# ANALYSIS OF RESIDUALS FROM $L_1$ NORM ESTIMATION


John Marshall
Surveying Engineering
Purdue University
U.S.A.
marshalj@ce.ecn.purdue.edu

James Bethel
Surveying Engineering
Purdue University
U.S.A.
bethel@ce.ecn.purdue.edu



**Key Words**: Surveying, Statitistics, Adjustment, Design, Estimation.



**ABSTRACT**

The detection of erroneous observations is a challenging skill that requires the dedicated energies of photogrammetrists and surveyors on a day-to-day basis. This paper focuses on a robust estimator for its erroneous observation detection because robust estimators often perform satisfactorily in the face of observations with blunders. Many current blunder detection methods are based on least squares which is not a robust estimator. The robust estimator examined is $L_1$ norm minimization and some of its unique properties related to network adjustment are presented. The $L_1$ norm residual sampling distribution is described both theoretically and empirically to illustrate the foundation for statistical inference of erroneous observations. Network reliability issues are presented in light of unique properties surrounding the $L_1$ norm. This analysis has application to many estimation problems such as those from geodesy, photogrammetry, and, in particular, close-range photogrammetry.


## 1 INTRODUCTION

An important asset to any photogrammetric or survey engineer is the ability to accurately detect erroneous observations based on accepted statistical premises. Traditionally erroneous observations have been identified by examining least squares ($L_2$ norm) residuals (Baarda, 1968), (Pope, 1976), (Kavouras, 1982) and others. Least squares estimation maintains many useful properties when the error vector $\varepsilon$, is normally distributed with zero mean, $\varepsilon \sim \mathrm{N}(0, \sigma^2)$. However, in cases where erroneous observations exist and the relatively stringent normality assumptions surrounding the $L_2$ norm are violated, a robust estimator such as the $L_1$ norm may be better suited to deal with these departures from normality. In addition to effectively managing the departures from normality, the $L_1$ norm may serve as an additional means of ruling whether borderline observations should be removed from a network adjustment. Used in this capacity, the $L_1$ norm is especially appealing to an engineer given the nonscientific method of selecting the significance level for statistical inference. The strengths and weaknesses of $L_1$ norm blunder detection are illustrated using Monte Carlo simulation.

## 2 OBSERVATION WEIGHTS

Accurate observation weights are necessary in surveying and photogrammetric applications, particularly when the observations come from different sources. For $L_1$ estimation the definition of the weight itself does not change, but we incorporate the weights differently in the computations. The function to be minimized under the $L_1$ criterion is,

$$\mathbf{t}^{\mathbf{T}}|\mathbf{v}| \longrightarrow \text{minimum} \tag{1}$$

where $\mathbf{t}$ is comprised of the diagonal elements of $\mathbf{T}$,

$$\mathbf{T} = \begin{bmatrix} 1/\sigma_1 & & & 0 \\ & 1/\sigma_2 & & \\ & & \ddots & \\ 0 & & & 1/\sigma_n \end{bmatrix},$$

and $\mathbf{v}$ is the vector of residuals. In the actual computations we implement the weighting by premultiplying the matrix equations by $\mathbf{T}$, i.e.,

$$\mathbf{B\Delta} \approx \mathbf{f} \tag{2}$$

becomes

$$\mathbf{TB\Delta} \approx \mathbf{Tf}. \tag{3}$$

## 3 THEORETICAL SAMPLING DISTRIBUTION OF $\mathbf{L_1}$ RESIDUALS

### 3.1 Mathematical Model

The theoretical sampling distribution of $L_1$ norm residuals forms the foundation for statistical inference concerning erroneous observations and it is based on *a priori* knowledge of network geometry, observation variances, and the method of estimation. In the case of $L_1$ norm estimation, many new challenges arise for the engineer who has relied exclusively on the $L_2$ norm for his erroneous observation detection because of the differing mathematical models associated with the $L_1$ and $L_2$ norm estimators. In terms of residuals and residual sampling distributions, the two most significant differences between the $L_1$ and $L_2$ estimators are,

1. The $L_2$ norm residuals can be expressed in terms of *all* the observations, whereas the $L_1$ norm residuals are expressed in terms of *subsets* of observations.

2. For the $L_2$ norm, linear combinations of Gaussian random variables give rise to Gaussian combinations (Menke, 1989), whereas for the $L_1$ norm, the results are not strictly linear combinations.

For now attention is directed to item 1 above and partitioning the $L_1$ mathematical model,

$$\underset{n,1}{\mathbf{v}} + \underset{n,u}{\mathbf{B}}\,\underset{u,1}{\Delta} = \underset{n,1}{\mathbf{f}} \Rightarrow$$

$$\left[\begin{array}{c}\underset{u,1}{\mathbf{v}_1}\\ \hline \underset{r,1}{\mathbf{v}_2}\end{array}\right]\left[\begin{array}{c}\underset{u,u}{\mathbf{B}_1}\\ \hline \underset{r,u}{\mathbf{B}_2}\end{array}\right]\underset{u,1}{\Delta} = \left[\begin{array}{c}\underset{u,1}{\mathbf{f}_1}\\ \hline \underset{r,1}{\mathbf{f}_2}\end{array}\right] \qquad (4)$$

where

n is the number of observations

u is the number of unknown parameters

r = n - u = the number of redundant observations

**v** is the vector of estimated residuals

**B** is the coefficient matrix of the unknown parameters

$\Delta$ is the vector of estimated $L_1$ norm parameters

**f** is the vector of condition equations

Given a subset of observations which satisfies the $L_1$ norm criterion, the $L_1$ norm parameter estimates follow as,

$$\Delta = \mathbf{B}_1^{-1}\mathbf{f}_1 \qquad (5)$$

Consequently, the $L_1$ norm residuals can be computed as,

$$\mathbf{v}_2 = \mathbf{f}_2 - \mathbf{B}_2\mathbf{B}_1^{-1}\mathbf{f}_1 \qquad (6)$$

$$\mathbf{v}_1 = \mathbf{f}_1 - \mathbf{B}_1\mathbf{B}_1^{-1}\mathbf{f}_1 = 0 \qquad (7)$$

### 3.2 Modification of Simplex Algorithm

Since the $L_1$ residuals are computed from two distinct subsets of observations, attention is directed to one of several processes used to create the two subsets, namely the simplex algorithm. By examining the simplex algorithm one can examine the partitioning process which leads to the probability distribution associated with the $L_1$ norm residuals.

Taylor and Basset have defined the sampling distribution of the $L_1$ estimator $\Delta$ based on a modification of the simplex algorithm (see Taylor in Zarembka, 1974), (Bassett, 1973). One major contribution resulting from their studies was identifying the role of the coefficient matrix $\mathbf{B}_1$ in the sampling distribution of the $L_1$ norm parameters. As Taylor points out, a small change to the condition equation vector **f** leads to a change in the matrix $\mathbf{B}_1$. Therefore, this *random* behavior exhibited by $\mathbf{B}_1$ introduces a new level of complexity in the theoretical sampling distribution of the residuals. Recall that the $L_2$ residual sampling distribution is based on a single coefficient matrix **B**. In the $L_1$ case, the behavior of $\mathbf{B}_1$ must be understood prior to arriving at the sampling distributions of the $L_1$ parameters and residuals.

Once $\mathbf{B}_1$ is understood, then Basset demonstrates that the probability density function of the $L_1$ estimator $\Delta$ has a closed form (Pg. 133). However it is not useful in terms of practical application since a "small" network with n = 20 observations and u = 5 unknown parameters requires more than 750 million steps to obtain the theoretical sampling distribution of the $L_1$ estimator.

### 3.3 Sampling Distribution of $L_1$ Residuals

The $L_1$ residual sampling distribution exhibits a symmetric and continuous sampling distribution, however this distribution differs from distributions commonly used in photogrammetric applications because of the spike located in the center of the distribution. This distribution is constructed by taking a linear combination of the dependent variables **f** and $\Delta$,

$$\mathbf{v} + \mathbf{B}\Delta = \mathbf{f} \Rightarrow$$

$$\left[\begin{array}{c}\mathbf{v}_1\ \text{(spike)}\\ \hline \mathbf{v}_2\ \text{(bell)}\end{array}\right] = \left[\begin{array}{c}\mathbf{f}_1\\ \hline \mathbf{f}_2\end{array}\right] - \left[\begin{array}{c}\mathbf{B}_1\\ \hline \mathbf{B}_2\end{array}\right]\Delta \qquad (8)$$

where

**f** has a normal distribution

$\Delta$ has a nonnormal distribution

**B** is a matrix of constants

## 4 EMPIRICAL SAMPLING DISTRIBUTION OF $L_1$ RESIDUALS

### 4.1 Monte Carlo Simulation

In contrast to the theoretical sampling distribution of $L_1$ residuals, the empirical distribution is conceptually easy to construct and it supplies the engineer with a clear picture of residual behavior under $L_1$ norm estimation. The empirical sampling distribution described herein is exclusively constructed by Monte Carlo simulations. In this approach, normally distributed random perturbations are added to synthetic observations and $L_1$ norm residuals are computed. This process of perturbing observations and computing residuals is performed many thousands of times to ultimately construct sampling distributions which yield valuable insight into residual behavior. Normally distributed perturbations have been chosen because surveying and photogrammetric equipment produce errors which resemble normal distributions. As an illustration, the Monte Carlo method was carried out on the trilateration network in Figure 1 where the perturbation magnitude for all observations was $\sigma = 0.220$ m.

After 100,000 Monte Carlo simulations, the residual sampling distribution is illustrated with histograms in Figure 2. The solid vertical line ("spike" ) in the center of the sampling distribution indicates the occurrence of zero residuals from the $\mathbf{v}_1$ vector in Equation 4. Further study is needed to determine the relationship between spike height and network geometry.

### 4.2 Nonnormality of $L_1$ Residuals

The normality of the residual sampling distribution under $L_2$ estimation plays an important role in determining how tests are performed on erroneous observations. If the errors are independent and normally distributed, then the engineer can rely on traditional tests based on common distributions such as the Standard Normal, Student $t$, and $\chi^2$ for example. Unfortunately, the $L_1$ norm residual sampling distribution is not normal, as illustrated in Figure 3.

Figure 1: Trilateration Network Diagram

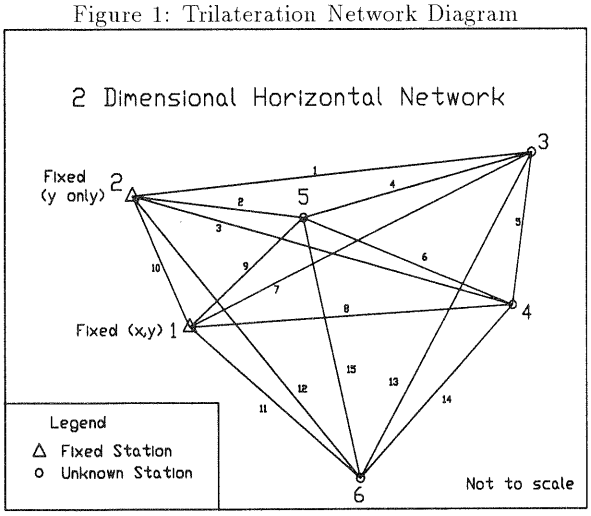



Figure 2: Empirical Residual Sampling Distribution for Figure 1

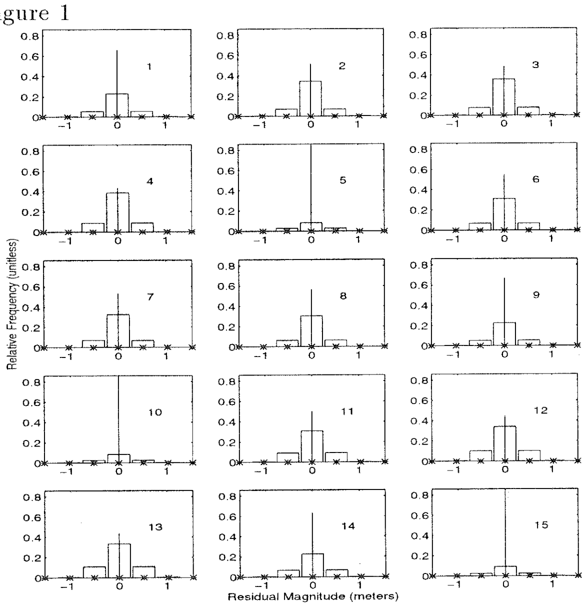
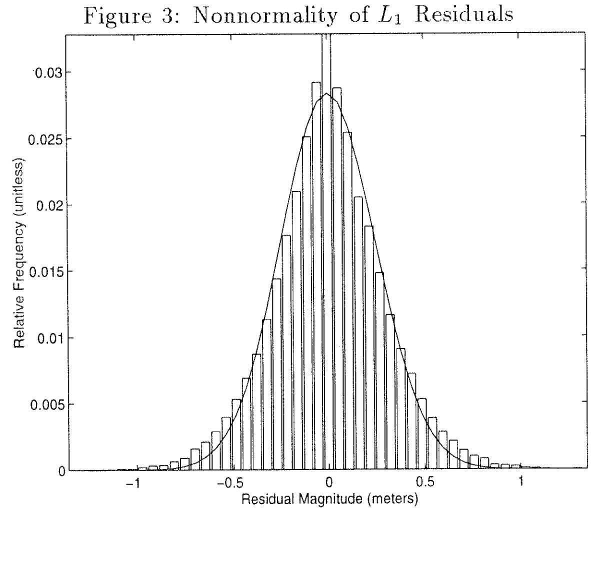



Figure 3: Nonnormality of $L_1$ Residuals




Normality was tested by fitting a Gaussian function to the Monte Carlo based sampling distribution and examining the fit. In this case, the empirical sampling distribution has fatter tails and a higher peak than the Gaussian distribution. The Gaussian curve is denoted by solid line in Figure 3. In addition to the high peak and fatter tails, the spike (which has been removed from Figure 3) also deviates sharply from the normality assumption. These departures from normality prevent the engineer from making statistical inference based on the common distributions mentioned previously. Instead, accurate statistical inference can be achieved using the Monte Carlo based sampling distributions described earlier in this section. This statistical inference may include critical value computation which is discussed in section 6. It should be emphasized that such an approach yields critical values which are problem dependent and even data dependent.

## 5 Preanalysis of Networks

### 5.1 Reliability

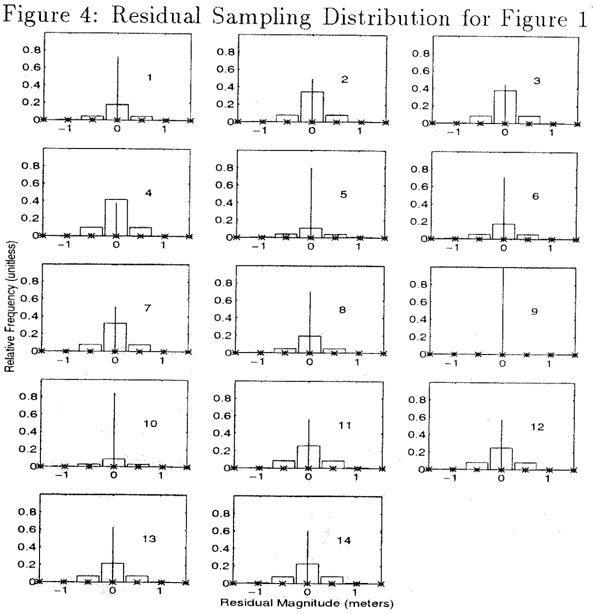
Preanalysis of networks is an important design tool used to evaluate network precision and sensitivity prior to executing fieldwork. Baarda has derived several reliability measures based on normally distributed random errors for the $L_2$ norm mathematical model (Baarda, 1968). Although counterparts to many of these $L_2$ reliability measures may exist under $L_1$, these derivations have not been completed at the time of this writing. In lieu of such derivations, the Monte Carlo method is used to evaluate network reliability. First, network internal reliability is examined to determine if a blunder can be detected under any circumstances, and secondly, network external reliability is examined to visualize the effects of a nonunique estimator on parameter estimates.

### 5.2 $L_1$ Internal Reliability

The underlying concept behind internal reliability is to determine how large an erroneous observation must be before it will be statistically detected through examination of the residuals. In this section, this concept is modified such that interest lies in detecting the erroneous observation under any conditions at all, statistically or otherwise. To examine the behavior of the $L_1$ norm under this scenario, the Monte Carlo method is used in conjunction with a geodetic trilateration network. A single observation is removed from the network to study its impact on blunder detection.

The initial trilateration network is presented in Figure 1 and has a total of 15 observations. After 100,000 Monte Carlo simulations, the sampling distribution of each residual is described by the histograms in Figure 2. Notice that the sampling distributions fluctuate in height and shape. The primary cause of these fluctuations is network geometry since the *a priori* $\sigma = 0.220$ m is identical for all observations as mentioned previously. The $L_2$ norm residual sampling distribution also experiences these fluctuations in height as a result of network geometry. However, when observation 15 is removed from the trilateration network and a new set of Monte Carlo simulations are computed, the residual sampling distribution changes substan-

Figure 4: Residual Sampling Distribution for Figure 1




tially as shown in Figure 4. The most significant change occurred in observation 9 which is illustrated by a "spike-only" distribution, indicating that residual 9 was equal to zero for all 100,000 simulations. Consequently, a blunder in observation 9 will never be found using $L_1$ norm estimation given the existing network geometry and the stated 14 observations. To remedy this deficiency, the network could be redesigned to incorporate new observations and/or new network stations into the network. Although the $L_1$ norm fails to identify an erroneous observation under this low redundancy scenario (redundancy = 14 - 9 = 5), the $L_1$ norm holds more promise at blunder detection given ample redundancy.

### 5.3 External Reliability

External reliability refers to the effect an erroneous observation has on the parameter estimates and has been studied by many authors (Baarda, 1968), (Mackenzie, 1985). Experimentation with external reliability and the $L_1$ norm indicates that in some instances an erroneous observation may have no effect on the $L_1$ parameter estimates at all. This unusual circumstance arises because $L_1$ norm minimization fails to provide unique parameter estimates in some cases. As an example, the simple trilateration network in Figure 5 was evaluated to illustrate the region where the $L_1$ norm has an infinite number of correct solutions. These solutions lie on the plane defined by ABCD in Figure 6. When examples such as these arise, the network could be redesigned to incorporate additional observations and/or stations which will reduce or eliminate the effects of the network deficiencies.

## 6 Statistical Evaluation of Residuals

Removing observations from a network must be well justified, especially given the extensive time, cost and labor involved in acquiring high quality observations. Therefore the purpose of this section is to determine whether an observation should be removed from a network based on the empirical sampling distribution presented in Section 4. The statistical test used to identify erroneous observations is based on Monte Carlo generated distributions for the null hypothesis.

Figure 5: Trilateration Network II

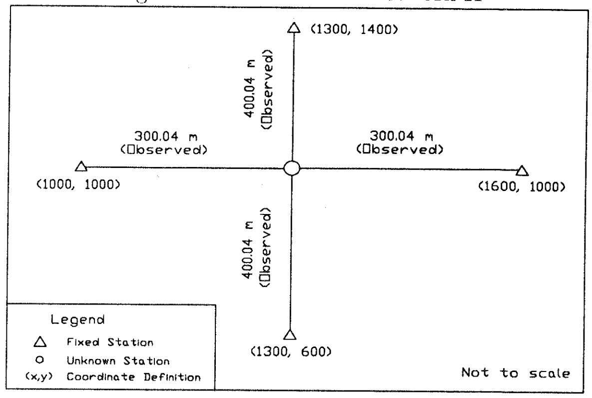
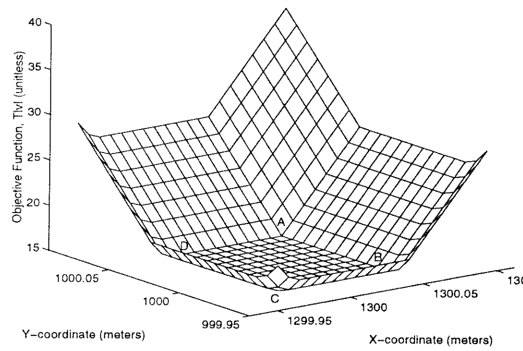



Figure 6: $L_1$ Objective Function Corresponding to Figure 5




### 6.1 Tests on Individual Residuals

In this section statistical inference is based on a technique of simple exponential curve fitting to the empirical sampling distribution of each residual. Using the area under the exponential curve, a straightforward computation of critical values results. The residual sampling distribution displayed in Figure 3 is symmetric and contains a sharp change in grade at the apex of the sampling distribution. Since the sampling distribution is symmetric, the curve fitting process will take place in the positive quadrant of the residual sampling distribution. The exponential equation used to fit the sampling distribution (with the spike removed) is,

$$y_i = ae^{bx_i^c} \qquad i = 1, 2, \cdots, n \tag{9}$$

where

$a$, $b$, $c$ are unknown parameters

$y_i$ is the relative frequency in the $i^{th}$ case

$x_i$ is the residual magnitude in the $i^{th}$ case.

In addition to accounting for the exponentially shaped portion of the sampling distribution, the sampling distribution must include the zero residuals (spike) in the basic set described in Equation 4, therefore an additional term is added to the curve fitting equation and is presented as,

$$\text{Prob}\{-\infty < v_i < \infty\} = 1$$

$$= 2\underbrace{\int_0^{\infty}\left(\hat{a}e^{\hat{b}x_i^{\hat{c}}}\right)dx}_{exponential\ curve} + \underbrace{\frac{n}{t}}_{spike} \qquad (10)$$

where

$v_i$ is $i^{th}$ residual

$\hat{a}$, $\hat{b}$, $\hat{c}$ are estimated parameters (via $L_2$ norm since erroneous observations are not expected)

n is the number of zero occurrences for the $i^{th}$ residual.

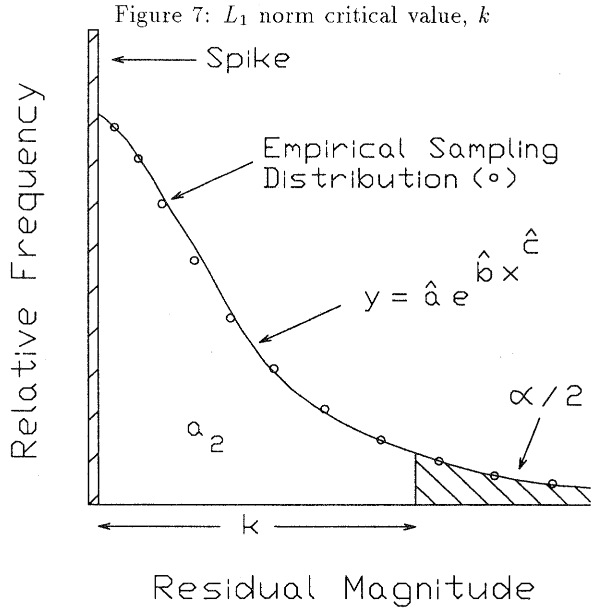
t is the total number of Monte Carlo simulations.

Based on Equation 10, a critical value $k$ is computed for a two-tailed test at a significance level $\alpha$ by solving for a percentage of the area

$$\text{Prob}\{0 < v_i < k\} = 0.5 - \alpha/2$$

$$= \int_0^k \left(\hat{a}e^{\hat{b}x_i^{\hat{c}}}\right)dx + \frac{n}{2t} \qquad (11)$$

Figure 7: $L_1$ norm critical value, $k$




Since a closed form expression for the integral in Equation 11 does not exist, the area under the exponential curve is determined using numerical integration. The critical value, $k$, can be found by the Newton-Rhapson Method (Press, 1992).

The test for individual outliers can be summarized as:

| Hypothesis | Critical Value |
|---|---|
| $H_o : v_i = 0$ | $k$ from Eqn. 11 |
| $H_a : v_i \neq 0$ | |

(12)

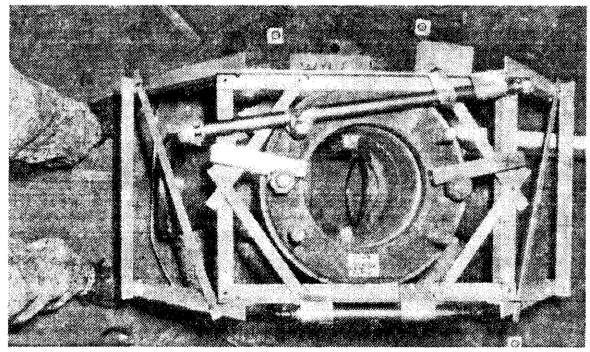
Decision Rule

If $|v_i| > k$ conclude $H_a$

# 7 Photogrammetric Example

To illustrate $L_1$ norm blunder detection in an engineering application, the blunder detection method described in this paper and the well known $L_2$ norm data snooping method were applied to the close-range photogrammetric measurements of the valve pictured in Figure 8.

Figure 8: Close-Range Photogrammetry Example. Photo Courtesy of Tennessee Valley Authority.



The photogrammetric network included four exposure stations, three fixed object points and seven unknown object points, yielding an overall network redundancy of 35. Multiple blunder detection was the focus of this example, therefore two observations were perturbed by 50$\mu$m (noted in Table 1) and the remaining observations were randomly perturbed by $\sigma = 5$ $\mu$m. The results of this limited example suggest that multiple blunders may be correctly identified after a *single* $L_1$ norm estimation whereas *repeated* estimations may be required to identify multiple blunders under $L_2$ estimation.

Table 1 describes the impact of observations with blunders on *one* unknown object point which is observed from four different exposure stations. It appears that the $L_1$ residuals are more consistent with the intentional errors, compared to the $L_2$ residuals. In addition to the statistically-based p-values in Table 1, a compelling reason for computing $L_1$ residuals is the ease with which large residuals can be identified based on a visual examination of residual plots as in Figures 9 and 10.

Table 1: P-values from $L_1$ and $L_2$ norm residuals

| Observation | Blunder Magnitude ($\mu$m) | $L_2$ redundancy number | $L_1$ p-value | $L_2$ p-value |
|---|---|---|---|---|
| $x_1$ | 0.050 | 0.36 | 0.00005 | 0.01097 |
| $y_1$ | 0.000 | 0.62 | 0.32227 | 0.00470 |
| $x_2$ | 0.000 | 0.36 | 0.06923 | 0.00001 |
| $y_2$ | 0.000 | 0.60 | 0.05638 | 0.01420 |
| $x_3$ | 0.000 | 0.59 | 0.01854 | 0.00001 |
| $y_3$ | 0.000 | 0.47 | 0.19880 | 0.25638 |
| $x_4$ | 0.000 | 0.61 | 0.30007 | 0.22081 |
| $y_4$ | 0.050 | 0.51 | 0.00001 | 0.00001 |

It is interesting to note that the initial network under study contained three exposure stations which were plagued by residual sampling distributions containing undesirable "spike-only" distributions. Adding a fourth exposure station to the network (and thereby increasing

Figure 9: $L_1$ Norm Residual Plot: Data Contains Two $50\mu m$ Blunders

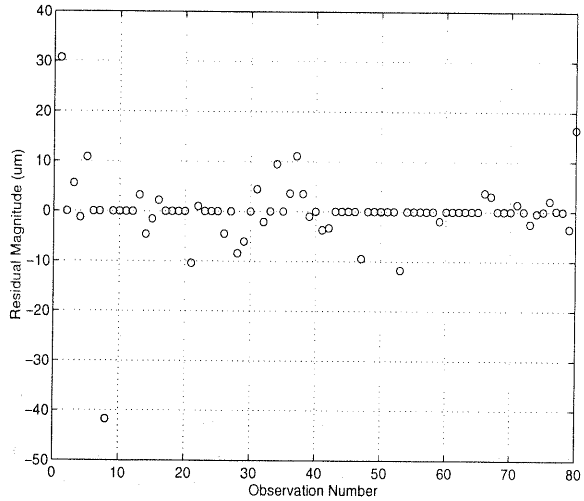
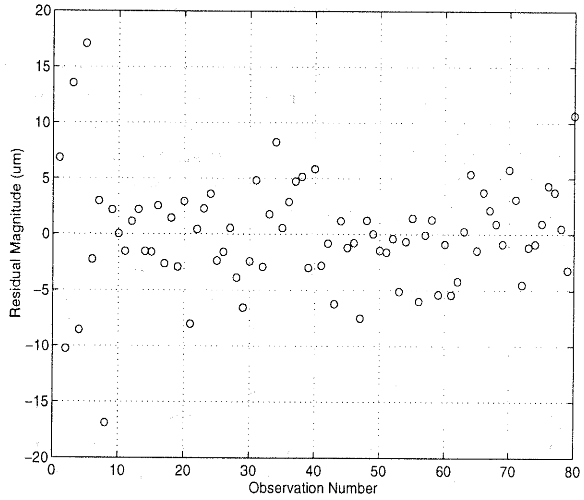



Figure 10: $L_2$ Norm Residual Plot: Data Contains Two $50\mu m$ Blunders




network redundancy) eliminated the "spike-only" residual sampling distributions.

# 8 Conclusions

Based on these preliminary investigations and numerical examples, it appears that analysis of $L_1$ residuals can be put on a sound statistical footing by Monte Carlo generation of sampling distributions. This together with the robust character of $L_1$ estimation makes it worthy of consideration for analysis of photogrammetric and geodetic networks. As pointed out, however, there are some potential pitfalls to avoid in network design when network redundancy is minimal.

# 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.

[2] Bassett, G., 1973. Some properties of the Least Absolute Error Estimator. PhD. Dissertation, University of Michigan.

[3] Kavouras, M., 1982. On the Detection of Outliers and the Determination of Reliability in Geodetic Networks. Technical Report No. 87, University of New Brunswick, Canada.

[4] Mackenzie, A., 1985. Design and Assessment of Horizontal Survey Networks. Master of Science Thesis, University of Calgary.

[5] Marshall, J. and Bethel, J., (in press 1996). Basic Concepts of $L_1$ Norm Minimization for Surveying Applications. Journal of Surveying Engineering.

[6] Menke, W., 1989. Geophysical Data Analysis: Discrete inverse theory, Academic Press, Inc. San Diego, CA.

[7] Mikhail, E., 1976. Observations and Least Squares, IEP, New York.

[8] Pope, A., 1976. The Statistics of Residuals and The Detection of Outliers, NOAA Technical Report NOS 65 NGS 1, Rockville, Md.

[9] Press, W., et. al 1992. Numerical Recipes in C, the Art of Scientific Computing, Second Edition, Cambridge University Press.

[10] Zarembka, P. ed., 1974. Frontiers in econometrics. Academic Press, New York, N.Y.