EARTH OBSERVATION AND GEOMATICS ENGINEERING
website: https://eoge.ut.ac.ir

# On the use of two L1 norm minimization methods in geodetic networks

**A. R. Amiri-Simkooei**
Department of Geomatics Engineering, Faculty of Civil Engineering and Transportation, University of Isfahan, Isfahan, Iran

**Article history:**
Received: 5 November 2017, Received in revised form: 1 March 2018, Accepted: 10 March 2018

**ABSTRACT**

L1norm adjustment is a powerful technique to detect gross errors in geodetic observations. This paper investigates the results of two formulations that provide the L1 norm adjustment of a linear functional model. The usual method for implementation of the L1 norm adjustment leads to solving a linear programming (LP) problem. The formulation of the L1 norm minimization is presented based on the LP problem for a rank deficient linear(ized) system of equations. Then, an alternative technique is explained based on the least squares residuals. The results are tested on both linear and non-linear functional models, which confirm the efficiency of both formulations. The results also indicate that the L1 norm minimization, compared to the weighted least squares method, is a robust technique for the detection of blunders in geodetic observations. Finally, this contribution presents a data snooping procedure to the residuals obtained by the L1 norm minimization method.

**KEYWORDS**

L1 Norm Minimization
Data Snooping Procedure
Linear Programming Problem

## 1. Introduction

The method of least squares is the standard method used to obtain the unique estimates for the unknown parameters from a set of redundant measurements. Because of the inherent stochastic properties of the observations, the redundant observations are not usually compatible (consistent) in the functional model. In other words, due to the random behavior of observations, any sufficient subset of functionally independent observations would yield a different result. Therefore, there is no unique solution that simultaneously satisfies all equations. In this situation, it is the basic principle of the adjustment that gives a unique estimate for all of the model parameters with certain optimum properties. The original set of observations, which is denoted by the vector and includes redundant observations, is replaced by another set of estimates which satisfies the model. The difference between these two sets is

$$\mathbf{e} = \hat{\mathbf{y}} - \mathbf{y} \tag{1}$$

which is termed in the classical theory as 'residuals'. The residual vector converts the system of observation equations into an underdetermined model that has an infinite number of solutions. Amongst all of the possibilities, there exists one set of estimates that satisfies a certain criterion referred to as the least squares principle. The least squares principle states that (Mikhail,1976)

$$\mathbf{e}^{\mathrm{T}}\mathbf{W}\mathbf{e} = \sum_{i=1}^{m} w_i e_i^2 \rightarrow \min \tag{2}$$

where the matrix is the weight matrix. The above criterion, as an objective function, states that the sum of the squares of the weighted residuals should be as small as possible. This criterion is referred to as L2 norm minimization of the residuals, as well. To formulate the least squares solution, consider the vector of the unknown parameters in a linear(ized) parametric model as

$$\begin{aligned} &\mathbf{y} + \mathbf{e} = \mathbf{A}\mathbf{x}, \quad \mathbf{W} = \mathbf{Q}_{\mathbf{y}}^{-1} \\ &\mathbf{D}\mathbf{x} = \mathbf{0} \end{aligned} \tag{3}$$

E-mail addresses: amiri@eng.ui.ac.ir (A.R. Amiri-Simkooei)
DOI: 10.22059/eoge.2018.256034.1021

where the vector , as before, is the vector of residuals, the vector is the vector of observations, the matrix is the design matrix, the matrix is the datum matrix of the network, and the matrix is the covariance matrix of observations. The least squares solution of Eq. (3) is given as (Chen,1983):

$$\hat{\mathbf{x}} = (\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{y} \qquad (4)$$

And correspondingly, the least squares residuals are obtained as

$$\hat{\mathbf{e}} = \mathbf{A}\hat{\mathbf{x}} - \mathbf{y} \qquad (5)$$

The covariance matrix of the estimated residuals can be obtained as follows:

$$\mathbf{Q}_{\hat{\mathbf{e}}} = \mathbf{Q}_{\mathbf{y}} - \mathbf{A}(\mathbf{A}^{\mathrm{T}}\mathbf{W}\mathbf{A} + \mathbf{D}^{\mathrm{T}}\mathbf{D})^{-1}\mathbf{A}^{\mathrm{T}} \qquad (6)$$

The above equations are derived by solving the famous normal equations. Such equations can be obtained, for example, from differentiating the objective function(2) .The advantage of the least squares algorithm is mainly based on the simple use of matrix notations and electronic computers for performing the computations. It is important to note that the application of the least squares does not require a priori knowledge of the distribution associated with the observations. This may be another advantage of the least squares technique. It can be shown that the least squares estimate is unbiased and of minimum variance. This means that the least squares estimator is the most efficient estimator of unknown parameters. It can also be proved that when the observations are normally distributed, the least squares estimates will give identical results to those obtained from the method of maximum likelihood.

## 2. Principles of $L_1$ norm minimization

The least squares adjustment is a standard and powerful method to estimate the unique and optimum solution for unknown parameters. It should be assumed, for better statistical interpretation of the results, that the observation errors are random and normally distributed. If these assumptions are violated, e.g., in cases where the observations are affected by gross errors in addition to random ones, the robust estimation techniques produce superior results. Such robust techniques may be useful for data screening and gross error detection before a final adjustment is made by the least squares. One such robust technique involves minimization of the $L_1$ norm of the residuals, namely

$$\mathbf{1}^{\mathrm{T}}|\mathbf{e}| = \sum_{i=1}^{m}|e_i| \rightarrow \min \qquad (7)$$

where $\mathbf{1}$ is an $m\times 1$ vector of ones (summation vector). For uncorrelated weighted observations, one can either enforce the weights by multiplying each equation by $w_i$, where $w_i$ is the weight associated with each observation, or modify the objective function as

$$\mathbf{w}^{\mathrm{T}}|\mathbf{e}| = \sum_{i=1}^{m} w_i|e_i| \rightarrow \min \qquad (8)$$

where $\mathbf{w}$ is an $m\times 1$ vector that contains the diagonal elements of the weight matrix $\mathbf{W}$. This is known as the objective function of the $L_1$ norm adjustment. It should be noted that the $L_1$ norm gives unbiased estimates like the least squares but they are not of minimum variance. This may destroy the statistical interpretation of the results.

The advantage of the $L_1$ norm minimization, compared to the least squares, is that it is less sensitive to gross errors. Therefore, the $L_1$ norm adjusted residuals may illustrate the blunders more straightforward than the least squares. For example, in repeated measurements of an unknown quantity, in order to estimate the population mean, the least squares yield the sample *mean*, whereas the $L_1$ norm minimization yields the sample *median*. It goes without saying that a gross error will affect the sample mean more than the sample median.

The $L_1$ norm minimization technique uses a special sufficient (and not redundant) subset of functionally independent observations, which minimizes the sum of the weighted absolute value of the other residuals. Accordingly, the technique would not result in an optimum estimation since all of the observations have not been used in this method. Regardless of the minimization of the objective functions (2) and (8), the parametric model and the datum constraints in Eq. (3) remain the same. The presence of an absolute sign in the criterion (8) precludes one from differentiating, which is customary in the least squares formulation. To convert Eq. (8) into something applicable, the usual strategy is based on the use of the 'linear programming'. For this purpose, the *slack* variables (vectors) should be introduced to guarantee non-negativity and therefore write the objective function without absolute signs (see section 4).

## 3. Linear programming problem

Linear programming (LP) is a branch of operations research that deals with the problem of minimizing a linear function in the presence of linear equality and/or inequality constraints. It is a well-known and fast developing technique in mathematics, with multiple scientific applications. The subject of linear programming, sometimes called linear optimization, in standard form, concerns with the following problem: For $p$ independent variables $z_1, z_2, ..., z_p$ minimize the function

$$z = c_1z_1 + c_2z_2 + ... + c_pz_p \qquad (9)$$

Subject to the primary constraints

$$z_1 \geq 0, z_2 \geq 0, ..., \ z_p \geq 0 \tag{10}$$

and simultaneously subject to $q$ additional constraints of the form

$$h_{i1}z_1 + h_{i2}z_2 + ... + h_{ip}z_p = b_i, i = 1,2,...,q \tag{11}$$

where $z$ in Eq. (9) is the objective function to be minimized. The constraints in Eq. (10) are the non-negativity constraints. A set of variables $z_1, z_2, ..., z_p$ satisfying all constraints is called a *feasible* point or a feasible vector. The set of all such points constitutes the feasible region or the feasible space. Using this terminology, the linear programming problem can be stated as follows: Among all feasible vectors, find one that minimizes the objective function (9). A linear programming problem can be stated in a more convenient form using matrix notations. Considering the column vectors $\mathbf{z}, \mathbf{c}$, and $\mathbf{b}$ and the $q \times p$ matrix $\mathbf{H}$, the 'standard form' of the LP problem can be written as

$$\text{minimize} \quad z = \mathbf{c}^T\mathbf{z} \tag{12}$$

$$\text{subject to} \quad \mathbf{Hz} = \mathbf{b}, \ \mathbf{z} \geq \mathbf{0} \tag{13}$$

Another form of the LP problem is called the 'canonical form'. A minimization problem is in the canonical form if all variables are non-negative and all constraints in Eqs. (11) and (13) are of the type $\geq$. By simple manipulations, this problem can be transformed to the standard form. There are different methods to obtain the solution of a linear programming problem. The *simplex* method which was first recommended by Dantzig is a smart procedure that moves from an extreme point to another one with a better (at least not worse) objective (Dantzig, 1963). It also discovers whether the feasible region is empty and whether the optimal solution is unbounded. In practice, the method only enumerates a small portion of the extreme points of a feasible region. The reader is referred to (Bazaraa et al.,1990). There are, however, new methods for solving an LP problem, among which we can mention the interior point methods. The reader is referred to (Roos et al.,1997) in which they present an interior point method (IPM) to both the theory of LP and its algorithms.

## 4. LP formulation of $L_1$ norm adjustment

As mentioned earlier, the $L_1$ norm minimization is an estimation method that minimizes the weighted sum of absolute residuals. To modify the objective function (8) and the constraints (3), the usual derivation of the least squares ($L_2$) estimates will not work for $L_1$ estimates. In other words, an analog way to the well-known normal equations is not possible for $L_1$ estimation. To transform Eqs. (3) and (8) into something applicable, the usual strategy is to borrow a trick from linear programming and introduce slack variables, which guarantee non-negativity. This allows one to rewrite the objective function without absolute value signs.

To set up the $L_1$ estimation problem by a linear programming solution, we need to formulate a mathematical model where all variables, both parameters and residuals, are non-negative. The development begins with the familiar parametric Eq. (3), and is then transformed into an $L_1$ estimation problem by adding slack variables. To convert these equations into a form of non-negative parameters and non-negative residuals, we introduce one slack vector $\boldsymbol{\alpha}$ and one slack variable $\beta$ for the parameters, and two slack vectors $\mathbf{u}$ and $\mathbf{v}$ for the residuals. The parameters as well as the residuals, in general, can become either positive or negative. Therefore, we can replace the unknown parameters and the residuals by (Marshall & Bethel ,1996 ; Amiri-Simkooei,2003)

$$\mathbf{x} = \boldsymbol{\alpha} - \boldsymbol{\beta} \ \equiv \ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} - \begin{bmatrix} \beta \\ \beta \\ \vdots \\ \beta \end{bmatrix}, \ \text{where} \ \boldsymbol{\alpha}, \boldsymbol{\beta} \geq \mathbf{0}$$

$$\mathbf{e} = \mathbf{u} - \mathbf{v} \ \equiv \ \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_m \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{bmatrix} - \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{bmatrix}, \ \text{where} \ \mathbf{u}, \mathbf{v} \geq \mathbf{0} \tag{14}$$

As can be seen, the number of unknown parameters and residuals is increased from $n$ to $n+1$, and from $m$ to $2m$, respectively. These will cause some singularities on the problem, i.e., the LP problem will give an infinite number of solutions (alternative optimal solutions). That is, if $e_j = u_j - v_j$ is a solution to the above system, then $e_j = (u_j + c) - (v_j + c)$ with $c > 0$ will be a solution as well. To overcome the singularity induced by the re-parameterization and to obtain a unique optimal solution for the residuals, the objective function (8) can be rewritten in terms of slack variables as

$$z = \mathbf{w}^T|\mathbf{e}| = \mathbf{w}^T|\mathbf{u} - \mathbf{v}| = \mathbf{w}^T(\mathbf{u} + \mathbf{v}) \rightarrow \min$$
$$\text{where} \quad u_i = 0 \ \text{ or } \ v_i = 0 \tag{15}$$

which is valid when we minimize the residuals. In this case, one of the components of $e_i$, either $u_i$ or $v_i$, becomes zero. To obtain a unique solution for the slack variable of the unknown parameters, one can minimize $\beta \rightarrow \min$. Because of the above objective functions are both positive, they can be lumped together to one objective function as

$$z = \mathbf{w}^T(\mathbf{u}+\mathbf{v}) + \beta \rightarrow \min \qquad (16)$$

If we closely, not exactly, follow (Amiri-Simkooei, 2003), the original parametric equation and datum constraints (3) can be rewritten, in terms of the slack vectors $\mathbf{u}$, $\mathbf{v}$, and $\boldsymbol{\alpha}$ as well as the slack variable $\beta$, as

$$z = \mathbf{c}^T\mathbf{z} \rightarrow \min, \text{ subject to } \mathbf{H}\mathbf{z} = \mathbf{b},\ \mathbf{z} \geq \mathbf{0} \qquad (17)$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{A} & -\mathbf{a} & \mathbf{I} & -\mathbf{I} \\ \mathbf{D} & -\mathbf{d} & \mathbf{O} & \mathbf{O} \end{bmatrix},\quad \mathbf{b} = \begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix}$$
$$\mathbf{z} = \begin{bmatrix} \boldsymbol{\alpha} \\ \beta \\ \mathbf{v} \\ \mathbf{u} \end{bmatrix},\quad \mathbf{c} = \begin{bmatrix} \mathbf{0} \\ 1 \\ \mathbf{w} \\ \mathbf{w} \end{bmatrix} \qquad (18)$$

where $\mathbf{O}$ is a zero matrix and the $m \times 1$ vector $\mathbf{a}$ and the $d \times 1$ vector $\mathbf{d}$ are the sum of columns of matrices $\mathbf{A}$ and $\mathbf{D}$, respectively

$$a_i = \sum_{j=1}^{n} a_{ij},\ \ i = 1,...,m$$
$$d_i = \sum_{j=1}^{n} d_{ij},\ \ i = 1,...,d \qquad (19)$$

The preceding equations are in fact of smaller size than those presented by (Amiri-Simkooei,2003). They are in the form of Eqs. (12) and (13), and therefore, a special problem of the operations research that can be solved by linear programming. Solving for vector $\mathbf{z}$ will yield vectors $\boldsymbol{\alpha}$, $\mathbf{u}$, $\mathbf{v}$ and variable $\beta$. Consequently, the solution vector $\mathbf{x} = \boldsymbol{\alpha} - \beta\mathbf{1}$ and the residual vector $\mathbf{e} = \mathbf{u} - \mathbf{v}$ can simply be obtained. When the functional model $\mathbf{A}(\mathbf{x})$ is non-linear, one can rely on the Taylor series expansion to linearize the model. In a similar manner to what we have in the standard least-squares, the procedure should be solved through iterations until the corrections to the solution vector $\mathbf{x}$ become zeros.

### 5. Alternative method for $L_1$ norm adjustment

We believe that the lack of attention paid to the $L_1$ norm adjustment in geodetic networks has been mainly due to the relative complexity of its implementation compared to the least squares. Although this complexity should not be of great importance in the presence of modern computing systems, in the present paper a simple technique will be reintroduced. The method was originally introduced by (Krarup et al.,1980) into geodetic applications. (Chen,1983 ; Secord,1985) used a similar technique by minimizing the L1 norm of the displacement vector to detect the stable reference points. For the $L_1$ norm adjustment, we have used the least squares residuals in an iterative procedure to reweight the observations. In this method, the weight matrix $\mathbf{W}$ in Eq. (4) has been modified in the ($k$+1)th iteration as follows:

$$\mathbf{W}^{(k+1)} = \text{diag}\left(\mathbf{W}^0(i,i) / |\, e_i^{(k)} \,|\right) \qquad (20)$$

where $\mathbf{W}^0(i,i)$ is the initial weight of the $i$th observation, and $e_i^{(k)}$ is the $i$th component of the estimated vector $\hat{\mathbf{e}}$ from Eq. (5) in the $k$th iteration. Modifying the weight matrix $\mathbf{W}$ in accordance with Eq. (20) and substituting into Eq. (4) will give a modified $\hat{\mathbf{x}}$ and subsequently a modified residual vector $\hat{\mathbf{e}}$. The iterative procedure continues until the absolute differences between the successive estimated residual components are smaller than a tolerance $\varepsilon$. That is, the iteration will stop after convergence to a constant residual vector $\hat{\mathbf{e}}$. During this procedure some $e_i^{(k)}$ will approach zero causing numerical instabilities because $1/e_i^{(k)}$ becomes very large. There is an approach to handle this problem. The expression can be replaced (20) by

$$\mathbf{W}^{(k+1)} = \text{diag}\left(\mathbf{W}^0(i,i) / (|\, e_i^{(k)} \,| + \delta)\right) \qquad (21)$$

where $\delta$ is a small positive number that precludes the problem from singularity. The above procedure provides a least squares adjustment on the observations where a set of the observations have very large weights compared to the others. In practice, this means that the adjustment has been conducted by only such minimum observations. This is the interpretation of the $L_1$ norm adjustment.

Authors' practical tests with this technique show that this process can be expected to converge to the $L_1$ norm minimization (although the theoretical reason for this is unknown). However, sometimes the convergence speed of this method is very low. In order to increase the convergence speed, after applying Eq. (21) for a few iterations, it is recommended to substitute the term $|\, e_i^{(k)} \,|$ by $|\, e_i^{(k)} \,|^2$ in Eq. (21). In section 7, two practical examples (both linear and non-linear models) have been used to test the efficiency of this technique.

## 6. Tests of individual residuals of $L_1$ and $L_2$ norms

The least squares estimation of residual vector and its covariance matrix can be obtained from Eqs. (5) and (6),

respectively. For data snooping, the null hypothesis $H_0$ for individual residuals is as follows (Baarda,1968):

$$H_0: \ \hat{e}_i \ \sim \ N(0, \sigma^2_{\hat{e}_i}), \ \text{where} \ \sigma^2_{\hat{e}_i} = (\mathbf{Q}_{\hat{e}})_{ii} \qquad (22)$$

The statistics chosen for testing the null hypothesis $H_0$ is as follows:

$$w_i = \frac{\hat{e}_i}{\sigma_{\hat{e}_i}} \qquad (23)$$

which under the null hypothesis $H_0$ is normally distributed with a zero mean and unit variance. Therefore, the standardized residual is distributed as

$$w_i / H_0 \sim \ N(0, 1) \qquad (24)$$

According to the principle of two-tailed test approach, given the significance level $\alpha_0$, $H_0$ is rejected if the following condition holds:

$$|w_i| > w_{\alpha_0/2} \equiv -w_{1-\alpha_0/2} \qquad (25)$$

where $w_{\alpha_0/2}$ and $w_{1-\alpha_0/2}$ are the upper and lower boundary values of the test statistics $w_i$ calculated from the standard normal distribution under the given significance level $\alpha_0$.

Data snooping can also be performed with the residuals from the $L_1$ norm adjustment. For this purpose, the $L_1$ norm residuals and the covariance matrix should be obtained. As mentioned before, the $L_1$ norm adjustment divides the observations into two groups, namely $\mathbf{y_1}$ and $\mathbf{y_2}$, where $\mathbf{y_1}$ is a special sufficient subset of functionally independent observations, which gives a unique solution, and $\mathbf{y_2}$ is the other part of $\mathbf{y}$. After a few simple operations, for the $L_1$ estimator, the solution can be obtained as

$$\hat{\mathbf{x}} = (\mathbf{A_1^T A_1} + \mathbf{D^T D})^{-1} \mathbf{A_1^T y_1}, \ \text{and} \ \mathbf{e_1} = \mathbf{0} \qquad (26)$$

where $\mathbf{A_1}$ is composed of the rows of $\mathbf{A}$ corresponding to the observations in $\mathbf{y_1}$. The residuals of the second part of the observations can be obtained using the following equation:

$$\hat{\mathbf{e}}_2 = \mathbf{A_2}\hat{\mathbf{x}} - \mathbf{y_2} = \mathbf{A_2}(\mathbf{A_1^T A_1} + \mathbf{D^T D})^{-1} \mathbf{A_1^T y_1} - \mathbf{y_2} \qquad (27)$$

where $\mathbf{A_2}$ is the second part of matrix $\mathbf{A}$ (other than $\mathbf{A_1}$). Applying the covariance propagation law to the preceding equation gives the covariance matrix of the residuals

$$\mathbf{Q}_{\hat{e}_2} = \mathbf{Q}_{y_2} + \mathbf{A_2}(\mathbf{A_1^T A_1} + \mathbf{D^T D})^{-1} \mathbf{A_1^T Q}_{y_1} \mathbf{A}_1 (\mathbf{A_1^T A_1} + \mathbf{D^T D})^{-1} \mathbf{A_2^T} \qquad (28)$$

where $\mathbf{Q}_{y_1}$ and $\mathbf{Q}_{y_2}$ are the covariance matrices of $\mathbf{y_1}$ and $\mathbf{y_2}$, respectively. For data snooping, the same tests as the least squares method can be used to detect the blunders in the observations.

## 7. Numerical results and discussions

For verification and comparison of the suggested formulation and the alternative technique, the results have been tested on a few simulated and real geodetic networks by the authors. Since presenting all of the examples are beyond the scope of the present study, we will consider only two examples. The first example is a linear model while the second one is a non-linear model.

### 7.1 Example 1

In the first example, a leveling network is assumed. The network consists of 6 points ($P_1$, $P_2$, $P_3$, $P_4$, $P_5$ and $P_6$) with 9 height difference observations. The degree of freedom of the network is $df = 4$. The datum of the network is provided by inner constraints. Therefore, $\mathbf{D} = [1 \ 1 \ 1 \ 1 \ 1 \ 1]$. Table 1 gives the list of the observations (columns 1 to 3) and their observed values (column 4). The precision of these observations is 1 mm. The network is evaluated in the following three steps:

In the first step, the initial observations are adjusted by the L2 norm (least squares) adjustment. Columns 5 and 6 give the results of computations for the estimated residuals and the normalized residuals, respectively. If we use a 5% level of significance then the normalized residuals should fall within the range –1.96 to +1.96. As can be seen from column 6, no normalized residual exceeds the critical value 1.96 which means that no observation has a blunder.

In the second step, it is assumed that the observed height differences No. 1 and 9, marked with asterisks, are erroneous by 1 cm. The erroneous observations are adjusted by the $L_2$ norm (least squares) criterion. Columns 7 and 8 give the results of the estimated residuals and the normalized ones, respectively. As can be seen from column 8, seven normalized residuals exceed the critical value 1.96 that may mean these observations have blunders, whereas we know only two observations are erroneous by 1 cm. For correct detection of these blunders, one may use the data snooping technique that was proposed by (Baarda,1968 and Pope ,1976, and Kok,1984). However, one of the goals of the present paper is the comparison of the $L_1$ and $L_2$ norms for outlier detection in their classical form. In the third step, the same observations as in the second step are assumed. The erroneous observations are adjusted by the $L_1$ norm criterion using two techniques, namely the linear programming (LP) formulation and the simple technique based on the least

squares residuals (LS). Columns 9 and 10 give the results of the $L_1$ norm residuals for the LS and the LP techniques, respectively. As can be seen, the results are neither exactly the same nor depart significantly from each other. This indicates that the solution is not unique as the absolute sum of the residuals is identical for both techniques (i.e., 20.66 mm). This is not a surprise, since the solution of the $L_1$ norm minimization is not guaranteed to be unique. Column 11 gives the standardized $L_1$ norm residuals (based on the LS). Only two normalized residuals (the first and ninth observations) exceed the critical value 1.96 that means these two observations have blunders. This means that the $L_1$ norm illustrates the blunders more straightforward than the $L_2$ norm. We believe that the lack of uniqueness of the $L_1$ norm solution seems not to destroy the efficiency of the technique for outlier detection.

**7.2 Example 2**

In the second example, a trilateration network is assumed. The network consists of 8 points with 28 distance observations. The degree of freedom of the network is $df = 15$. Columns 1, 2, and 3 of Table 2 give the list of observations. Column 4 gives the observations. The precision of observations is 1 mm. The datum of the network is provided by inner constraints. Again, the network is adjusted in three steps: In the first step, the initial observations are adjusted by the L2 norm (least squares) adjustment. Columns 5 and 6 give the estimated residuals and the normalized ones, respectively. Again, if we use a 5% level of significance, the normalized residuals should fall within the range –1.96 to +1.96. As can be seen, no normalized residual exceeds the critical value 1.96, which means that there is no blunder in the observations. In the second step, it is assumed that the observations No. 1 and 28 are erroneous by 1 cm and the observations No. 14 and 23 are erroneous by -1 cm. The erroneous observations are adjusted by $L_2$ norm (least squares) criterion. Columns 7 and 8 give the results. In this case, fifteen normalized residuals exceed the critical value 1.96 that might imply these observations have blunders.

Table 1. Heights difference observations as well as $L_1$ and $L_2$ norm standardized residuals in example 1.

| Obs. No. | Height Difference | | Observed Values | ($L_2$ Norm) Residuals | | ($L_2$ Norm) Residuals | | ($L_1$ Norm) Residuals | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | From | To | (m) | I. [a] | SR [e] | E. [b] | SR [e] | LS [c] | LP [d] | SR [e] |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) |
| 1 * | 1 | 2 | 1.2484 | 0.690 | 1.03 | -4.87 | -7.30 | **-7.53** | **-8.06** | **-3.76** |
| 2 | 2 | 3 | 3.8099 | 0.530 | 0.79 | -3.92 | -5.87 | -0.56 | 0 | -0.23 |
| 3 | 3 | 4 | -7.0507 | 0.720 | 1.08 | 1.83 | 2.75 | 0 | 0 | 0 |
| 4 | 4 | 5 | 11.5890 | 0.750 | 1.12 | 2.97 | 4.46 | 2.40 | 1.84 | 1.20 |
| 5 | 5 | 6 | -2.6912 | 0.910 | 1.36 | 2.02 | 3.03 | 0 | 0.53 | 0 |
| 6 | 6 | 1 | -6.9097 | 0.710 | 1.07 | -3.73 | -5.59 | 0 | 0 | 0 |
| 7 | 1 | 4 | -1.9904 | 0.030 | 0.04 | 1.14 | 1.71 | 0 | 0.03 | 0 |
| 8 | 2 | 5 | 8.3501 | 0.160 | 0.24 | -0.95 | -1.43 | 0 | 0 | 0 |
| 9 * | 3 | 6 | 1.8496 | -0.19 | -0.29 | -5.75 | -8.62 | **-10.17** | **-10.20** | **-5.09** |
| **Absolute Sum:** | | | -- | -- | -- | -- | -- | 20.66 | 20.66 | -- |

[a] I. = $L_2$ norm residuals of initial observations (mm)

[b] E. = $L_2$ norm residuals of erroneous observations (mm)

[c] LS = $L_1$ norm residuals using the technique of least squares residuals (mm)

[d] LP = $L_1$ norm residuals using the technique of linear programming (mm)

[e] SR = Standardized residuals

However, we know that this is not the case because only four observations have blunders. In the third step, the same observations as in the second step are assumed. The erroneous observations are now adjusted by the $L_1$ norm criterion using two techniques, namely the linear programming (LP) formulation and the simple technique based on the least squares residuals (LS). Columns 9 and 10 give the estimated residuals of the LS and LP techniques, respectively. Here, the results are almost the same. The slight difference for the residuals is due to the use of the LS method in which we needed quite a large number of iterations before we could obtain acceptable results. This can be considered as a drawback of this simple method. Column 11 gives the standardized $L_1$ norm residuals. It can be seen that only four normalized residuals (No. 1, 14, 23, and 28) exceed the critical value 1.96 that means these observations have gross errors. This implies that the $L_1$ norm residuals are more robust than the $L_2$ norm residuals. Therefore, the superior

method for the outlier detection is based on the $L_1$ norm minimization.

## 8. Concluding remarks

There are many procedures for adjusting the data and detecting the presence of blunders in geodetic observations. Most of such techniques involve minimizing the $L_2$ norm of the residuals (least squares technique), which are referred to as 'data snooping techniques'. There is a robust procedure called the $L_1$ norm minimization, which minimizes the absolute sum of the weighted residuals. This is a powerful technique used in geodetic networks to detect the blunders in observations. In this paper, the formulation and implementation of the $L_1$ norm minimization were presented for rank deficient Gauss-Markov models that leads to solving a linear programming problem. Then, a simple technique was also used based on the least squares residuals. For verification of the given formulation and the simple technique, two examples were presented on both linear and non-linear models. The results showed that the $L_1$ norm minimization was more sensitive than the least squares for outlier detection. The $L_1$ norm minimization was more efficient than the $L_2$ norm when we had more blunders in the observations (or equivalently when the redundancy of the model was low). One disadvantage of the $L_1$ norm adjustment is that the unique solution is not guaranteed. This, however, seems not to be a serious problem for outlier detection (see example 1). It seems that the lack of attention paid to the $L_1$ norm minimization in geodetic applications is mainly due to the relative complexity of its implementation compared to the least squares. Although this complexity

Table 2. Distance observations as well as $L_1$ and $L_2$ norm standardized residuals in example 2.

| Obs. No. | Measured Distances | | Observed Values | ($L_2$ Norm) Residuals | | ($L_2$ Norm) Residuals | | ($L_1$ Norm) Residuals | | |
|---|---|---|---|---|---|---|---|---|---|---|
| (1) | From (2) | To (3) | (m) (4) | I. (5) | SR (6) | E. (7) | SR (8) | LS (9) | LP (10) | SR (11) |
| 1 * | 1 | 2 | 837.5544 | -0.81 | -1.09 | -6.50 | -8.72 | **-9.97** | **-9.96** | **-4.76** |
| 2 | 1 | 3 | 1880.0688 | 0.64 | 0.88 | 2.33 | 3.22 | 0.00 | 0.01 | 0.00 |
| 3 | 1 | 4 | 1177.0744 | -0.68 | -1.14 | -1.85 | -3.11 | -0.44 | -0.45 | -0.21 |
| 4 | 1 | 5 | 1485.0330 | 0.87 | 1.17 | 2.03 | 2.73 | 0.21 | 0.21 | 0.07 |
| 5 | 1 | 6 | 2250.6640 | -0.60 | -0.73 | 0.07 | 0.08 | 0.00 | 0.00 | 0.00 |
| 6 | 1 | 7 | 688.6537 | 0.24 | 0.33 | -0.85 | -1.20 | 0.00 | 0.00 | 0.00 |
| 7 | 1 | 8 | 1482.1634 | 0.14 | 0.18 | 3.57 | 4.59 | 0.68 | 0.69 | 0.40 |
| 8 | 2 | 3 | 1043.2772 | -0.29 | -0.39 | -2.80 | -3.78 | -1.74 | -1.74 | -0.85 |
| 9 | 2 | 4 | 1350.9739 | -0.35 | -0.48 | -0.64 | -0.88 | 0.00 | 0.00 | 0.00 |
| 10 | 2 | 5 | 1143.9659 | -0.32 | -0.46 | -0.90 | -1.32 | 0.00 | 0.00 | 0.00 |
| 11 | 2 | 6 | 1537.8888 | 0.18 | 0.23 | -1.18 | -1.55 | 0.21 | 0.21 | 0.12 |
| 12 | 2 | 7 | 794.9992 | 0.89 | 1.23 | 3.65 | 5.08 | 2.27 | 2.27 | 1.30 |
| 13 | 2 | 8 | 707.8153 | -0.28 | -0.37 | -0.78 | -1.06 | 0.00 | 0.00 | 0.00 |
| 14 * | 3 | 4 | 2047.7722 | 0.00 | 0.00 | 6.48 | 8.02 | **8.53** | **8.53** | **5.17** |
| 15 | 3 | 5 | 1420.7470 | -0.25 | -0.34 | -3.89 | -5.30 | 0.00 | 0.00 | 0.00 |
| 16 | 3 | 6 | 923.0041 | 0.30 | 0.58 | 1.28 | 2.55 | 0.00 | 0.00 | 0.00 |
| 17 | 3 | 7 | 1648.1903 | 0.00 | 0.00 | 0.78 | 1.00 | 0.00 | 0.00 | 0.00 |
| 18 | 3 | 8 | 553.9795 | -0.17 | -0.22 | -3.69 | -4.71 | -0.58 | -0.58 | -0.30 |
| 19 | 4 | 5 | 798.2518 | -1.07 | -1.47 | -1.18 | -1.62 | -3.25 | -3.25 | -1.10 |
| 20 | 4 | 6 | 1884.7847 | 0.16 | 0.23 | -4.70 | -6.61 | 0.00 | 0.00 | 0.00 |
| 21 | 4 | 7 | 584.0447 | 0.50 | 0.69 | -1.12 | -1.56 | 0.00 | 0.00 | 0.00 |
| 22 | 4 | 8 | 1495.0593 | 0.99 | 1.27 | 0.93 | 1.20 | 0.00 | 0.00 | 0.00 |
| 23 * | 5 | 6 | 1089.7829 | 1.20 | 1.66 | 6.26 | 8.68 | **13.22** | **13.22** | **5.05** |
| 24 | 5 | 7 | 823.3101 | 1.13 | 1.56 | 3.78 | 5.23 | 0.78 | 0.77 | 0.27 |
| 25 | 5 | 8 | 890.9380 | -1.31 | -1.88 | -1.35 | -1.94 | -1.15 | -1.15 | -0.55 |
| 26 | 6 | 7 | 1744.3977 | -1.38 | -1.78 | -1.03 | -1.33 | 0.00 | 0.00 | 0.00 |
| 27 | 6 | 8 | 837.4764 | 0.49 | 0.69 | 0.19 | 0.27 | 0.00 | 0.00 | 0.00 |
| 28 * | 7 | 8 | 1128.5570 | 0.17 | 0.21 | -5.82 | -7.50 | **-8.98** | **-8.98** | **-5.21** |
| **Absolute Sum:** | | | -- | -- | -- | -- | -- | **52.08** | **52.07** | -- |

should not be that important in the presence of modern computing techniques, in this paper, a simple technique was used, and its efficacy was tested on both linear and non-linear models. The method could give identical results to the standard linear programming (LP) formulation. However, this method can sometimes be time-consuming as it requires many numbers of iterations before obtaining a reasonable solution. On the other hand, for the LP formulation, there are standard procedures to solve a linear programming problem. This technique becomes more efficient if we rely on the new methods such as the interior point methods rather than the classical simplex method.

**References**

Amiri-Simkooei, A. (2003). Formulation of L 1 norm minimization in Gauss-Markov models. *Journal of Surveying Engineering*, *129*(1), 37-43.

Baarda, W. (1968). *A Testing procedijre for use in geodetic networks*. Publication on Geodesy, New series 2, No. 5, Netherlands Geodetic Commission, Delft.

Bazaraa, M. S., Jarvis, J. J., & Sherali, H. D. (2011). *Linear programming and network flows*. John Wiley & Sons.

Chen, Y. Q. (1984). *Analysis of deformation surveys-A generalized method*. Tech. Report No. 94, Dept. of Surveying Engineering, University of New Brunswick.

Dantzig, G. B. (1963). Linear Programming and Extensions. Princeton landmarks in mathematics and physics. Princeton University Press, Princeton, N. J., 1963.

Kok, J. J. (1984). *On data snooping and multiple outlier testing*. US Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service, Charting and Geodetic Services.

Krarup, T. (1980). Gotterdammerung over least squares adjustment. In *Proc. 14th Congress of the International Society of Photogrammetry* (Vol. 3, pp. 369-378).

Marshall, J., & Bethel, J. (1996). Basic concepts of L 1 norm minimization for surveying applications. *Journal of surveying engineering*, *122*(4), 168-179.

Mikhail, E. M., & Ackermann, F. E. (1976). Observations and least squares. IEP-A dun-Donnelley Publisher.

Pope, A. J. (1976). *The statistics of residuals and the detection of outliers* (No. NOS-65-NGS-1), NOAA Tech. Report NOS 65 NGS 1, National Geodetic Survey, Rockville, Md.

Roos, C., Terlaky, T., & Vial, J. P. (1997). *Theory and algorithms for linear optimization: an interior point approach*. John Wiley & Son Ltd.

Secord, J. M. (1985). *Implementation of a generalized method for the analysis of deformation surveys*. Department of Surveying Engineering, University of New Brunswick.