INCREASING THE RELIABILITY OF GPS GEODETIC NETWORKS

By

Dr. Dalal S. Alnaggar [1]   Eng. Gomaa M. Dawod [2]

ABSTRACT

The use of outlier detection is an important step in the statistical analysis of precise GPS surveys. A special GPS adjustment program was developed with a built-in automatic outlier detector which utilizes the τ (TAU) statistical test. Using a 20-station GPS network, the developed program was examined and the quality of outlier detection was investigated. Significant improvements were obtained after identifying and removing erroneous observations. Standard deviations of the station coordinates were in the range of millimeters.

1. INTRODUCTION

Precise measurements, such as the Global Positioning System (GPS) satellite positioning data, require a careful screening of the field data to detect and identify those observations that may have been affected by gross errors. Including these erroneous observations in the computation process may significantly affect the final results and may give misleading interpretations. Also, it is a matter of fact that gross errors are not avoidable in measurements. Consequently, detecting outliers in high-precision GPS surveys becomes a must.

A great deal of research has been carried out in past years on the development of statistical and numerical techniques to detect outliers in precise engineering measurements. Barada (1968) and Pope (1976) are examples of such researches in geodesy. El-Hakim (1981) provides applications in photogrammetry. However, not much research has focused

[1]   Director of the Survey Research Institute, National Center for Water Researches, 308 Al-Ahram St., Giza 12111, Egypt.
[2]   Assistant Researcher, the Survey Research Institute.

on the applications of outliers identification in GPS data which posses special characteristics.

The main objective of this study was to design an efficient computer program that adjusts GPS networks and, during each iteration, checks for the presence of outliers and identifies them. The wide use of GPS data within the Survey Research Institute (SRI) activities has made this task very important in order to obtain reliable results.

Section Two of this paper presents an overview of the statistical theory behind outliers identification. The developed computer program is described in section Three. Section Four demonstrates the obtained results and the improvements in a small GPS network that was carried out by SRI. Concluding remarks are given in section Five.

2. STATISTICAL TESTS FOR OUTLIERS DETECTION

Gross errors may be defined as the results of a malfunctioning of either the instrument, or the surveyor. It is naturally expected that outliers are caused by gross errors. But, what is an outlier?. Caspary (1987) defines an outlier as " a residual which, according to some test rule, is in contradiction to assumptions on the stochastic properties of the residuals". Therefore, the detection of outliers depends on the selected risk level, the assumed distribution, and the test procedure.

Generally, the methods used in identifying outliers may be grouped according to two basic concepts of modeling the outliers (Chen et al. 1987):
(i) outliers have a mean shift model with the distribution of
   $N(\mu+\delta,\sigma^2)$ instead of $N(\mu,\sigma^2)$ where μ is the expectation, σ^2
   is the variance, and δ is the mean shift value.
(ii) outliers come from a variance inflation model with the
   distribution $N(\mu,a^2\sigma^2)$ where $a^2>1$.

Barada (1968) follows the first concept and develops the so-called data-snooping method under the assumption that the a priori standard error of unit weight (σ_o) is known. Pope (1976) presents another test strategy considering σo to be unknown in practice, but its a posterior estimate ($\hat{\sigma}_o$) is available. Krarup et al (1980) prefer the second concept and develop a strategy based on the idea that large residuals indicate less accurate observations and they use an iterative method for reweighting the observations. Pope's method was used in this study in detecting outliers in GPS data.

In the least squares adjustment of full rank observation equations

$$A X + L = \varepsilon \tag{1}$$

where,

$$L \sim N (AX , \sigma^2_o \Sigma) \tag{2}$$

$$\varepsilon \sim N (0 , \sigma^2_o \Sigma) \tag{3}$$

the nx1 vector L represents the observations, the nxn matrix A is the design matrix, X is the mx1 vector of unknowns, $\Sigma = \sigma^2_o P^{-1}$ is the nxn covariance matrix of the observations, P is the nxn weight matrix, ε is the nx1 vector of observation errors, n is the number of observations, and m is the number of unknown parameters.

After the adjustment, the residuals and their estimated covariance matrix are:

$$V = [I - A (A^T P A)^{-1} A^T P] L \tag{4}$$

$$\hat{\Sigma}_V = [P^{-1} - A (A^T P A)^{-1} A^T] \hat{\sigma}^2_o \tag{5}$$

Instead of the residual (V), another quantity

V' = normalized (or standardized) residual = $|V| / \sigma_v$

is used in the test statistics of the i-th observation:

$$T_i = \frac{|V|}{\sigma_v} \sim \tau(f) \tag{6}$$

using the τ (tau) distribution with f degrees of freedom.

Pope (1976) provides an algorithm which computes the critical τ value. Tables of the τ distribution are also given. Although the τ-distribution is used in geodetic applications, it is not found in some statistical literature. However, it should be stated that the τ distribution can be transformed from the known Student's t-distribution.

Some characteristics of the τ distribution are (Patterson 1985):

1- The population mean is zero.
2- The variance is unity for all degrees of freedom.
3- The τ distribution coverages is distribution to N(0,1).
4- The τ distribution has finite limits.
5- For high degrees of freedom the τ distribution is greatly   resemble the normal distribution.
6- Due to property 4, the τ distribution differs markedly from   the normal distribution for low degrees of freedom.

The null hypothesis of the τ-test assumes that all observations are normally distributed with E(L) = AX, so that the expectation of the residuals is zero:

$$Ho : E(V_i) = 0 \tag{7}$$

The alternative hypothesis is:

$$Ha : E(V_i) \neq 0 \quad \text{for one residual.} \tag{8}$$

Ho is rejected for a residual V_j, if

$$Tj = \frac{|V_j|}{\sigma_{V_j}} > \tau_{\alpha,f} \tag{9}$$

for a certain type I error percentile (α).
From the above methodology, it can be said that the use of normalized residuals is much more meaningful than the use of the residuals themselves. If a residual (even with small magnitude) is much larger than its standard deviation, then it is likely that the corresponding observation is an outlier.

3. THE DEVELOPED ADJUSTMENT PROGRAM

One of the most important characteristic of assessing adjustment programs is how the program deals with possible numerical problems that sometimes may arise from the algorithm used to reduce, solve, and invert the system of normal equations. In this regard, the developed adjustment program (NADO: Network Adjustment and Detecting Outliers) possesses several properties that make it an efficient program.

The program utilizes the inner constraint technique, known also as free

network adjustment, in order to detect the internal precision and consistency of the field observations. For GPS networks, it is essential to define only an origin for the (WGS-84) datum since the orientation and scale are implicitly known from the phase observables because the coordinates of the GPS satellites are assumed to be known. Consequently, to overcome the (three) expected datum defects, one station is held fiducial to its coordinates deduced from the observations (i.e., in the WGS84 coordinate system). By this method, problems due to errors in the coordinates of fixed stations are avoided.

When solving the normal equations system, NADO implements the concept of observation equations with weighted parameters which can be found in details in several adjustment texts (e.g. Uotila 1986, pp. 104). Dawod (1991, pp. 32) demonstrates that this method is just a simplified least-squares adjustment in the Gauss-Markoff model with pseudo observations (i.e., the prior information about the coordinates of the fiducial station).

Inverting the normal equation matrix (N=ATPA) requires a rigorously stable algorithm to remove numerical problems. The Cholesky factorization algorithm is employed in the developed program and hence makes the solution three times faster than any other similar inversion procedure (Press et al. 1989). The normal equation matrix, N, is decomposed to the product of a lower triangular matrix, S, and its transpose:

$$N = S\ S^{T} \tag{10}$$

The inverse of N could be written as
$$N^{-1} = (S\ S^{T})^{-1} = (S^{T})^{-1}\ S^{-1} = (S^{-1})^{T}\ S^{-1} \tag{11}$$

To invert the lower triangular matrix, S, less execution time is required when compared to the inversion of the full matrix N.

Another advantage of the Cholesky factorization algorithm is that it reveals the defects in the normal equation matrix. The number of generated zeros on the diagonal represents the number of defects. This property is important in detecting outliers and removing observations because the removed observations might create observational defects in the network. This case happened in one of the iterations explained in section Four. In such situations, NADO informs the user about the locations of the new defects in the normal equation matrix. Therefore, singularities are properly handled by not solving for the unknowns causing those singularities.

Considering the covariance matrix of the observations, Σ, is a diagonal matrix, both the (A^TPA) matrix and the (A^TPL) vector can be formed by summing or accumulating the contribution of the observations one by one. Therefore, there is no need to store the full A(nxm), Σ(nxn), or L(nx1) in the computer memory. This concept is used in the developed program in order to reduce the required computer memory and, hence, adjust larger GPS networks in a PC environment.

NADO also converts the estimated parameters, i.e., the Cartesian coordinates of the network stations, to a geodetic coordinate system (latitude, longitude, and height) along with their covariance matrices.

For detecting outliers, the program uses equation 6. Normalized residuals are computed for each GPS vector component. Each base line therefore has 3 normalized residuals, one for each of its 3 components. The largest value is compared against the computed critical τ value using a risk level of 5%. Equation 9 is then used to flag outlying observations. The structure of the program is depicted in Figure 1.

4. RESULTS AND ANALYSIS

A small GPS network was examined from the point of view of detecting outliers in GPS networks, where coordinate differences from individual base line processing are used as observables in a network configuration. The project is a 20-station GPS network designed and observed by SRI in 1992 (Fig. 2). The main purpose of the project was to establish a control GPS network in Aswan area and around lake Nasser for crustal movement monitoring. The network covers an area extending approximately 50 Km in the north-south direction and a width ranging from 2 to 10 Km. The network was observed with 5 Trimble GPS receivers, two of them are dual-frequency receivers.

After processing the raw GPS phase observations, 55 fixed solutions were obtained and used as observables for the developed adjustment program. Therefore, the number of observations in the first adjustment was 165 with 108 degrees of freedom. The relatively large estimated standard deviation of unit weight may be thought of as a global indicator that the observations may be contaminated by gross errors. Applying the built-in subroutine for detecting outliers shows that 11 base lines are flagged as candidate outliers. In the subroutine, 5% type I error was used. Results of the initial adjustment are shown in table 1.

Table 1
Results of the First Adjustment

Number of base lines	55
Number of observations	165
Number of stations	20
Number of unknowns	60
Degrees of freedom	108
Estimated $\hat{\sigma}^2_o$	19.2
Critical τ value	3.52
Flagged outliers	11

Based on these results, it was decided to delete only the base line that had the largest normalized residual. Most of the literature dealing with outlier detection stressed that any algorithm should not be used as a black box which automatically cleans the observations. This warning may be explained when we remember the original assumptions behind the τ -test for outlier detection: (1) all observations are normally distributed; and more important (2) only one outlier is assumed to be present in the data set. Consequently, the whole theory breaks down if the observations include two or more gross errors.

The following rational approach is followed in order to obtain as much correct results as possible:

* Initial adjustment is carried out using all observations.
* If more than one normalized residual exceed the critical τ value, only the base line with the largest normalized residual is deleted.
* The adjustment is repeated again with n-3 observations leading to new residuals and new $\hat{\sigma}^2_o$
* This process is repeated until all outliers are flagged.

Applying this methodology, 13 different adjustments were carried out. Results of the last two adjustment are shown in table 2.

Table 2
Results of the Last Two Adjustments

	Adjustment 12	Adjustment 13
No. of Observations	102	99
No. of Base Lines	34	33
Degrees of Freedom	45	42
Estimated $\hat{\sigma}^2_o$	1.299	0.483
No. of Outliers	4	1

Although four outliers were flagged in the iteration 12, it was decided to accept its results as the optimum. Two reasons may legitimize that acceptance: (1) the estimated standard deviation of unit weight of iteration 13 goes below one; and (2) the last GPS base line deleted in iteration 13 creates a new observation defect because it was the only remaining base line connecting station 18 to the rest of the network. NADO is capable of detecting any new observation defects after fixing the datum. This is done by observing singularities (i.e., zeros on the diagonal) when inverting the lower triangular matrix, S, of the Cholesky factorization algorithm. This property is very important and should be applied in any adjustment program. NADO displays on screen the locations of the newly created defects in an interactive mode with the user. Therefore, the results of adjustment 12 were accepted as final for the GPS network stations to avoid additional observations.

In order to investigate the improvements in the estimated parameters, a comparison is carried out between the results of the first and the twelfth adjustments. Some results are given in table 3. Significant improvements in terms of the estimated standard deviations of the station coordinates are clearly visible after removing outliers. Standard deviations of the stations coordinates are in the range of millimeters.

Table 3
Improvements After Detection Outliers

Adj.	St.	X	Y	Z
1 12	2	4896330.813±0.091 30.822±0.008	3167453.249±0.096 53.231±0.009	2575259.755±0.101 59.752±0.009
1 12	3	4900749.351±0.080 49.353±0.008	3165047.939±0.107 47.921±0.011	2569831.582±0.125 31.549±0.013
1 12	4	4899907.405±0.123 07.358±0.024	3164614.389±0.089 14.384±0.017	2571940.935±0.085 40.925±0.016
1 12	5	4899546.297±0.078 46.303±0.010	3166777.201±0.103 77.204±0.013	2569984.087±0.125 84.204±0.016

5. CONCLUSIONS

Significant consideration has been given to the detection of outliers in the field of precise surveys. The τ-test, among other statistical tests, offers statistical tools used to identify those erroneous observations that may be contaminated by gross errors. Using the τ criteria, a special GPS adjustment program with a built-in subroutine for outlier detection was developed. A GPS network that was carried out by SRI in 1992 was used to examine the new program. The program shows efficiency in solving the system of normal equations and identifying suspected observations.
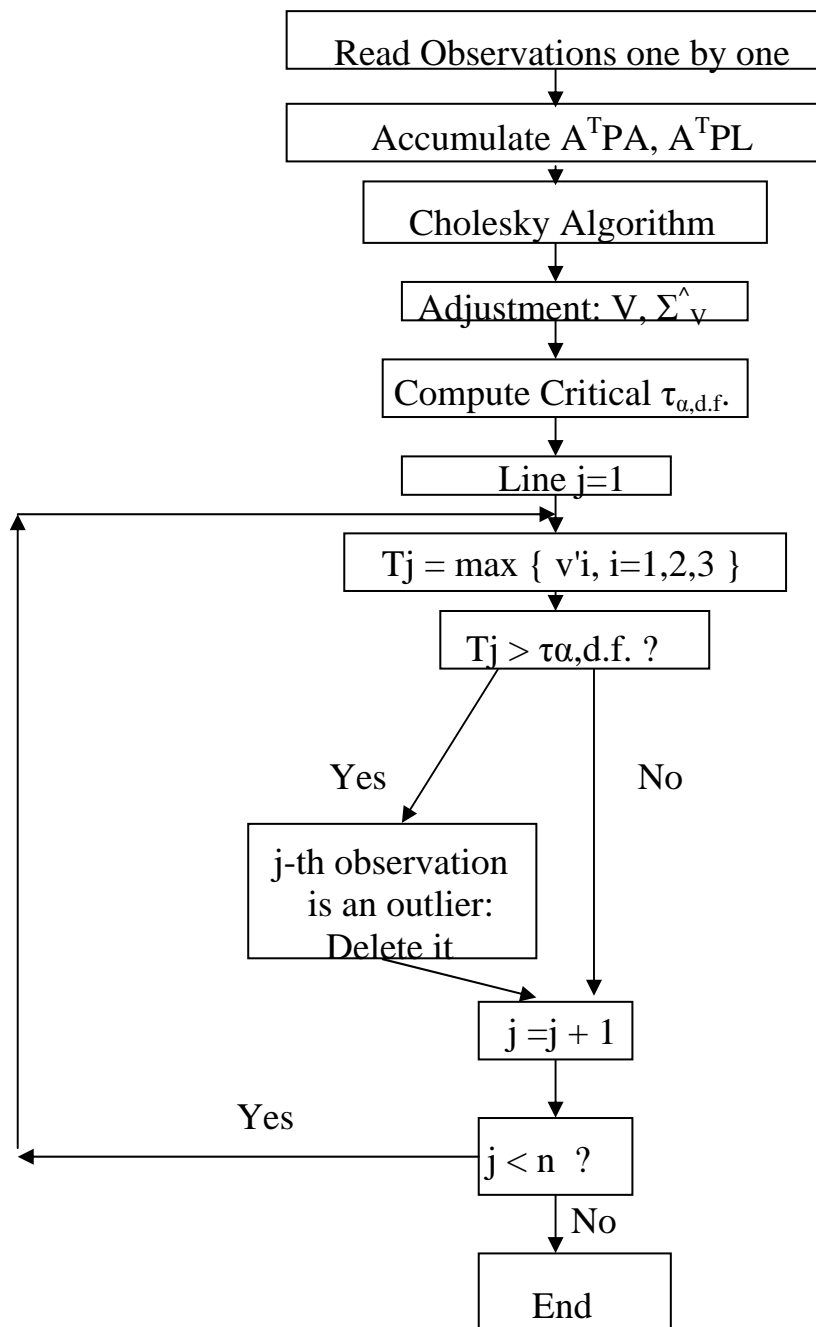
The τ test is based on the assumption that only one outlier is present in the data set. A difficulty with this procedure is encountered when two or more observations are flagged as outliers. The approach used in this study was to delete the observation that has the largest normalized residual in each iteration. Then, the adjustment is repeated with n-3 observations leading to new residuals. Because of the assumption behind the τ test, it is emphasized that only one observation is deleted every time the adjustment is repeated. It is also stressed that any outlier detection algorithm should be used in an interactive way with the user and not in an automatic mode of cleaning the observations.

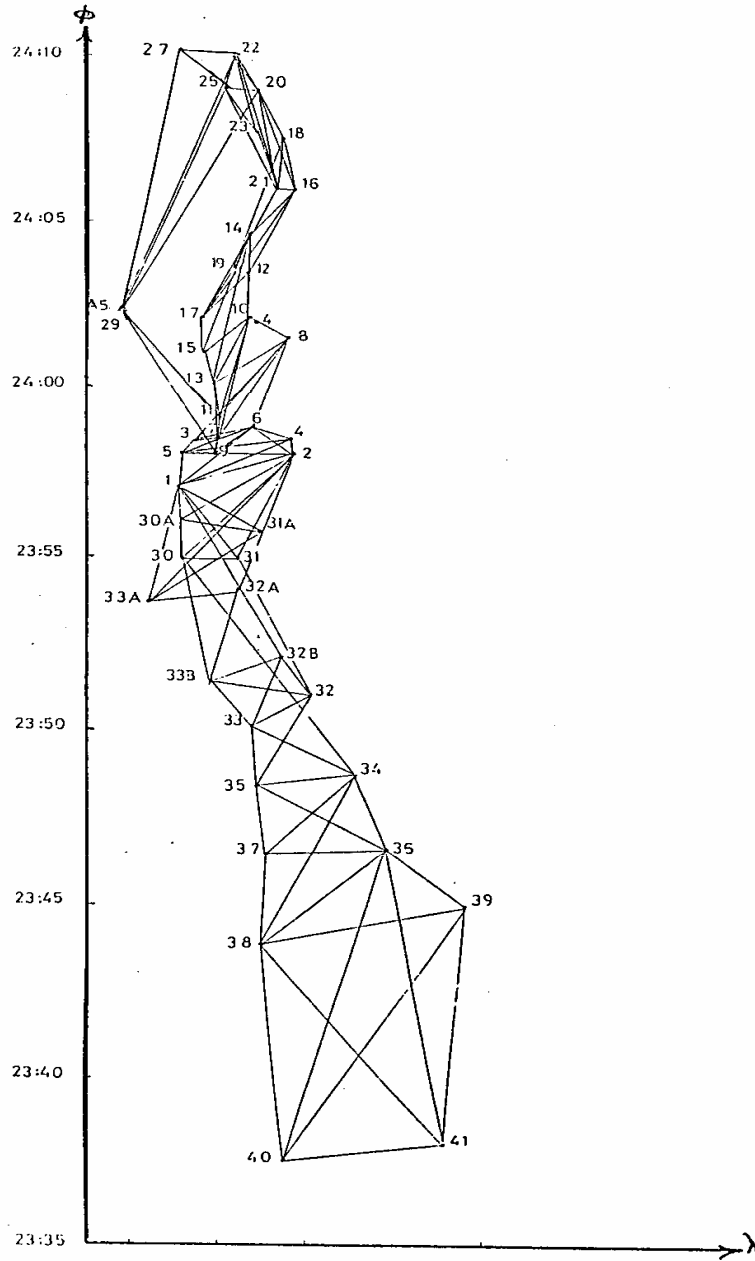In the GPS network used in this study, 13 different adjustments were

carried out. Even having four remaining outliers, iteration 12 was accepted as the final step because deleting any new outlier will create observation defects and cause some stations to be disabled from the network. It was accepted to avoid having to obtain new observations. A comparison between the first and last adjustments shows significant improvements in the estimated parameters after detecting and removing outliers. Millimeter level standard deviations of the stations coordinates were obtained after identifying outlying observations. It is recommended that outlier detection procedures should be implemented in the adjustment of precise GPS surveys in order to obtain reliable results and proper interpretations.

LIST OF REFERENCES

Barada, W., 1968, A testing procedure for use in geodetic networks, Publications on Geodesy, New Series, Netherlands Geodetic Commission, V. 2, N. 5, Delft, The Netherlands.

Caspary, W., 1987, Concepts of networks and deformation analysis, Monograph No. 11, School of Surveying, University of New South Wales, Australia.

Chen, Y., Kavouras, M., Chrzanowski, A., 1987, A strategy for detection of outlying observations in measurements of high precision, The Canadian Surveyor, V. 41, N. 4, pp. 529-540.

Dawod, G., 1991, Some considerations in the adjustment of GPS-derived base lines in the network mode, MSC thesis, Department of Geodetic Science and Surveying, The Ohio State University, Ohio, USA.

El-Hakim, S., 1981, A practical study of gross-error detection in bundle adjustment, The Canadian Surveyor, V. 35, N. 4, pp. 373-386.

Krarup, T., Kubik, T., Juhl, J., 1980, Gotterdammerung over least squares adjustment, International Archives of Photogrammetry, V. XXIII, Part B3, Commission III, ISP, Hamburg, pp. 369-378.

Patterson, R., 1985, Outlier detection: Some distribution theory, Australian Journal of Geodesy, Photogrammetry, and Surveying, N. 43, pp. 1-30.

Pope, A, 1976, The statistics of residuals and the detection of outliers, NOAA Technical Report NOS 65 NGS 1, U.S. Department of Commerce, Washington, D.C., USA.

Press, W., Flannery, B., Teukolsky, S., Vetterling, W., 1989, Numerical recipes: The art of scientific computing, Cambridge University Press.

Uotila, U., 1986, Notes on adjustment computations: Part I, Lecture notes, Department of Geodetic Science and Surveying, The Ohio State University, Ohio, USA.

Read Observations one by one

↓

Accumulate A^TPA, A^TPL

↓

Cholesky Algorithm

↓

Adjustment: V, $\hat{\Sigma}_V$

↓

Compute Critical $\tau_{\alpha,\text{d.f.}}$

↓

Line $j=1$

↓

$T_j = \max \{ v'_i, i=1,2,3 \}$

↓

$T_j > \tau_{\alpha,\text{d.f.}}$?

Yes ↙   No ↓

j-th observation
is an outlier:
Delete it

↓

$j = j + 1$

↓

$j < n$?

Yes ←   No ↓

End

Fig. 1
Structure of the Developed Program

Fig 2
The GPS Network