GROSS ERROR DETECTION AND CONVERGENCE ANALYSIS IN PHOTOGRAMMETRIC NETWORKS

John Marshall
Geomatics Engineering
School of Civil Engr.
Purdue University
U.S.A.
marshalj@ecn.purdue.edu

James Bethel
Geomatics Engineering
School of Civil Engr.
Purdue University
U.S.A.
bethel@ecn.purdue.edu

Commission III, Working Group 1

KEY WORDS: Statistics, Adjustment, Design, Robust Estimation.

ABSTRACT

Robust methods of parameter estimation are often employed in multivariate applications where gross errors affect the data, however robust methods commonly lack pre- and post-analysis measures enjoyed by least squares estimation. In terms of pre-analysis, we describe a mathematically rigorous method for determining redundancy numbers based on L_1-norm minimization and for post-analysis we outline an L_1-norm-based method for detecting gross errors in photogrammetric observations. Additionally, we describe a graphical method for interpreting the convergence robustness of nonlinear parameter estimators and apply this method to single photo resection.

1 INTRODUCTION

Robust methods of gross error detection are sometimes viewed in a negative light by scientists because they lack the statistical rigor which is enjoyed by least squares. Further, robust methods are associated with high computational expense as well as a lack of pre- and post-analysis measures which are commonly used to design and analyze photogrammetric networks. In response to some of these criticisms, we show (1) that "L_1-norm redundancy numbers" exist which closely resemble those previously derived for least squares methods, (2) the nonlinear parameter convergence rates of two robust methods of estimation are superior to the conventional least squares method when gross errors corrupt the data, and (3) an outline for a rigorous hypothesis test for detecting gross errors by examining L_1-norm residuals.

2 MATH MODELS

Existing L_2-based methods of gross error detection effectively identify *small* gross errors (Baarda, 1968), (Pope, 1976), however these nonrobust methods are often incapable of accommodating extremely large gross errors which arise from a variety of sources. Consequently, alternative robust methods have been employed (Levenberg-Marquardt method, L_1-norm minimization) to handle situations when large gross errors corrupt the data. A mathematical description of these methods is provided in the following subsections.

2.1 L$_1$ norm

Boscovich developed L_1-norm minimization in 1757, however L_1-norm minimization has only received significant attention in recent years due to efficient computer algorithms (Barrodale and Roberts, 1973) and interest in robust estimation methods. The solution to L_1-norm problems is obtained by minimizing the following weighted L_1-norm objective function Φ_{L_1}

$$\Phi_{L_1} = \mathbf{s}^T|\mathbf{v}| \longrightarrow \text{minimum} \tag{1}$$

where $\mathbf{s}$ is a nx1 vector with entries $1/\sigma_i$, $\mathbf{v}$ is the residual vector, and $|\bullet|$ refers to the absolute value operator. Linear programming techniques are used to minimize the objective function for solutions with nonlinear parameters, the details of which are provided in Marshall (1998).

2.2 L$_2$ norm (Newton)

Contemporary photogrammetric problems are typically solved using the method of least squares with Newton-type corrections due to its favorable computation speed and computational simplicity. The well-known L_2-norm objective function is (Moffit and Mikhail, 1980)

$$\Phi_{L_2} = \mathbf{v}^T\mathbf{W}\mathbf{v} \longrightarrow \text{minimum} \tag{2}$$

where $\mathbf{W}$ is an nxn weight matrix. For estimation of nonlinear parameters, the objective function is minimized by employing the normal equations

$$(\mathbf{B}^T\mathbf{W}\mathbf{B})\mathbf{\Delta} = \mathbf{B}^T\mathbf{W}\mathbf{f} \tag{3}$$

until the corrections, $\mathbf{\Delta}$, to the nonlinear parameters are negligible

$$\mathbf{x}_{\text{new}} = \mathbf{x}_{\text{old}} + \mathbf{\Delta} \tag{4}$$

The least squares solution may fail to converge when approximate parameters are supplied which lie outside the vicinity of convergence or when large gross errors corrupt the data.

2.3 L$_2$ norm (Levenberg-Marquardt)

The Levenberg-Marquardt method is a technique for least squares estimation of nonlinear parameters and it minimizes the same objective function, Φ_{L_2} as the Newton method (equation (2)). The Levenberg-Marquardt method employs a computational strategy to minimize the objective function which combines desirable attributes from both the gradient and Newton methods. As Marquardt (1963) states, the strategy combines the gradient methods' ability to converge from an initial guess which may be outside the region of convergence for other methods with the efficiency of the Newton method.

3 REDUNDANCY NUMBERS

Redundancy numbers are commonly used for photogrammetric pre-analysis to determine each observation's contribution to the total network redundancy and to determine the relationship between the residuals and true errors. It is typically more difficult to detect gross errors in observations with small redundancy numbers. The well-known derivation for least squares redundancy numbers is founded on linear combinations of the residuals and the true errors, whereas L_1-norm redundancy numbers are probability based. We summarize the derivations and provide a numerical example. For the L_2 case, the relationship between the true errors, ε, and the residuals, $\mathbf{v}$, is

$$\mathbf{v} + \mathbf{B}\boldsymbol{\Delta}_{\mathrm{L.S.}} = \boldsymbol{\ell} - \mathbf{d} = \mathbf{f} \quad (5)$$

$$-\varepsilon + \mathbf{B}\boldsymbol{\Delta}_{\mathrm{true}} = \boldsymbol{\ell} - \mathbf{d} = \mathbf{f} \quad (6)$$

The least squares parameter estimates, $\boldsymbol{\Delta}_{\mathrm{L.S.}}$, are obtained using the normal equations

$$\boldsymbol{\Delta}_{\mathrm{L.S.}} = \left(\mathbf{B}^T\mathbf{W}\mathbf{B}\right)^{-1}\mathbf{B}^T\mathbf{W}\mathbf{f} \quad (7)$$

and the corresponding L_2 redundancy numbers arrive from the following linear relationship

$$\mathbf{v} = \left(-\mathbf{Q}_\ell + \mathbf{B}\left(\mathbf{B}^T\mathbf{W}\mathbf{B}\right)^{-1}\mathbf{B}^T\right)\mathbf{W}\varepsilon \quad (8)$$

$$\mathbf{v} = -\mathbf{Q}_v\mathbf{W}\varepsilon \quad (9)$$

$$\mathbf{v} = -\mathbf{R}\varepsilon \quad (10)$$

The diagonal elements of the redundancy matrix, $\mathbf{R}$, correspond to the redundancy numbers, r_{ii}, $(i = 1, \ldots, n)$. Two characteristics of the singular and idempotent matrix $\mathbf{R}$ are: 1) $0 \le r_{ii} < 1$ and 2) trace$(\mathbf{R}) = r$. Networks with homogeneously large redundancy numbers are desirable since large redundancy numbers indicate the true errors will be better approximated by the residuals.

As mentioned, L_1-norm redundancy numbers are probability-based quantities which are comparable measures to the L_2-norm redundancy numbers. Although we do not know of a formal mathematical proof, "L_1-norm redundancy numbers" are obtained by determining the probability that an L_1-norm residual is not equal to zero which is expressed by the following equation

$$r_j = 1 - P(V_j = 0) \quad (j = 1, \ldots, n) \quad (11)$$

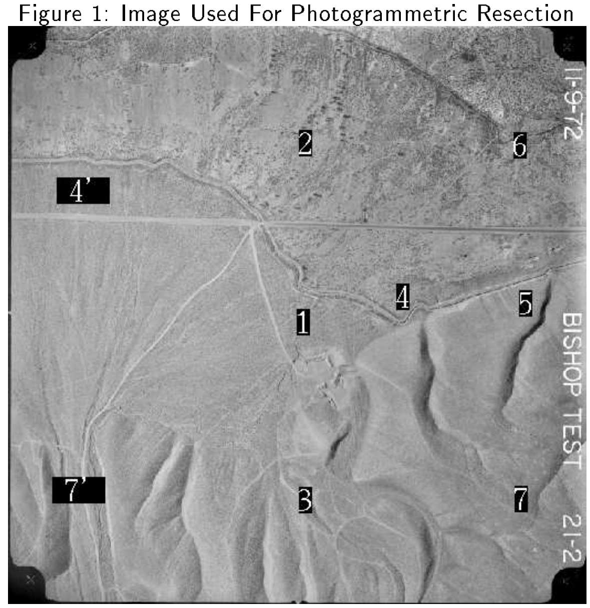
where $P(V_j = 0)$ is the probability that the jth residual is zero. Similar to the L_2-norm case, the sum of the individual redundancy numbers equals the total network redundancy or

$$\sum_{i=1}^{n} r_i = r \quad (12)$$

A comparison of L_1- and L_2-norm redundancy numbers for Figure 1 is provided in Table 1. We have made similar comparisons for land surveying trilateration networks and leveling networks, and in each case we find that the magnitude of L_1 and L_2 redundancy numbers tend parallel each other, however they are not identical since they are based on significantly different estimators.

The nominal parameters for the single image in Figure 1 are

Figure 1: Image Used For Photogrammetric Resection



ω (rad)	ϕ (rad)	κ (rad)
-1.64125457e-02	2.425529472e-03	1.600359980

X_L (m)	Y_L (m)	Z_L (m)	f (mm)
670760.583	223932.397	2233.001	152.0

Fixed Ground Stations			
Station	X (m)	Y (m)	Z (m)
1	670765.720	223897.990	1252.140
2	670319.290	223890.510	1240.610
3	671224.090	223939.800	1288.920
4	670750.020	224165.850	1247.520
5	670742.330	224444.720	1286.340
6	670306.790	224447.740	1235.060
7	671185.360	224399.430	1367.290

4 BREAKDOWN POINT

Breakdown point can be roughly described as a quantity assigned to an estimator that provides a measure of the estimator's quantitative robustness. Breakdown point indicates the fraction *observations* can be corrupted with gross errors before the estimator breaks down and yields unreliable results.

Breakdown point has been formally described by Hodges (1967) and Hampel (1971). Recently, the finite sample breakdown point was introduced by Donoho and Huber (1983) and it has received significant attention from others (Rousseeuw, 1983), (Bloomfield and Steiger, 1983), (Rousseeuw and Bassett, 1991). The mathematical description begins by denoting $\boldsymbol{\ell}$ as the nx1 vector of randomly perturbed observations and $\boldsymbol{\ell}_q$ as the nx1 vector of grossly contaminated observations. The vector $\boldsymbol{\ell}_q$ is generated by replacing any q of the original observations in $\boldsymbol{\ell}$ with arbitrary, contaminated observations. If T denotes any estimator, then let $||T(\boldsymbol{\ell}_q) - T(\boldsymbol{\ell})||$ denote the Euclidean distance between evaluations of the estimator at $\boldsymbol{\ell}_q$ and $\boldsymbol{\ell}$, respectively. Then when $\boldsymbol{\ell}_q$ is allowed to range over all contaminated data sets, $bias(q; T, \boldsymbol{\ell}) =$

Table 1: Redundancy Numbers and Observations for Photogrammetric Resection

Observation	Observations (mm)	L_1 norm Redundancy Number (unitless)	L_2 norm Redundancy Number (unitless)
x_1	-2.875	0.78	0.80
y_1	-1.080	0.65	0.64
x_2	-1.925	0.65	0.59
y_2	67.198	0.03	0.19
x_3	1.468	0.54	0.53
y_3	-75.180	0.55	0.43
x_4	38.675	0.87	0.85
y_4	0.124	0.80	0.80
x_5	85.554	0.81	0.75
y_5	0.059	0.74	0.72
x_6	83.606	0.45	0.48
y_6	66.831	0.34	0.39
x_7	83.084	0.35	0.40
y_7	-78.207	0.45	0.45
$\sum r_i$	n/a	8.00	8.00

$sup\|T(\boldsymbol{\ell}_q) - T(\boldsymbol{\ell})\|$ denotes the supremum or least upper bound of the estimator T for any values in the contaminated vector $\boldsymbol{\ell}_q$. The finite sample breakdown point, $\epsilon^*(T, \boldsymbol{\ell})$ of the estimator T is defined as,

$$\epsilon^*(T, \boldsymbol{\ell}) = min\left\{\frac{q}{n} \mid bias(q; T, \boldsymbol{\ell}_q) \text{ is infinite}\right\} \tag{13}$$

In simpler terms, the finite sample breakdown point is the smallest fraction of gross errors, q/n, which cause the estimator T to assume arbitrarily large values, i.e., values that are infinite. The finite sample breakdown point is illustrated in two univariate cases next.

The finite sample breakdown point for the sample mean is $\epsilon^* = 1/n$ because a single large gross error in the ith observation can cause the sample mean to approach infinity. By contrast, the sample median can tolerate large gross errors in about 50% of the observations without driving the sample median to infinity. Multivariate finite sample breakdown point is described next.

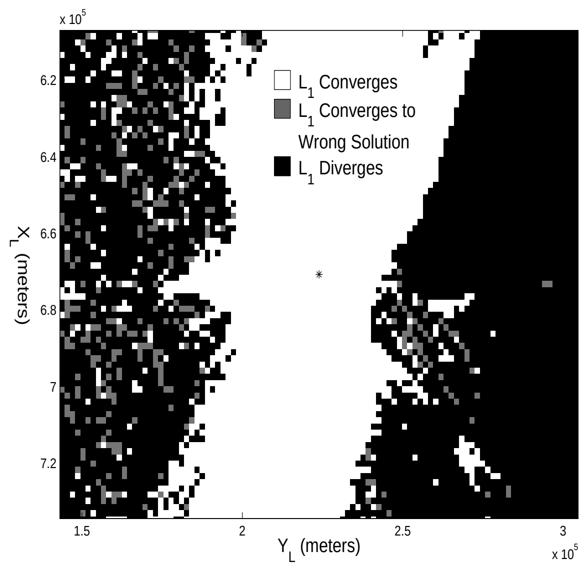
The finite sample breakdown point for L_1 regression has been examined in the literature (Bloomfield and Steiger, 1983), (Rousseeuw and Leroy, 1987) and this concept of breakdown point can be extended to photogrammetric networks. The finite sample breakdown point for L_1 regression can be described in two cases: (1) corruption only in the dependent variable, y, and (2) corruption in both the dependent and independent variable, y and x. The breakdown point for the first case is $\epsilon^* = 1/2$ like the sample median, whereas the breakdown point in the second case is $\epsilon^* = 1/n$ since a single gross error can have large effects on the L_1-norm parameter estimates (Donoho and Huber, 1983). The intuitive reason for the undesirable breakdown point in case (2) is that single points may be moved far away from the bulk of the points in the x direction causing the estimated regression line to pass through this point. Using a similar argument, a single point in a photogrammetric network can be moved far away from the bulk of the points to cause the L_1-norm estimator to break down.

5 CONVERGENCE REGIONS

It is well-known that estimation of nonlinear parameters requires a set of approximate parameters which are sufficiently close to the final parameter estimates, otherwise the solution will diverge. Consequently, another method for assessing the robustness of estimators is to examine the regions where the nonlinear solution is successful (converges) and where it fails (diverges). This method of analysis differs from breakdown point because here the approximate *parameters* are intentionally contaminated over a specific range, whereas the *observations* are contaminated in the breakdown point analysis.

The following examples are based on a single frame resection (Figure 1) using the observations in Table 1. In each of the examples, all six parameters $(\omega, \phi, \kappa, X_L, Y_L, Z_L)$ are estimated, however only the approximate values for the parameters X_L and Y_L are contaminated. The first example illustrates the convergence regions for the L_1, L_2, and Levenberg-Marquardt methods, respectively with no contamination to the observations (Figures 2, 3, 4). The blank (white) areas in these figures indicate the solution converged, the grey areas indicate the solution converged to a point distant from the nominal position (*) in the center of the figures, and black indicates the solution diverged. The figures associated with the first example show that the Levenberg-Marquardt method converges more often than either the L_1 or L_2 methods, indicating that it is less sensitive to corrupted approximate parameters for this example.

Figure 2: L_1-norm Convergence Region with No Gross Errors



The second example we illustrate uses the same mathematical model for single frame resection, however now three observations are perturbed as well. Specifically, points 4 and 7 in the image are moved to points 4' and 7' respectively, $(x'_4, y'_4) = (-87.676, 30.124), (x'_7, y'_7) = (-83.084, -78.207)$ (Figure 1). The impact of these three intentional gross errors causes the L_2-norm solution to diverge more often than it did without gross errors (Figure 6). The Levenberg-Marquardt and the L_1 norm, on the other hand, appear to converge more frequently than

Figure 3: L_2-Convergence Region with No Gross Errors

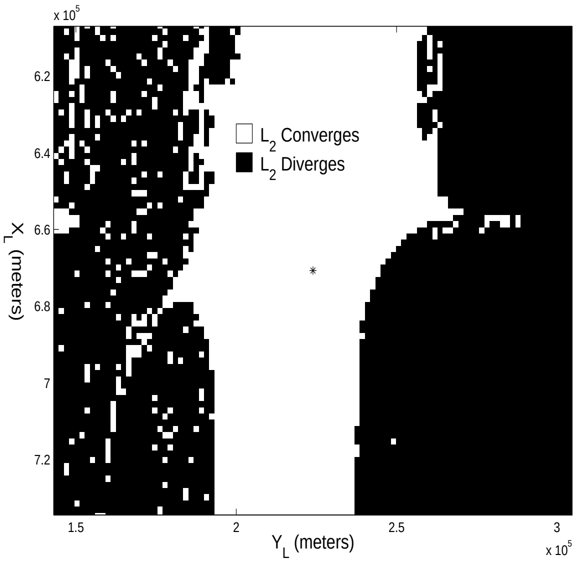
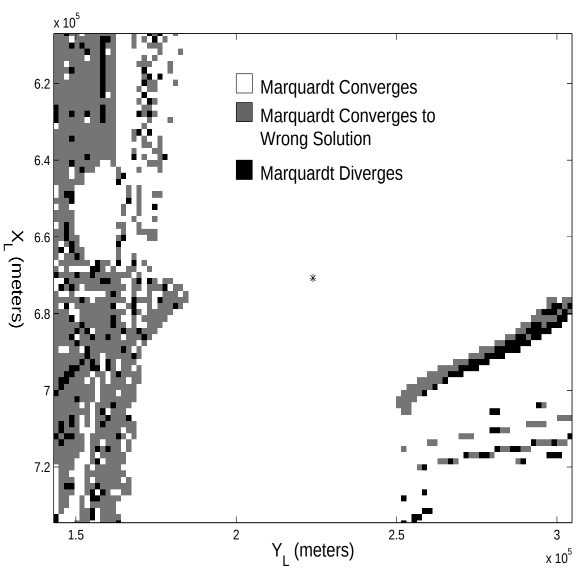


Figure 4: Levenberg-Marquardt Convergence Region with No Gross Errors



the L_2-Newton method. (Figures 7 and 5). Based on our limited studies, it can not be stated whether Levenberg-Marquardt outperforms L_1-norm minimization or vice versa. From the figures we show, the L_1 appears to converge to the correct solution more frequently than does the Levenberg-Marquardt method when large gross errors exist (Figures 7 and 5).

Figure 5: L_1-norm Convergence Region with Large Gross Error

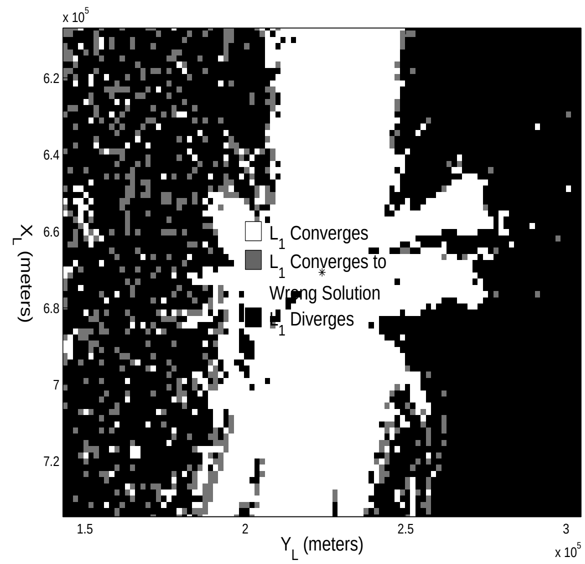
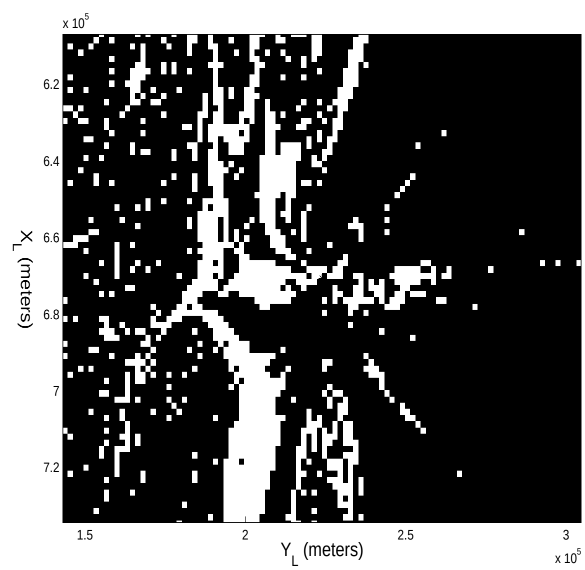


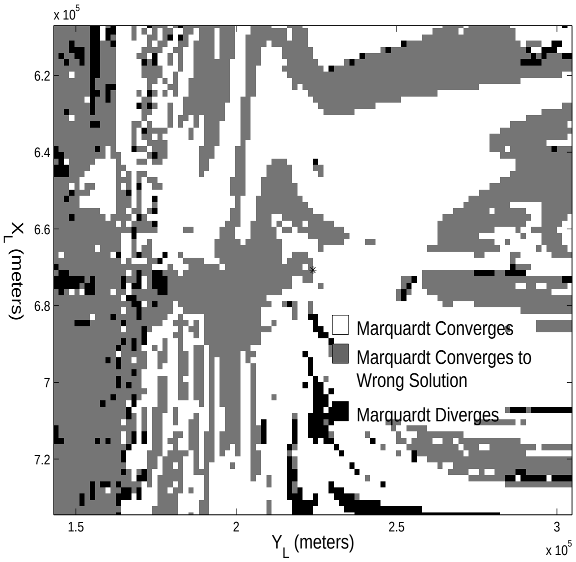
Figure 6: L_2-norm Convergence Region with Large Gross Error



6 HYPOTHESIS TESTING

Several existing L_2-based gross error detection methods exist for analyzing photogrammetric observations (Baarda, 1968), (Pope, 1976). These methods are mathematically rigorous and have proven themselves in numerous applications, however they are based on L_2-norm minimization

Figure 7: Levenberg-Marquardt Convergence Region with Large Gross Error



which is inherently nonrobust and the methods assume that the observations are normally distributed. These existing methods can be rigorously applied to either the L_2-Newton or L_2-Levenberg-Marquardt methods since the methods minimize the same objective function, Φ_{L_2}. L_1-norm gross error detection requires a different strategy which is described next.

It is well-known that the expected value of the L_2-norm residual vector is zero when the observations are normally distributed and when the observations are free of systematic errors, i.e., the gross error vector, $\boldsymbol{\nabla \ell}$, is zero. Likewise, the expected value of the L_1 residual vector is also zero. This fact leads to the L_1 multivariate hypothesis which is tested at the α significance level.

$$
\begin{array}{ll}
H_o : E(V_i) = 0 & H_a : E(V_i) \neq 0 \\
i = 1, \ldots, n & \text{for a single } i
\end{array}
\tag{14}
$$

The joint distribution of the n residuals is rather complicated and therefore the test in equation (14) is carried out in r (r = redundancy) simultaneous univariate tests. A total of r simultaneous tests take place since u of the L_1 residuals are zero. The $j = 1, \ldots, r$ univariate hypotheses take the form

Hypotheses

$$
\begin{array}{l}
H_{o_j} : E(V_j) = 0 \\
H_{a_j} : E(V_j) \neq 0
\end{array}
\tag{15}
$$

The specific details (i.e., test statistic, critical values, etc.) are being developed in current research. We believe that Bonferroni's inequality may provide desirable (large) simultaneous significance levels since testing occurs on r nonzero residuals.

7 CONCLUSIONS

This paper has described two tools which may simplify the use and interpretation of robust estimators in photogrammetric networks. First, a technique was described for analyzing the breakdown of nonlinear estimators though the use of convergence maps which graphically indicate where approximate parameters are inadequate for convergence. And secondly, a post-adjustment hypothesis test was outlined for detecting gross errors in photogrammetric networks by examining L_1-norm residuals. Together, these mathematical techniques begin to provide the necessary tools for rigorously applying robust estimation methods to photogrammetric networks.

REFERENCES

[1] Baarda, W. (1968). "A testing procedure for use in geodetic networks." *Neth. Geod. Comm., Publ. on Geod., New Series 2, No. 5.* Delft, The Netherlands.

[2] Barrodale, I. and Roberts, F. (1973). "An improved algorithm for discrete L_1 linear approximation." *SIAM Journal Numerical Analysis,*10(5), 839-848.

[3] Bloomfield, P. and Steiger W. (1983). *Least absolute deviations: theory, applications and algorithms.* Birkhäuser Boston, Inc., U.S.A.

[4] Donoho, D. and Huber, P. (1983). "The notion of breakdown point," in *A Festschrift for Erich Lehmann,* edited by P. Bickel, K. Doksum, and J.L. Hodges, JR., Belmont: Wadworth.

[5] Hampel F. (1971). "A general qualitative definition of robustness." *Ann. Math. Stat.* 42(6), 1887-1896.

[6] Hodges, J. (1967). "Efficiency in normal samples and tolerance of extreme values for some estimates of location." *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* Vol. 1, 163-186. Univ. of Calif. Press.

[7] Johnson R. and Wichern D. (1988). *Applied multivariate statistical analysis,* second edition. Prentice Hall.

[8] Marshall, J. (1998) "Preanalysis of survey networks by L_1-norm minimization." Proceeding of American Congress and Surveying and Mapping, Baltimore, MD.

[9] Moffit, F. and Mikhail, E. (1980). *Photogrammetry,* Third Edition, Harper and Row.

[10] Marquardt, D. (1963). "An algorithm for least-squares estimation of nonlinear parameters." *J. Soc. Indust. Appl. Math.* 11(2), 431-441.

[11] Pope, A., (1976). "The statistics of residuals and the detection of outliers." *NOAA Technical Report NOS 65 NGS 1,* Rockville, Md.

[12] Rousseeuw, P. (1983). "Multivariate estimation with high breakdown point," *Proc. of the 4th Pannonian Symp. on Math. Stat.* Edited by W. Grossmann, G. Pflug, I. Vincze, and W. Wertz. D. Reidel Publishing Co.

[13] Rousseeuw, P. and Bassett, G. (1991). "Robustness of the p-subset algorithm for regression with high breakdown point," in *Directions in Robust Statistics and Diagnostics, Part II.* Edited by W. Stahel and S. Weisberg. Springer-Verlag.

[14] Rousseeuw, P. and Leroy, A. (1987). *Robust regression and outlier detection.* Wiley.