THE PROBLEM OF MINIMIZATION OF AN UNEVEN FUNCTION OF SEVERAL VARIABLES IN THE NORM l_1 AND l_∞ BY MEANS OF NEURAL NETWORKS

VAIRĀKU ARGUMENTU FUNKCIJAS MINIMIZĀCIJA l_1 UN l_∞ NORMĀS AR NEIRONU TĪKLU PALĪDZĪBU

JÓZEF GIL

University of Zielona Góra, Department of Civil and Environmental Engineering
65-246 Zielona Góra, Poland
Phone: + 4868 3282416, *fax*: + 4868 3247292, *e-mail*:J.Gil@ib.zgora.pl

Abstract. *The paper presents the problem of the minimization of an uneven function of several variables $F: R^m \to R^n$. In general, the minimization of an uneven function is difficult in terms of numerical application, especially when there is no information about the character of the unevenness of the function or any precise data about the distribution of observation errors. This paper presents two algorithms for the minimization of an uneven function by means of neural networks: solutions to an overdetermined system of linear equations according to the criteria of the norm l_1 and according to the criteria of the norm l_∞ (the Chebyshev norm) with the use of a square and exact penalty function. The results of particular solutions have been compared with the solution via the method of the least squares. The equalization tasks have been performed with minimum restrictions of extents of freedom.*

Keywords: *minimization of an uneven function in the norm l_1 and l_∞.*

Introduction

In a wide class of technical and scientific problems, especially in terms of the equalization of geodesic observations, the most frequent procedure for the estimation of the components of the parameters vector of linear models (Gauss-Markov models) is the method of the least squares. The use of the average-square solution to systems of linear equations with a specific redundancy, optimum in terms of the norm l_2 is closely connected to the assumption that observation errors are subject to the Gauss distribution.

Although, the average-square solution often effectively leads to the solution in other norms [3], research into robust statistics proves [5] that for the distribution of errors subject to uniform distribution, the most suitable criterion for minimization is the residuum norm $\|\mathbf{v}\|_\infty$ called the Chebyshev norm, and for the Cauchy distribution the optimum minimization criterion is used in the form $\|\mathbf{v}\|_1$. Moreover, the norm $\|\mathbf{v}\|_1$ is preferred when there is not enough information about the distribution of errors.

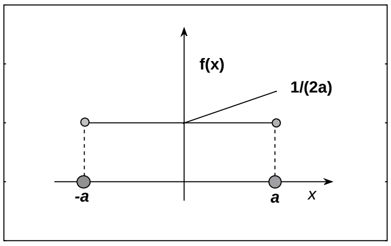

Fig. 1. **The density of a uniform distribution**

A uniform distribution (fig. 1), within the range of $[-a, a]$ is called distribution density.

ISBN 9984 – 779 – 06 – 8

$$f(x)=\frac{1}{2a} \qquad \text{for } |x| \le a \qquad (1)$$

and

$$f(x)=0 \qquad \text{for } |x|>a \qquad (2)$$

The desity of probability for the Cauchy distribution centered at the origin of the system has the form

$$f(x)=\frac{1}{\pi}\frac{t}{t^2+x^2}, \qquad (-\infty<x<+\infty) \qquad (3)$$

where $t>0$ is the scale parameter. The diagram of the density of the Cauchy distribution (fig. 2) resembles the diagram of a normal distribution, but it approximates the x axis very slowly.

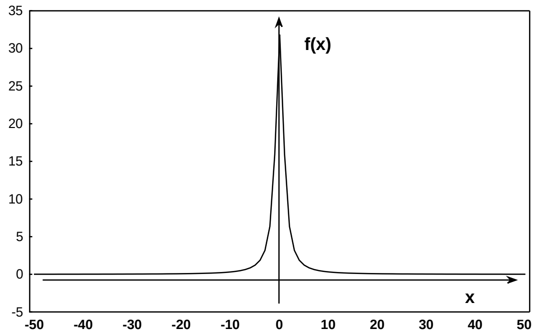

Fig. 2. **The density of the Cauchy distribution**

Because of the fact that it is difficult to determine the distribution density of the probability of the random vector which burdens the observation vector, the alternatives to the method of the minimization of the sum of the squares of corrections to the observation are the method of the minimization of the sum of modules (the minimization of the norm $\|\boldsymbol{v}\|_1$) and the method of minimax (the minimization of the Chebyshev norm $\|\boldsymbol{v}\|_\infty$). Up till now the abovementioned methods have been used mainly in procedures for the estimation of parameters of non-differentiable functions.

The purpose of this paper is to show the possibilities of determining the estimator $\hat{\boldsymbol{x}}$ of the vector of parameters $\boldsymbol{x}$ and the residuum $\boldsymbol{v}=\boldsymbol{A}\hat{\boldsymbol{x}}-\boldsymbol{l}$ of the linear model $\boldsymbol{Ax}\cong\boldsymbol{l}$ by means of neural networks according to the criterion in the form of the minimization of the norm $\|\boldsymbol{v}\|_1$ and the norm $\|\boldsymbol{v}\|_\infty$.

Formulation of the problem

First, let us cosider the general problem of solving an overdetermined system of linear equations $\boldsymbol{Ax}=\boldsymbol{l}+\boldsymbol{v}$ where $\boldsymbol{A}$ is the linear representation assigning the observation vector $\boldsymbol{l}\in R^m$ $(m\ge n)$ to the parameter vector $\boldsymbol{x}\in R^n$ according to the norm criterion $\|\boldsymbol{v}\|_k$, defined in the form of an objective function in the form [2]

$$F_k(x)=1/k\sum_{i=1}^{m}|v_i(\boldsymbol{x})|^k \qquad \text{for } (1\le k<\infty), \quad (4)$$

$$\text{and } v_i(\boldsymbol{x})=\sum_{i=1}^{n}a_{ij}x_j-l_i \quad (i=1,2,...,m).$$

The solution to the problem of the minimization of the function $F_k(\boldsymbol{x})$ boils down to the implementation of a set of differential equation

$$\frac{dx_j}{dt}=-\mu_j\frac{\partial F_k(\boldsymbol{x})}{\partial x_j} \qquad (5)$$

ISBN 9984 – 779 – 06 – 8

where $\mu_j > 0 \ \ (j = 1,2,...,n)$ is the learning coefficient.
The process of the minimization of the function $F_k(\boldsymbol{x})$ is a stationary process, where the change of the direction of the function gradient vector $F_k(\boldsymbol{x})$

$$\nabla F_k(x) = \left[\frac{\partial F_k(\boldsymbol{x})}{\partial x_1}, \frac{\partial F_k(\boldsymbol{x})}{\partial x_2}, ..., \frac{\partial F_k(\boldsymbol{x})}{\partial x_n}\right] \tag{6}$$

is dynamic in character, and the components of the vector gradient are activation functions in the form

$$g[v_i(\boldsymbol{x})] = /v_i(\boldsymbol{x})/^{k-1} \ sgn[v_i(\boldsymbol{x})], \tag{7}$$

and

$$\mathrm{sgn}[v_i(\boldsymbol{x})] = \begin{cases} 1 & for \quad v_i(\boldsymbol{x}) > 0 \\ -1 & for \quad v_i(\boldsymbol{x}) < 0 \end{cases} \tag{8}$$

The procedure for the minimization of the function $F(\boldsymbol{x})$ according to the criterion in the form of the norm $\|\boldsymbol{v}\|_1$ and according to the criterion in the form of the norm $\|\boldsymbol{v}\|_\infty$ is a little different. According to the criterion of the norm $\|\boldsymbol{v}\|_1$, the system of differential equations (5) for $k = 1$ (cf. formula (8)) is

$$\frac{dx_j}{dt} = -\mu_j \sum_{i=1}^{m} a_{ij} \ sgn[v_i(\boldsymbol{x})] \quad (i = 1,2,...,m) \tag{9}$$

where the activation function is the sign function.

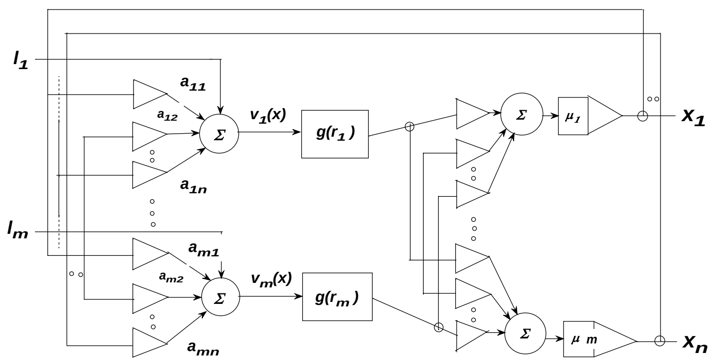

Fig. 3. **The block diagram of signal flow in a neural network for the solution to the task of minimization according to the criterion of the norm** $\|\boldsymbol{v}\|_k \ (1 \le k < \infty)$

The process of the estimation of parameters in the form of the block diagram of the signal flow, corresponding to the system of differential equations (9), has been illustrated in fig. 3.
Hence, it can be seen that the coordinates of the input vector are the coordinates of the column vectors of the matrix $\boldsymbol{A}$ and the coordinates of the vector of free terms $\boldsymbol{l}$. The correction vector $\boldsymbol{v}$ is calculated on the basis of the approximate values of the coordinates of the parameter vector $\boldsymbol{x}$, then the value of the signum activation function $\boldsymbol{v}(\boldsymbol{x})$, whose sign corresponds to the direction of the correction of the function $F_k(\boldsymbol{x})$ at the point $\boldsymbol{x}$ is also calculated. The following steps are in compliance with the method of the greatest descent, bearing in mind the scalar learning coefficient μ_j (the step coefficient) ($j = 1,2,...n$).

ISBN 9984 – 779 – 06 – 8

The norm $\|v\|_\infty$ - the Chebyshev norm as the criterion for the minimization of the maximum absolute error

We set the task of such a choice of the coordinates $x_j\ (j = 1,2,...,n)$ of the parameter vector $\boldsymbol{x}$, that the functional

$$V = \max_{1 \le i \le m} [/v_i(\boldsymbol{x})/] \qquad v_i(\boldsymbol{x}) = \sum_{j=1}^{n} a_{ij} x_j - l_i \tag{10}$$

which is the representation of each of the coordinates of the vector $\boldsymbol{x}$ separately, should reach the minimum value. The problem of optimization will be formulated by the introduction of the additional variable δ, and the original problem will be transformed to the problem modified by the minimization of the variable δ, including the constraint

$$/v_i(\boldsymbol{x})/ - \delta \le 0 \tag{11}$$

Thus, the solution to the problem of minimization according to the diagram in fig. 3, boils down to finding the least

$$\delta^{\bullet} = F_\infty(\boldsymbol{x}^{\bullet}) \ge 0 \tag{12}$$

so that the following condition is fulfilled

$$/v_i(\boldsymbol{x})/ \le \delta^{\bullet} \tag{13}$$

In order to minimize the value δ we will use, according to [7], the method of the square and exact penalty function [4], used in non-differentiable minimization. For the square penalty function, the energetic function (objective function) is:

$$F(\boldsymbol{x},\delta) = \alpha\delta + \frac{\beta}{2}\sum_{i=1}^{m} \{[\delta + v_i(\boldsymbol{x})]_-^2 + [\delta - v_i(\boldsymbol{x})]_-^2\} \tag{14}$$

where

$$[w]_- = \min\{0,w\} \tag{15}$$

The penalty coefficients $\alpha > 0, \beta > 0$ specify the participation of the penalty component.
If we mark as

$$C_{i1} = \delta + v_i(\boldsymbol{x}) \tag{16}$$

$$C_{i2} = \delta - v_i(\boldsymbol{x}) \tag{17}$$

and the following steps will depend on the gradient strategy as the activation function $f(C_{i1,i2})$, which assumes only two binary values 0 and 1, then according to the condition (15) we can write

$$f(C_{i1,i2}) = 0 \ \text{ if } \ \delta + v_i(\boldsymbol{x}) \ge 0 \ \text{ or } \ \delta - v_i(\boldsymbol{x}) \ge 0 \tag{18}$$

$$f(C_{i1,i2}) = 1 \ \text{ if } \ \delta + v_i(\boldsymbol{x}) < 0 \ \text{ or } \ \delta - v_i(\boldsymbol{x}) < 0 \tag{19}$$

The problem of the minimization of the value δ boils down solving the system of the two differential equations (after the initial values of the coordinates of the vector $\boldsymbol{x}$, the coefficient δ and the penalty coefficients α and β have been adopted).

$$\frac{d\delta}{dt} = -\eta\{\frac{\alpha}{\beta} + \sum_{i=1}^{m} [(\delta + v_i(\boldsymbol{x}))f(C_{i1}) + (\delta - v_i(\boldsymbol{x}))f(C_{i2})]\} \tag{20}$$

ISBN 9984 – 779 – 06 – 8

$$\frac{dx_j}{dt} = -\eta_j \sum_{i=1}^{m} \{a_{ij}[(\delta + v_i(\boldsymbol{x}))f(C_{i1}) - (\delta - v_i(\boldsymbol{x}))f(C_{i2})]\} \qquad (21)$$

When the exact penalty function [4] is used, the energetic function is defined as

$$F(\boldsymbol{x},\delta) = \alpha\delta - \beta\sum_{i=1}^{m}\{[\delta + v_i(\boldsymbol{x})]_- + [\delta - v_i(\boldsymbol{x})]_-\} \qquad (22)$$

The parameter δ and the coordinates x_j $(j = 1,2,...,n)$ of the parameter vector $\boldsymbol{x}$ are obtained from the solution to the set of differential equations

$$\frac{d\delta}{dt} = -\eta\left[\frac{\alpha}{\beta} - \sum_{i=1}^{m}[f(C_{i1}) + f(C_{i2})]\right] \qquad (23)$$

$$\frac{dx_j}{dt} = -\eta_j \sum_{i=1}^{m} a_{ij}[f(C_{i2}) - f(C_{i1})] \qquad (24)$$

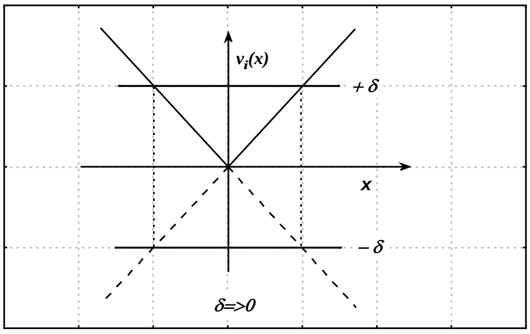

Fig. 4. **The graphical representation of the solution to the task of the minimization of the function** $F(x,\delta)$

The forms of the differential equations (20), (21), (23) and (24) result from the partial derivatives calculated and equated to zero of the energetic functions (14) and (22) with relation to the coefficient δ and the components x_i $(i = 1,2,...,n)$ of the parameter vector $\boldsymbol{x}$. The solution to the task of the minimization of the energetic functions (14) i (22) in compliance to the equation (11) has been presented in fig. 4.

The analysis of the algorithms under discussion used for the estimation of parameters of a vertical geodesic network

The abovementioned methods of estimation do not precisely determine the choice of a method of calculations. In the paper [9] the author clearly remarks that the method of the least squares is used with the assumption that particular observations are in an independent normal distribution. Also K.F. Gauss had his doubts about the optimality of this method and suggested an alternative procedure according to the rule of the least modules [6]. The verification of the linear model optimised via the method of the least squares requires a test concerning the normality of the distribution of the observation vector $\boldsymbol{l}$ with an application of the adequate way of classification of the random vector coordinates described in the paper [1].

In practice the minimization methods under discussion have been used for the assessment of the displacement of a building founded on expansive soil (clay). The overdetermined system of linear equations $\boldsymbol{Ax} = \boldsymbol{l} + \boldsymbol{v}$ put into a configuration of 17 lines $(m = 17)$ and 11 columns $(n = 11)$ has been solved in four variants:

- in the norm $\|\boldsymbol{v}\|_2$ (the standard deviation $m_0 = 0{,}28\ mm$),

ISBN 9984 – 779 – 06 – 8

- in the norm $//v//_1$ (the coefficient of accuracy for the solution $m_1 = 0{,}36\ mm$),
- in the norm $//v//_\infty$ with the use of a square penalty function (the accuracy coefficient for the solution $m_2 = 0{,}29\,mm$),
- in the norm $//v//_\infty$ with the use of an exact penalty function (the accuracy coefficient for the solution $m_3 = 0{,}33\ mm$).

In all the four cases the accuracy indices for the solution reached almost identical values, but the values of parameters estimated indicated certain differences. The best approximation to the average square solution proved to be the solution in the norm l_∞ with the use of a square penalty function (fig. 5) because of the values of the accuracy indices for the minimization and because of the value of the correlation coefficient r of the residuum vectors, which turned out to be equal $r_{v//x//2-v//x//\infty} = 0{,}96$.

The displacements specified via the solution to the task in the norm l_1 are different in value from the displacements in the norms l_2 and l_∞. The reasons for this discrepancy can be found in the density of probability of the random variable $\boldsymbol{1}$ different from the density of probability of the Cauchy distribution.

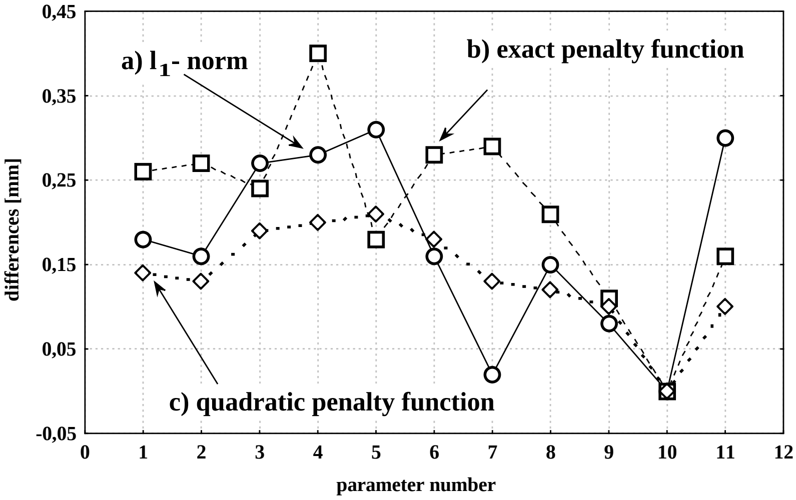

Fig. 5. **The diagram of the differences between the displacement values obtained from equalization in the norm l_1 and l_∞ and the results of equalization in the norm l_2**

Conclusions

The problem of the minimization of absolute residuum values is the counterpart to linear programming, similarly the ordinary method of the least squares can be regarded as a particular case of square programming. The most important difference in the approach to the method of minimization in the norm l_1 and the norm l_2 results from the form of energetic functions. The results of the minimization of an energetic function in the norm l_1 are usually similar to the results of the minimization of even functions [2]. The average values of parameters are obtained from the solution to the task via the method of the least squares, and the parameters resulting from the solution in the norm l_1 or in the norm l_∞ assume median values with the assumption that the matrix $\boldsymbol{A}$ is a full rank. For this reason the square of the norm l_1 will always be greater in value than the square of the norm l_2.

Optimization according to the criterion of the norm l_∞ which consists in the minimization of the maximum values of a function, is suitable for observations without disturbances such as impulses i.e. observations being in uniform distribution. The norm l_∞ becomes considerably

ISBN 9984 – 779 – 06 – 8

important in the process of the approximation of a uniform function, specified within a certain range or a discrete set of points, by means of the Chebyshev polynomials. The above example and the opinions on this subject presented in publications prove that uniform approximations are close to approximations via methods of the least squares [8].

References

1. Brillouin L., Nauka a teoria informacji, PWN, Warszawa 1969.
2. Cichocki A., Unbehauen R., Neural Networks for Solving Systems of Linear Equations – Part II:minimax and Least Absolute Value Problems, IEEE Transactions on Circuits and Systems, Vol. 39, No.9, 1992.
3. Dahlquist G., Björck Ä., Metody numeryczne, PWN, Warszawa 1983.
4. Fletcher R., Practical Methods of Optimization. New York: Wiley-Interscience, 1986.
5. Huber P.,J., Robust Statistics, New York:Wiley, 1981.
6. Kadaj R., Modele, metody i algorytmy obliczeniowe sieci kinematycznych w geodezyjnych pomiarach przemieszczeń i odkształceń obiektów, monografia, Wydawnictwo AR im. H. Kołłątaja w Krakowie, Kraków 1998.
7. Kennedy M.P., Chua L.O., Neural networks for nonlinear programming, IEEE Trans. Circuits Syst., vol. CAS-34, 1987.
8. Korn Granino,A., Korn Teresa, M., Matematyka dla pracowników naukowych i inżynierów, cz. II, PWN, Warszawa 1983.
9. Nowak E., Badanie przemieszczeń drogą statystycznej weryfikacji poprawności modelu i obserwacji, Mat. Sympozjum „Informatyka w Geodezji", Warszawa 1982.

ISBN 9984 – 779 – 06 – 8