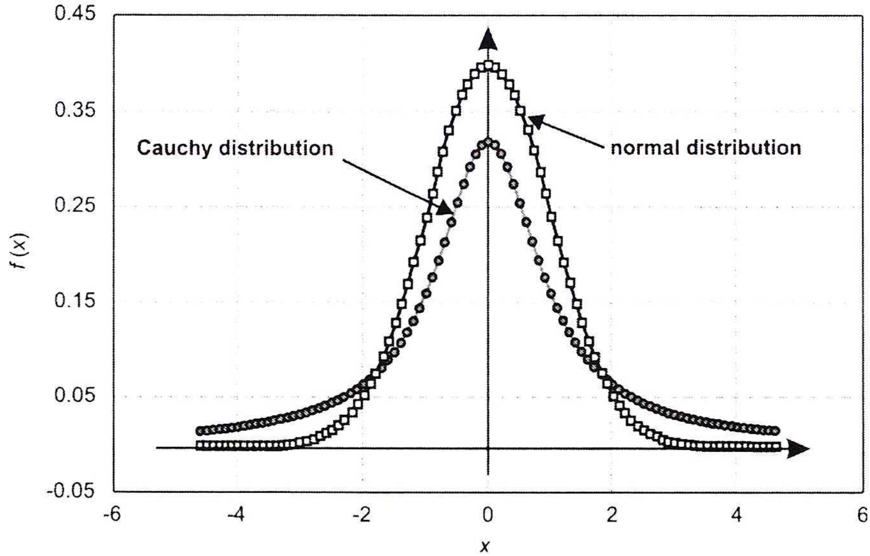

Fig. 1. Probability density of the normal distribution and the Cauchy distribution

The probability density of the Cauchy distribution, centred at the origin of the coordinate system is described by the formula

$$f(x) = \frac{1}{\pi} \frac{t}{t^2 + x^2} \quad (-\infty < x < +\infty) \tag{2}$$

The diagram of the probability density for the Cauchy distribution resembles the diagram for the normal distribution (Fig. 1), but in infinity it approaches the x-axis more slowly. The expected value $E\{X\}$, and consequently the variance $Var\{X\}$, of the random variable X, subject to the Cauchy distribution, do not exist.

In order to identify and reduce the effect of outlying observations on the solution, methods of robust estimations are used, formulated by means of modifying the objective function in the form of a weighting function assigned to each observation. Outlying observations considerably distort the values of parameters estimated by means of traditional least squares. In this case, the gradient of the standard objective function in the form of the square of the norm of the vector of residuals is a linear function of that vector. It means that the influence of the value of the vector of residuals on the estimated vector of parameters is linear. The methods of robust estimations are connected to the method of the greatest likelihood (Wiśniewski, 1982) and the so called "rule of the choice of an alternative" (Kadaj, 1984).

The purpose of this paper is to show the features of "resistance" to outlying observation errors of objective functions modified by restructuring it to increase its effectiveness and the use of the method of determining the estimator $\mathbf{x}^*$ of the vector of parameters $\mathbf{x}$ and the vector of residuals $\mathbf{v} = \mathbf{A}\hat{\mathbf{x}} - \mathbf{l}$ by means of neural networks of a circuit structure. The estimation strategy, regardless of the function type, does not introduce numerical limitations due to the fact that the Hessian is not positive definite, what is required in the Newton's method.

2. Formulating the problem

Formulation of robust estimation problems in the aspect of solving them by means of neural networks requires constructing a suitable energy function (objective function) $F(\mathbf{x})$, whose lowest energy state corresponds to the optimum value $\mathbf{x}^*$ of the vector of parameters $\mathbf{x}$. The energy function should not be associated with the physical meaning of energy; its name results from similarities between their properties. If the variations of the energy function during the operation of the algorithm are not positive, then the energy function is the Lapunov function (Korn and Korn, 1983).

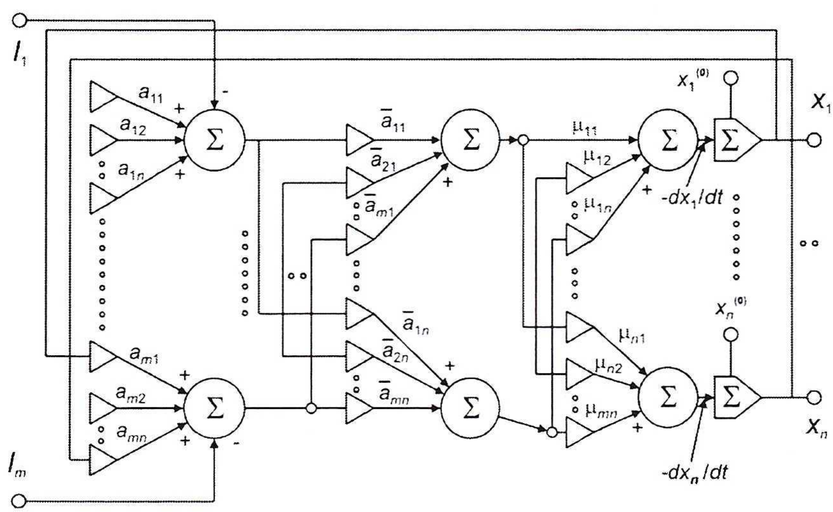

Fig. 2. A general scheme of the architecture of a neural network for solving overdetermined systems of linear equations of the form $\mathbf{Ax} = \mathbf{l}$

In general, the problem of estimating the vector of parameters $\mathbf{x}$ of the system of observation equations (1) consists in minimizing the energy function

$$F(\mathbf{x}) = \sum_{i=1}^{m} \omega_i[v_i(\mathbf{x})] \tag{3}$$

where $v_i(\mathbf{x}) = a_i^{\mathrm{T}}\mathbf{x} - l_i = \sum_{j=1}^{n} a_{ij}x_j - l_i$ are coordinates of the vector of residuals $\mathbf{v}$, and $\omega_i(v_i)$ represents an arbitrarily chosen convex function of $\mathbf{x}$ in the whole space R^n, that ensures the convergence of the minimization algorithms (Cichocki and Unbehauen, 1992). The choice of that function whose form is essential in the process of estimating parameters is arbitrary, providing it is differentiable, because in the process of minimization by means of neural networks the gradient of the objective function as the activation function must be known. The diagram of the architecture of the neural network used for solving systems of linear equations has been presented in Figure 2 (Cichocki and Unbehauen, 1992).

Gradient methods, whose operation bases are known values of parameters of the gradient vector of the objective function, belong no doubt to effective optimisation methods. Considering the effectiveness of gradient methods as the basis for further discussion, the problem of estimating parameters of energy functions consists in solving a system of differential equations

$$\frac{d\mathbf{x}}{dt} = -\eta(t)\nabla F(\mathbf{x}) \tag{4}$$

where $\mathbf{x} = [x_1, x_2, ..., x_n]^{\mathrm{T}}$, and

$$\nabla F(\mathbf{x}) = \left[\frac{\partial F(\mathbf{x})}{\partial x_1}, \frac{\partial F(\mathbf{x})}{\partial x_2}, ..., \frac{\partial F(\mathbf{x})}{\partial x_n}\right] = \mathbf{A}^{\mathrm{T}}(\mathbf{A}\mathbf{x} - \mathbf{l}) \tag{5}$$

The matrix of learning coefficients $\eta(t) = \eta_{ij}(t)$ is a diagonal $n \times n$ matrix. Considering that

$$F[\mathbf{x}(t)] > 0 \tag{6}$$

and

$$\frac{d}{dt}F[\mathbf{x}(t)] = \sum_{j=1}^{n} \frac{\partial F}{\partial x_j}\frac{dx_j}{dt} = [\nabla F(\mathbf{x})]^{\mathrm{T}}\frac{d\mathbf{x}}{dt} \leqslant 0 \tag{7}$$

one can notice that the function F is the Lapunov function, which for the solution $\mathbf{x}(t)$ is a real function of the C^1 class such that $F(0) = 0$. The solution of (4) is thus asymptotically stable.

Parameters of the function will be estimated by solving a system of differential equations in the form

$$\frac{dx_j}{dt} = -\sum_{p=1}^{n} \eta_{jp}\left\{\sum_{i=1}^{m}\left[a_{ip}\varphi_i\left(\sum_{k=1}^{n} a_{ik}x_k - l_i\right)\right]\right\} \tag{8}$$

where the function

$$\phi_i[v_i(\mathbf{x})] = \phi_i\left(\sum_{i=1}^{n} a_{ik}x_k - l_i\right) \tag{9}$$

is an activation function, i.e. an influence function.

With the weighting function $\omega_1[v_i(\mathbf{x})] = v_i^2$, the values of estimators obtained as a result of minimizing the function

$$F_1(\mathbf{x}) = \sum_{i}^{m} v_i^2(\mathbf{x}) = \|\mathbf{Ax} - \mathbf{l}\|_2 \tag{10}$$

correspond to the least squares estimators. In the case of unequally accurate observations one has to deal with the standard energy function in the form (Coleman et al., 1990)

$$F_2(\mathbf{x}, \mathbf{P}) = \sum_{i=1}^{m} p_i v_i^2(\mathbf{x}) = (\mathbf{Ax} - \mathbf{l})^{\mathrm{T}}\mathbf{P}(\mathbf{Ax} - \mathbf{l}) \tag{11}$$

with the weight matrix $\mathbf{P} = \mathrm{diag}(p_1, p_2, ..., p_m)$. For this case the weighting function is

$$\omega_2[v_i(\mathbf{x})] = p_i v_i^2(\mathbf{x}) \tag{12}$$

It is known from subject literature that in order to identify outlying observations and eliminate their deformable effect on the estimation results, a number of modifications to the energy function have been introduced in the form of the weighting functions, which exhibit properties resistant to outlying observation errors.

The activation functions $\phi_i[v_i(\mathbf{x})]$ mentioned in the following part of the paper are the derivatives of the m-argument functional $F(\mathbf{x})$ in the direction of $\mathbf{v}(\mathbf{x})$. When formulating weighting functions the activation function of the resistant estimator is assumed continuous and bounded; the breakdown point α^* as a specified limit of random errors, used to obtain the expected solution of the problem of estimation is their characteristic feature.

Let G be a particular family of cumulative distribution functions and F – a real-valued function determined on G, i.e. a functional. For the discrete distribution $E(G) = \sum_{i=1}^{m} x_i p_i$ (p_i – the probability of adopting the value x_i from the set $A \in R$). Replacing G with the empirical cumulative distribution function G_m, and x_i with $v_i(\mathbf{x})$, leads to (Ostasiewicz, 1999)

$$E(G_m) = \int v_i dG_m[v_i(\mathbf{x})] = \mathbf{v}(\mathbf{x}) \tag{13}$$

because $dG_m[v_i(\mathbf{x})] = 1/m$, $i = 1, 2, ..., m$; (m – number of observations).

Assume further that the set of observations contains $100\alpha\%$ of outlying observations, located at point $\mathbf{v}(\mathbf{x})$. Then the cumulative distribution function of the observations will be

$$G_{v_i(\mathbf{x}),\alpha}(y) = (1-\alpha)G_m(y) + \alpha\Delta_{\mathbf{v}(\mathbf{x})}(y) \tag{14}$$

where $G_m(y)$ is the a cumulative distribution function corresponding to observations free of gross errors, and

$$\Delta_{\mathbf{v}(\mathbf{x})} = \begin{cases} 1 \text{ when } y \geqslant v_i(\mathbf{x}) \\ 0 \text{ when } y < v_i(\mathbf{x}) \end{cases} \tag{15}$$

is the cumulative distribution function of observations affected with gross errors. It indicates that the cumulative distribution function $G_{v_i(\mathbf{x}),t}$ is located in the neighbourhood $t \leqslant \alpha$ of the cumulative distribution function $G_m(y)$.

From the above, the activation function of the functional F for the cumulative distribution function G_m is defined as the derivative of F at G_m in the direction of $\Delta_{\mathbf{v}(\mathbf{x})}$.

$$IF[\mathbf{v}(\mathbf{x}); G_m] = \lim_{\alpha\to 0} \frac{F(G_{\mathbf{v}(\mathbf{x}),\alpha}) - F(G_m)}{\alpha} \tag{16}$$

The influence function IF describes the local resistance of the estimator to outlying observations. From the definition of an activation function it results that the breakdown point α corresponds to the minimum participation of point disturbances at $\mathbf{v}(\mathbf{x})$, and the functional $F(G_{v_i(x),\alpha})$ is unlimited because of an undefined number of observations. If the minimum α does not exist, then the value α_m^* is adopted on the basis of the precision of the measurement instrument and the method used.

The mean is not a good resistant estimator. For the mean one has

$$E(G_{\mathbf{v}(\mathbf{x}),\alpha}) = (1-\alpha)E(G) + \alpha\mathbf{v}(\mathbf{x}) \tag{17}$$

In that case the form of the influence function is as follows

$$IF[\mathbf{v}(\mathbf{x}); E, G] = \lim_{\alpha\to 0}\left\{\frac{\alpha[\mathbf{v}(\mathbf{x}) - E(G)]}{\alpha}\right\} = \mathbf{v}(\mathbf{x}) - E(G) \tag{18}$$

and $\alpha^* = 0$, because the influence function is unbound. For the empirical cumulative distribution function G_m, the $\alpha_m^* = 1/m$, what means that even one outlying observation will change the value of the estimator. For the median and for the M-estimator (an example of the greatest likelihood estimator) $\alpha^* = 0.5$ (Ostasiewicz, 1999), which means that the breakdown of the estimator will take place only when the number of outlying observations reaches at least half of all observations carried out. It is worth to mention that the form of the M-estimator is not explicit, but it results from the solution (usually with the iterative method) of a particular equation. In practice, a restricted activation function is replaced with its smooth approximation.

3. Activation functions of chosen estimators

A convex function of resistant properties, used in linear algebra and research concerning robust statistics is the function

$$\omega_3[v_i(\mathbf{x})] = \frac{\alpha}{\beta} \ln\{\cosh[\beta v_i(\mathbf{x})]\} \tag{19}$$

suggested by Karayanninis (Karayiannis and Venetsanopoulos, 1992). The corresponding activation function has the form

$$\varphi\phi_3[v_i(\mathbf{x})] = \frac{\partial\{\frac{\alpha}{\beta} \ln[\cos[(\alpha v_i(\mathbf{x}))]]\}}{\partial[v_i(\mathbf{x})]} = \alpha \tanh[\beta v_i(\mathbf{x})] \tag{20}$$

For $\alpha = 1$ and large values of β, the function

$$\tanh[\beta v_i(\mathbf{x})] \approx 1 - 2e^{-\beta v_i(\mathbf{x})} \tag{21}$$

and then the function (20) converges to $|v_i(\mathbf{x})|$ (Fig. 3). After switching the values of parameters α and β, the estimators obtained approach classic least squares estimators.

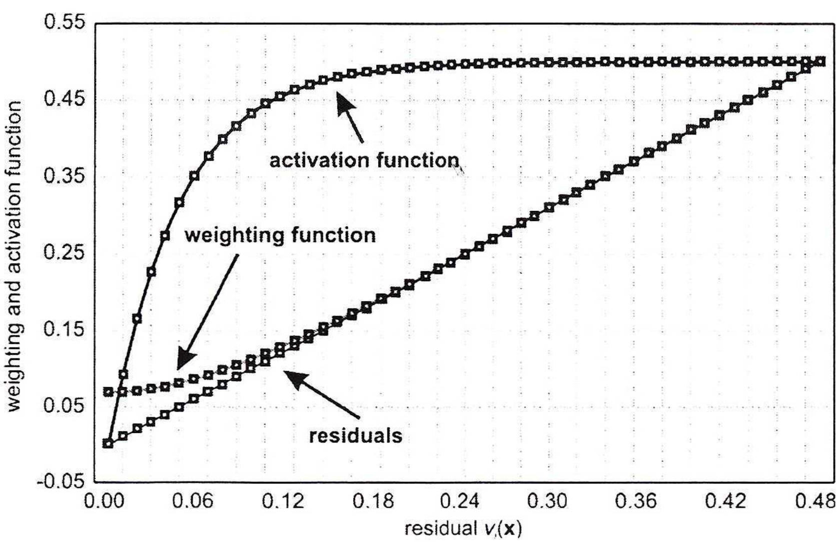

Fig. 3. Weighting function and the activation function according to Karayiannis

The following function used sometimes in digital signal processing

$$\omega_4[v_i(\mathbf{x})] = |v_i(\mathbf{x})| \tag{22}$$

also belongs to the class of convex functions. Its form corresponds to the rule of minimum absolute residuals, as "natural" robust estimation. The function (22) is irregular and in numerical application it requires the use of special procedures of mathematical programming (Abdelmalek, 1980; Kadaj, 1998) or the use of numerically simple algorithm by means of neural networks. The activation function is defined as

$$\phi_4[v_i(\mathbf{x})] = \mathrm{sgn}[v_i(\mathbf{x})] \tag{23}$$

The function $\mathrm{sgn}[v_i(\mathbf{x})]$ determines the sign of the derivative from the left or from the right in the neighbourhood of the point $\mathbf{x}$ (the function (22) is a continuous function, but not differentiable).

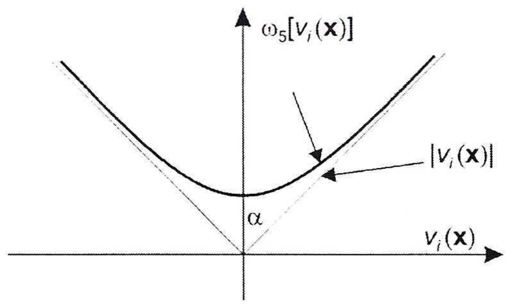

Fig. 4. The hyperbolic function

An alternative is the hyperbolic function (Fig. 4) in the form (Kadaj, 1988)

$$\omega_5[v_i(\mathbf{x})] = \sqrt{v_i^2(\mathbf{x}) + \alpha^2} \tag{24}$$

and the activation function corresponding to it

$$\phi_5[v_i(\mathbf{x})] = \frac{\partial \sqrt{v_i^2(\mathbf{x}) + \alpha^2}}{\partial v_i(\mathbf{x})} = \frac{v_i(\mathbf{x})}{\sqrt{v_i^2(\mathbf{x}) + \alpha^2}} \tag{25}$$

approaches 1 for $v_i(\mathbf{x}) \to \infty$ and a particular value of α (Fig. 5).

Considering

$$\lim_{v_i(\mathbf{x}) \to -\infty} \frac{\sqrt{v_i^2(\mathbf{x}) + \alpha^2}}{v_i(\mathbf{x})} = -1$$

and

$$\lim_{v_i(\mathbf{x}) \to +\infty} \frac{\sqrt{v_i^2(\mathbf{x}) + \alpha^2}}{v_i(\mathbf{x})} = +1$$

the function (24) has two sloping asymptotes: $\omega_5[v_i(\mathbf{x})] = |v_i(\mathbf{x})|$ (Fig. 4). For $\alpha \to 0$ the function approaches the function resulting from the rule of the minimum of the sum of absolute residuals (cf. (22)), and with growing α the estimators of parameters of the system approach the values of the least squares estimators.

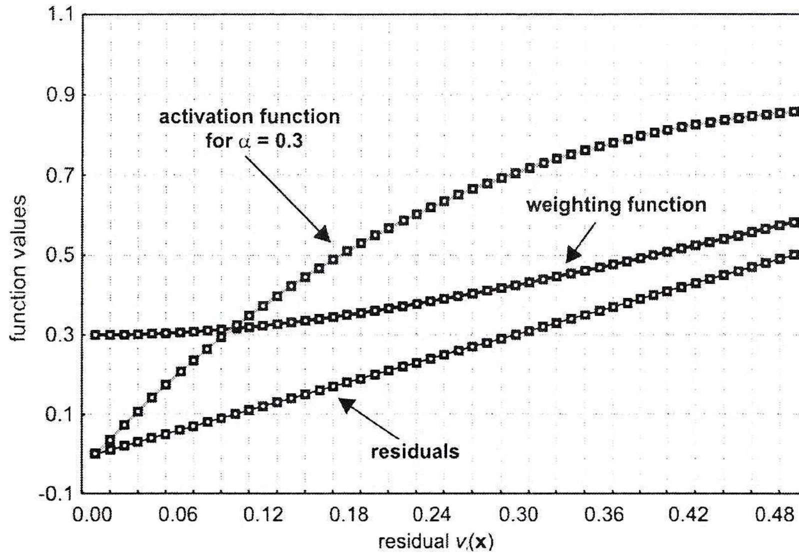

Fig. 5. Weighting function and the activation function according to Kadaj

Favourable results of identifying outlying observations can also be obtained by adopting a modification of the objective function in the form of the logarithmic function (Liano, 1994)

$$\omega_6[v_i(\mathbf{x})] = \log[1 + 0.5v_i^2(\mathbf{x})] \tag{26}$$

whose activation function

$$\phi_6 = \frac{\partial \log[1 + 0.5v_i^2(\mathbf{x})]}{\partial v_i(\mathbf{x})} = \frac{v_i(\mathbf{x})}{1 + 0.5v_i^2(\mathbf{x})} \tag{27}$$

is the non-linear function of $v_i(\mathbf{x})$. For $v_i(\mathbf{x}) \rightarrow \infty$

$$\lim_{v_i(\mathbf{x})\to\infty} \frac{\frac{v_i(\mathbf{x})}{1+0.5v_i^2(\mathbf{x})}}{v_i(\mathbf{x})} = \lim_{v_i(\mathbf{x})\to\infty} \frac{1}{1 + 0.5v_i^2(\mathbf{x})} = 0$$

what indicates an increase in suppression of growing gradient values, and as a result of an increase in suppression of the effect of outlying observations on the estmation results (Fig. 6).

Another function modifying the objective function is the commonly known weighting function used in the "Danish" method of robust estimation

$$\omega_7[v_i(\mathbf{x})] = \begin{cases} [v_i(\mathbf{x})]^2 & \text{for } |v_i(\mathbf{x})| \leqslant k \\ \exp\{-\alpha[|v_i(\mathbf{x})| - k]^\beta\} & \text{for } |v_i(\mathbf{x})| > k \end{cases} \tag{28}$$

where k is a limit of random errors, and α, β are suppression parameters.

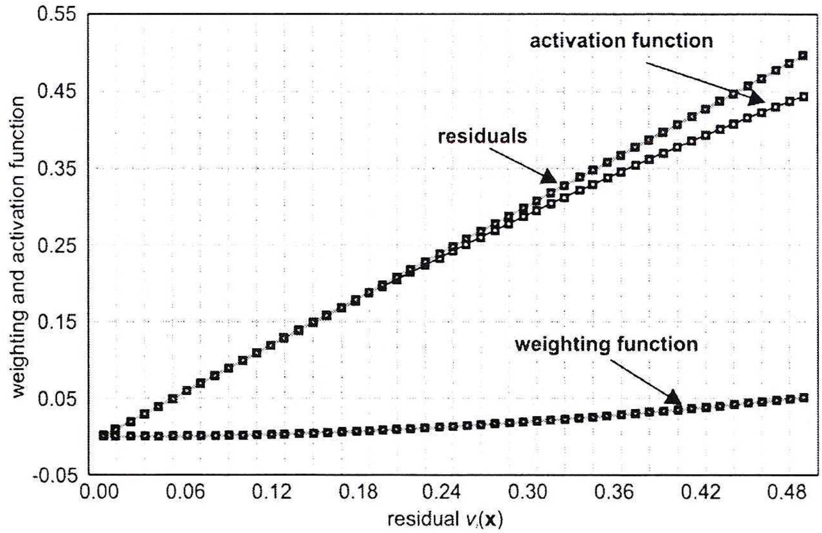

Fig. 6. Logarithmic weighting function and the activation function

For $|v_i(\mathbf{x})| \leqslant k$ the activation function has the form

$$\phi_7[v_i(\mathbf{x})] = 2v_i(\mathbf{x}) \tag{29}$$

while for $|v_i(\mathbf{x})| > k$ it is as follows

$$\phi_7[v_i(\mathbf{x})] = \exp\{-\alpha[|v_i(\mathbf{x})| - k]^\beta\} \times \{\alpha\beta[|v_i(\mathbf{x})| - k]^{\beta-1}\,\mathrm{sgn}[v_i(\mathbf{x})]] \tag{30}$$

The use of the function (28) (there are a number of forms of this function (Szczepański, 2004)) results in increasing suppression of gradient values for large residuals $v_i(\mathbf{x})$, what causes suppression of the influence of outlying observations (Fig. 7), because for $v(\mathbf{x}) \to \infty$

$$\lim_{|v_i(\mathbf{x})| \to \infty} \frac{\exp\{-\alpha[|v_i(\mathbf{x})| - k]^\beta\}}{|v_i(\mathbf{x})|} \to 0$$

and

$$\lim_{|v_i(\mathbf{x})| \to \infty} \frac{\exp\{-\alpha[|v_i(\mathbf{x})| - k]^\beta\} \times \{\alpha\beta[|v_i(\mathbf{x})| - k]^{\beta-1}\} \times \mathrm{sgn}[v_i(\mathbf{x})]}{|v_i(\mathbf{x})|} \to 0$$

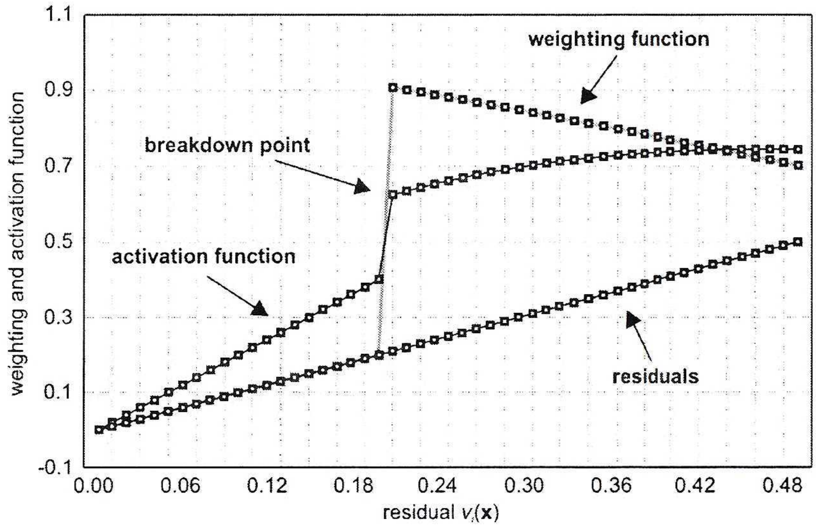

Fig. 7. The "Danish" weighting function and the activation function

Using the algorithm for minimizing the energy function by means of neural networks the author has simplified the "Danish" function excluding the coefficient β with no harm to the minimization result.

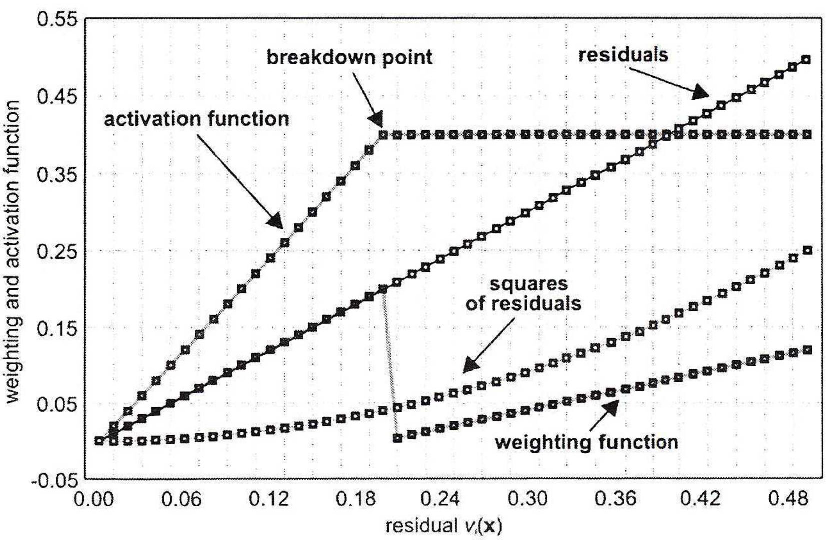

Fig. 8. The weighting function and the activation function according to Huber

The most popular modification of the energy function is the Huber weighting function (Fig. 8) of the form (Huber, 1981)

$$\omega_8[v_i(\mathbf{x})] = \begin{cases} [v_i(\mathbf{x})]^2 & \text{for } |v_i(\mathbf{x})| \leqslant \alpha \\ \alpha\,|v_i(\mathbf{x})| - \alpha^2 & \text{for } |v_i(\mathbf{x})| > \alpha \end{cases} \tag{31}$$

where the coefficient α like in the case of e.g. the hyperbolic function (24) is the limit of random errors, adapted to the specificity of a task considered. In this case it is the spline function (a parabola and two half-lines tangent to it). When the residual equals α, the value of the activation function (a derivative of the weighting function)

$$\phi_8[v_i(\mathbf{x})] = \begin{cases} -\alpha & \text{for } v_i(\mathbf{x}) < -\alpha \\ v_i(\mathbf{x}) & \text{for } |v_i(\mathbf{x})| \leqslant \alpha \\ \alpha & \text{for } v_i(\mathbf{x}) > \alpha \end{cases} \tag{32}$$

equals α. Huber has shown that if the function $\omega_8[v_i(\mathbf{x})]$ is bound by the absolute value of the constant α, then the estimator resulting from that function is an estimator with the least variance in a class of functions of that bounding (the asymptotic feature of robust estimators). It is also true for other estimators satisfying that assumption.

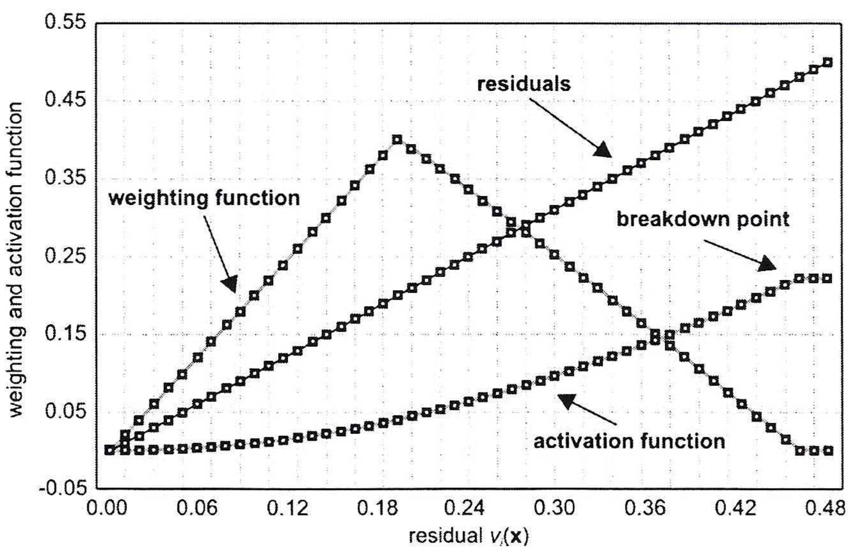

Fig. 9. The weighting function and the activation function according to Hampel

Favourable minimization results in terms of robust estimations can be obtained by applying the energy function that uses the Hampel weighting function (Hampel, 1971; Hampel et al., 1986).

$$\omega_9[v_i(\mathbf{x})] = \begin{cases} v_i[(\mathbf{x})]^2 & \text{for } |v_i(\mathbf{x})| \leqslant a \\ \gamma \ln[\cosh(\delta v_i(\mathbf{x}))] & \text{for } a \leqslant |v_i(\mathbf{x})| \leqslant b \\ v_i(\mathbf{x}) = \text{const} & \text{for } |v_i(\mathbf{x})| > b \end{cases} \tag{33}$$

It is an extension of the Huber function and its derivative (activation function) with respect to $v_i(\mathbf{x})$ is

$$\phi_9[v_i(\mathbf{x})] = \begin{cases} 2v_i(\mathbf{x}) & \text{for } \ |v_i(\mathbf{x})| \leqslant a \\ \gamma \tanh\{\delta[b - |v_i(\mathbf{x})|]\} & \text{for } \ a \leqslant |v_i(\mathbf{x})| \leqslant b \\ 0 & \text{for } \ |v_i(\mathbf{x})| > b \end{cases} \tag{34}$$

where a and b represent the ranges of the Hampel function, and γ and δ are certain constants (Fig. 9).

The Hampel function consists of three functions: a square function within the range of $[-a, a]$, a hyperbolic function within the range of $[|a|, |b|]$, and a constant function. Within the range of $[-a, a]$, the residuals affect linearly the values of the activation function. With an increase of the values of residuals, the influence of the value of the activation function decreases, and after the threshold value b is exceeded, activation function reaches 0. The state of suppression of the influence of outlying observations with a simultaneous retention of those observations whose values do not exceed a certain threshold value results from the course of the function.

4. Characteristics of methods of robust estimations

The least squares method in its classic version is not resistant to large blunders, that distort the estimated parameters of the observation equation system. The estimation of parameters of overdetermined systems of linear equations, including modifications of the energy function, and the estimation of parameters of those systems in the norm l_1, belong to the class of robust estimations.

It is worth to mention that values of parameters corresponding to average observation values are obtained from adjustment based on the l_2 norm, and parameters obtained from adjustment based on the l_1 norm correspond to values of observation medians on the assumption that $\mathbf{A}$ matrix is of a full rank.

The effectiveness of the discussed methods of estimating parameters, in terms of detecting outlying observations, will be checked for a vertical geodetic network consisting of 56 points and 87 height differences observed, including a simulation of one outlying observation. The test for randomness in the form $u = v/m_v$ (m_v – residual error) has given $u = 3.0$ for that observation. The levelling network was adjusted with minimum limits of degrees of freedom because of no restrictions on the observations.

In order to evaluate the effectiveness of the methods discussed in the aspect of detecting observations affected with a non-random error, random variables (residuals) have been classified, and a respective histogram has been made. The classification of the values of residuals has been done on the basis of an optimum choice of lengths of classification cells (lengths of class units), that makes possible to obtain maximum information in a cell, according to the formula (Brillouin, 1969)

$$s = \left[t + \ln \frac{tT}{s(s-1)} \right] \tag{35}$$

where T is the range of the feature investigated, t is the length of the class unit, and s is a value of the variable classified.

The solution of (35) is the number $k1 \approx t/s$, that expresses an optimum ratio of the length of the class unit to the value of the variable classified. If the size of cells is too small in comparison to the "size" of the data analysed, then the classification is ambiguous; if the size is too big, the effectives of the classification is low.

Statistical information for data (components of vectors of residuals) in the number of 696, which has been obtained on the basis of tests of the methods examined, has been accumulated in 148 class units, and a graphical representation of this statistical material has been presented in Figure 10.

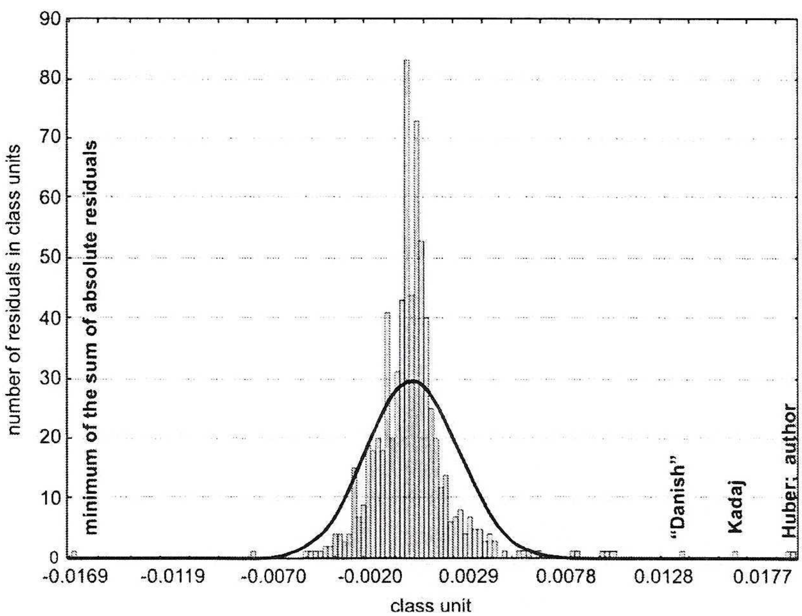

Fig. 10. A histogram of residuals on the basis of the minimization of energy functions with 148 class units

Following approximate practical procedures in order to determine the number of class units t, for example by means of the formula

$$t \leqslant 5 \log n \tag{36}$$

where n is the number of the sample, one should emphasize certain global features of the data, but anomalies in the data cannot be modelled (Fig. 11). It results from the diagrams presented in Figures 10 and 11 that the distribution of residuals determined by the adjustment of a vertical network by means of neural networks with the use of robust estimations is not a normal distribution (a considerable rise near zero), but it is close to the Laplace distribution or to the bilateral gamma distribution. Because of the existing discrepancies in the estimation of the distribution of probability of random variables it is possible to use more sophisticated testing methods (e.g. Law and Kelton, 1982).

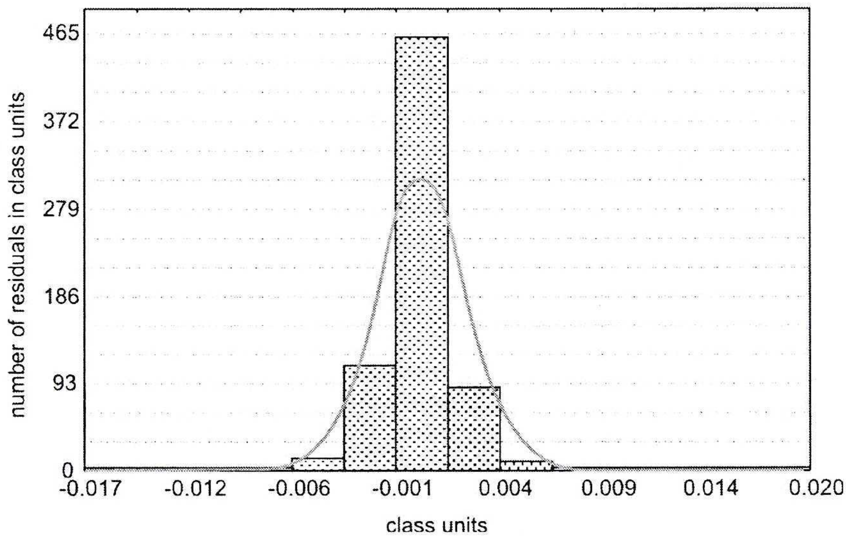

Fig. 11. A histogram of residuals on the basis of the minimization of energy functions with 14 class units

As far as the evaluation of methods of robust estimations is concerned, it is possible to say that in general their use is limited to checking and diagnosing an overdetermined system of observation equations. The results of the tests carried out by the author indicate that as far as detecting of outlying observations is concerned, the best results are provided by estimation methods based on the use of the Huber, Kadaj, the author's and "Danish" functions as well as (which should be emphasized) the estimation method according to the rule of the minimum sum of modules. The other modifications of energy functions rank a bit lower in this classifications, and their effectiveness in identifying an observation affected with a non-random error is very similar. However, due to the features of the algorithm for minimizing the objective function, the way it is modified, and the values adopted as the limit for random errors, particular estimation methods provide slightly different results.

Let us also note that as a result of the modification of the energy functions in order to obtain resistance features of the objective function, parameters obtained slightly differ in terms of value (Fig. 12).

In the author's opinion the differences result from the way of modifying the objective function and the value adopted as the limit α of random errors (in this paper the limits of random errors are also denoted as $k1$, a, b). Moreover, it should be mentioned that the solution of the task of minimizing the objective function by means of neural

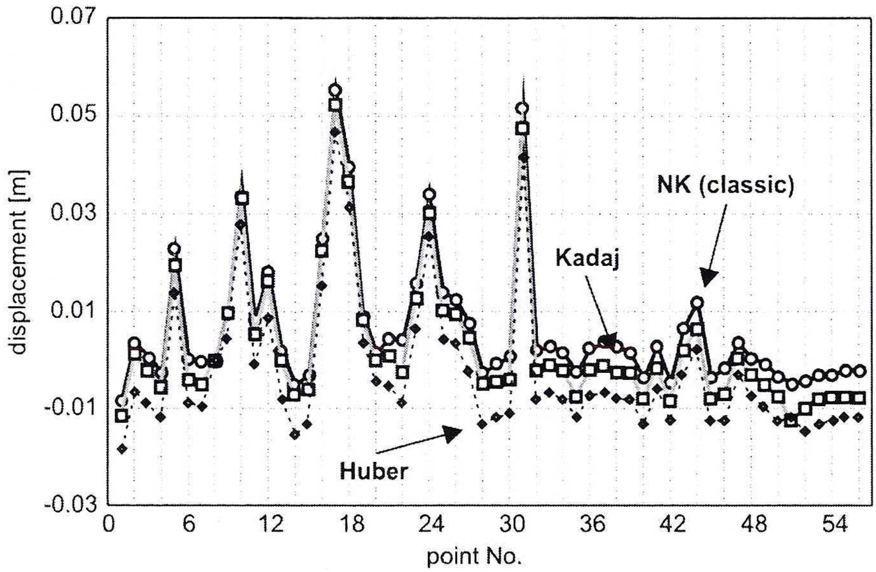

Fig. 12. Vertical displacements obtained by means of the minimization of selected energy functions

networks is an approximate solution (iterative methods), and the minimum obtained is local.

5. Conclusions

The above discussion on the resistance of selected objective functions to considerable observation errors and the results of minimizing these functions by means of neural networks make it possible to conclude that the robust estimators obtained are less vulnerable to outlying observations. From the practical point of view the outlying observation should be associated with a considerable error, that largely exceeds the range of probable estimations. An outlying observation for which the ratio v_i/m_{vi} (m_{vi} – a correction error) does not exceed the limit error G = 2 may remain unnoticed in the process of identifying observations with gross errors, especially when among the observations used to form the observation equation system there are several ones with errors exceeding several times the limit of random errors. Such a situation will appear when the value of the constant α (constants: $k1$, a, b) is too large, and for a small limitation of the weighting function on terms of its absolute value, observations without gross errors can be qualified as outlying observations. The limit value of the weighting function becomes important because of the possibility of obtaining different resistance measures. Outlying observations specified by means of methods of robust estimations should be eliminated one by one because each residual is a function of all observations and in consequence the outlying observation deforms all calculated residuals, especially those referred to the neighbouring observations (Nowak, 1982).

Eliminating a single outlying observation provides a chance to identify all remaining outlying observations by means of methods of robust estimations, and consequently to eliminate them. The evaluation of asymptotic characteristics of robust estimators remains an open question in the problem discussed.

Acknowledgements

The author of the article expresses his deep thanks to the anonymous reviewer for his inspiring participation as well as precious remarks and directions, which considerably increased the scientific meaning and quality of the article. The author expresses special gratitude to the Editor-in-chief Prof. Jan Krynski for his help and support in preparing the final version of the manuscript.

References

Abdelmalek N.N., (1980): *L_1 solution of overdetermined systems of linear equations*, ACM Trans. Math. Software 6, No 2, pp. 220-227.

Andrews D.F., (1974): *A robust method for multiple linear regression*, Technometrics 16, pp. 523-531.

Brillouin L., (1969): Science and information theory (in Polish), PWN, Warsaw, pp. 106-113.

Cichocki A., Unbehauen R., (1992): *Neural Networks for Solving Systems of Linear Equations and Related Problems*, IEEE Transactions on Circuits and Systems, Vol.39, No 2, pp. 124-127.

Coleman D., Holland P., Kaden N., Klema V., (1990): *A system of subroutines for iteratively reweighted least squares computation*, ACM Trans. Math. Software, pp. 327-336.

Dahlquist G., Björck Ä., (1983): *Numerical methods* (in Polish), PWN, Warsaw.

Hampel F.R., (1971): *A general quantitative definition of robustness*, Ann., Math. Statist., Vol. 42, pp. 1887-1896.

Hampel F.R., Roussew P.J., Ronchetti E.M., Stahel W., (1986): *Robust Statistics – the Approach Based on Influence Function*, J. Wiley, New York.

Huber P.J., (1981): *Robust Statistics*, J. Wiley, New York.

Kadaj R., (1984): *Die Methode der besten Alternative: Ein Ausgleichungsprincip für Beobachtungssysteme*, Zeitschrift für Vermessungswesen, H3, 113J, pp. 301-307.

Kadaj R., (1988): *Eine Klasse von Schätzverfahren mit praktischen Anwendungen*, Zeitschrift für Vermessungswesen, H8, pp. 157-165.

Kadaj R., (1998): *Models, methods and algorithms for computing kinematic networks in deformation surveying* (in Polish), Publications of Agriculture Academy in Cracow, Cracow, pp. 126-135.

Karayiannis N., Venetsanopoulos A., (1992): *Fast learning algorithms for neural networks, IEEE Trans. Neural Netwoworks*, Vol. 39, pp. 453-474.

Korn G.A., Korn T.M., (1983): *Mathematics for researchers and engineers* (in Polish), part 1, PWN, Warsaw, pp. 423-426.

Law A.M., Kelton D., (1982): *Simulation Modeling and Analysis*, New York: McGraw- Hill.

Liano K., (1994): *A robust approach to supervised learning in neural network*, IEEE Proc. ICNN, Orlando, pp. 513-516.

Nowak E., (1982): *Deformation analysis based on statistical verification of measurements and models used* (in Polish), Mathematical Symposium "Informatics in Geodesy", Warsaw.

Ostasiewicz W., (ed.) (1999): *Statistical methods of data analysis* (in Polish), Publications of Oscar Lange Economic Academy in Wroclaw, Wroclaw, pp. 312-326.

Szczepański J., (2004): *A method of geodetic network adjustment resistant on outlying tie points* (in Polish), Geodesia et Descriptio Terrarum, Geodesy and Cartography 3(1-2), Publications of Agriculture Academy in Wroclaw, Wroclaw, pp. 20-25.

Wiśniewski Z., (1982): *Estimation of geodetic object parameters and deterministic distortions of measurements using maximum likehood method* (in Polish), Proceedings of Agriculture-Technical Academy in Olsztyn, Geodezja i Urządzenia Rolne 12, pp. 51-69.

Zagadnienie rozwiązywania układów równań liniowych za pomocą sieci neuronowych

Józef Gil

Wydział Inżynierii Lądowej i Środowiska
Instytut Budownictwa
Uniwersytet Zielonogórski
ul. Szafrana 2, 65-516 Zielona Góra
e-mail: j.gil@ib.uz.zgora.pl

Streszczenie

W pracy omówiono zagadnienie rozwiązywania nadokreślonych układów równań liniowych z zastosowaniem metod estymacji mocnych, które eliminują niekorzystny wpływ obserwacji odstających na wyniki estymacji. Proces estymacji wektora parametrów został zrealizowany za pomocą sieci neuronowych o strukturze obwodowej. Formułowane zagadnienia w aspekcie ich rozwiązywania, wymagały sformułowania funkcji energetycznej (funkcji celu), której postać modyfikowano przez zastosowanie określonej funkcji wagowej.

W końcowej części pracy dokonano oceny skuteczności opisanych metod w zakresie kontroli i diagnostyki nadokreślonego układu równań obserwacyjnych. Artykuł stanowi jedynie przyczynek do szeroko pojętego zagadnienia geodezyjnych zastosowań estymatorów mocnych.