Magical Least Squares
or
When is One Least Squares Adjustment Better Than Another?

© Earl F. Burkholder, PS, PE
NMSU Dept of Surveying Engineering - Las Cruces, NM 88003
September 2005

Introduction

Least squares adjustment, that is, the sum of the squares of the residuals will be a minimum, has been proven and accepted as the best possible method for adjusting survey data.

Within reason, it is also true that, depending upon how weights are selected, you can get any answer you want using least squares. Therefore, the issue in discussing "How good are my results?" switches from the choice of the tool (least squares) to how the tool is used. Of course, the input data must first be checked and verified blunder-free.

Given blunder-free survey data and a specific statement of how weights are selected, all least squares packages should provide the same answers. Differences from one brand software to another will have to do with the survey data input (formats, weights etc) and what information is included in the report after the adjustment is completed. This article looks at 3 different weighting assumptions on a small network and compares the results.

The example used in this paper is a GPS network based upon two A-order HARN points. Station "Reilly" is located in the central horseshoe of the NMSU campus and Station "Crucesair" is located at the Las Cruces airport some 16 kilometers west of campus. The network consists of 7 independent baselines connecting 4 additional points to the existing HARN stations as shown in Figure 1.

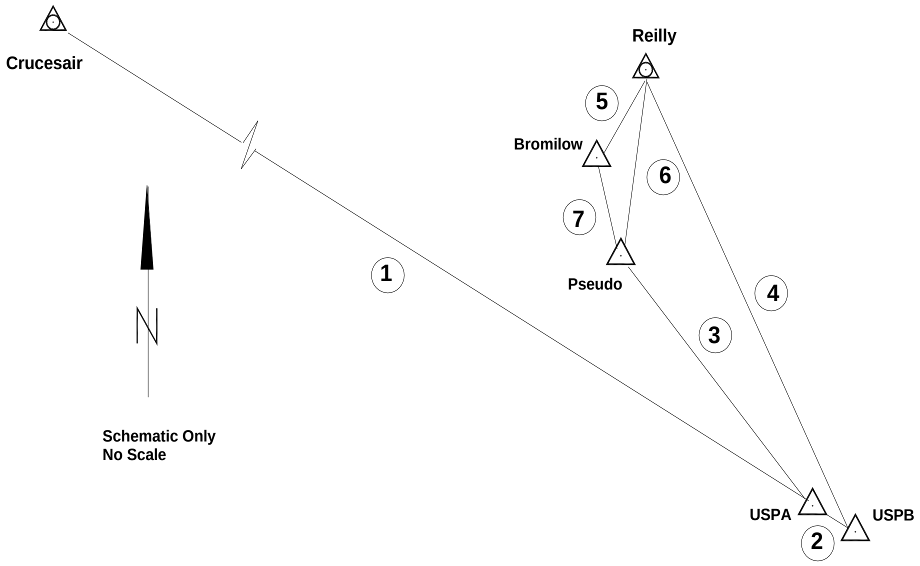

Figure 1 GPS Network at NMSU

The GPS baselines shown and used were collected on four different dates over a period of 5 years. These are not the only baselines on campus nor are they the only observations between the points in question. These baselines were selected because they show excellent consistency, are independent, and include often used points. The results can also be used to show the difference between network accuracy and local accuracy.

Control Values and Observed Vectors

The NAD83 geocentric X/Y/Z coordinates for A-order HARN stations “Reilly” and “Crucesair” are as published by the National Geodetic Survey (NGS) and were held fixed in this exercise. They are:

Station Reilly	Station Crucesair
X = -1,556,177.615 m	X = -1,571,430.672 m
Y = -5,169,235.319 m	Y = -5,164,782.312 m
Z = 3,387,551.709 m	Z = 3,387,603.188 m

Single frequency Trimble GPS receivers were used to collect static data, 57 minutes being the shortest common observation time for any of the 7 baselines. The baseline components and the covariance matrix for each observed baseline as determined by Trimble software using default processing parameters are:

Baseline 1 – Crucesair to USPA – observed 3/28/02 (use subscript CA):

			Sxx	Syy	Szz
ΔX_{CA} =	15,752.080 m	Sxx	6.321492E-06		
ΔY_{CA} =	-5,179.102 m	Syy	1.545948E-05	4.739877E-05	
ΔZ_{CA} =	-903.089 m	Szz	-1.061303E-05	-3.184780E-05	2.388036E-05

Baseline 2 – USPA to USPB – observed 11/12/03 (use subscript AB):

			Sxx	Syy	Szz
ΔX_{AB} =	14.964 m	Sxx	1.412453E-06		
ΔY_{AB} =	-15.365 m	Syy	1.285418E-06	4.653209E-06	
ΔZ_{AB} =	-16.664 m	Szz	-5.669127E-07	-1.658118E-06	1.872469E-06

Baseline 3 – USPA to Pseudo – observed 3/28/02 (use subscript AP):

			Sxx	Syy	Szz
ΔX_{AP} =	-528.036 m	Sxx	9.505016E-08		
ΔY_{AP} =	560.657 m	Syy	8.957064E-08	3.729339E-07	
ΔZ_{AP} =	585.897 m	Szz	-5.022282E-08	-2.221975E-07	3.363763E-07

Baseline 4 – USPB to Reilly - observed 3/28/02 (use subscript BR):

			Sxx	Syy	Szz
ΔX_{BR} =	-514.003 m	Sxx	3.650165E-07		
ΔY_{BR} =	741.438 m	Syy	9.024127E-07	2.796189E-06	
ΔZ_{BR} =	868.293 m	Szz	-6.189027E-07	-1.881145E-06	1.410196E-06

Baseline 5 – Bromilow to Reilly - observed 12/10/98 (use subscript MR):

			Sxx	Syy	Szz
ΔX_{MR} =	32.134 m	Sxx	2.762550E-07		
ΔY_{MR} =	51.175 m	Syy	3.200312E-07	6.870545E-07	
ΔZ_{MR} =	94.198 m	Szz	-2.008940E-07	-4.006259E-07	4.661596E-07

Baseline 6 – Pseudo to Reilly – observed 1/23/02 (use subscript PR):

				Sxx	Syy	Szz
ΔX_{PR}	=	29.000 m	Sxx	1.325760E-07		
ΔY_{PR}	=	165.422 m	Syy	1.317165E-07	5.265054E-07	
ΔZ_{PR}	=	265.719 m	Szz	-7.253348E-08	-3.020965E-07	5.006575E-07

Baseline 7 – Bromilow to Pseudo – observed 1/23/02 (use subscript MP):

				Sxx	Syy	Szz
ΔX_{MP}	=	3.136 m	Sxx	3.367818E-07		
ΔY_{MP}	=	-114.242 m	Syy	3.937476E-07	8.766570E-07	
ΔZ_{MP}	=	-171.527 m	Szz	-5.186521E-07	-8.977932E-07	1.446501E-06

Blunder Checks

In order to verify the absence of blunders in the baselines, misclosures are computed for each component (X/Y/Z) as follows:

Traverse including baselines 1, 2, and 4 (from “Crucesair” to “Reilly”):

	X	Y	Z
Station Crucesair	-1,571,430.672 m	-5,164,782.312 m	3,387,603.188 m
Baseline 1	15,752.080 m	-5,179.102 m	-903.089 m
Baseline 2	14.964 m	-15.365 m	-16.664 m
Baseline 4	-514.003 m	741.438 m	868.293 m
Computed value	-1,556,177.631 m	-5,169,235.341 m	3,387,551.728 m
Station Reilly	-1,556,177.615 m	-5,169,235.319 m	3,387,551.709 m
Misclosures	**-0.016 m**	**-0.022 m**	**0.019 m**

Loop including baselines 2-3-7-5-4 (being careful to preserve sign convention):

Baseline 2	-14.964 m	15.365 m	16.664 m
Baseline 3	-528.036 m	560.657 m	585.897 m
Baseline 7	-3.136 m	114.242 m	171.527 m
Baseline 5	32.134 m	51.175 m	94.198 m
Baseline 4	514.003 m	-741.438 m	-868.293 m
Misclosures	**0.001 m**	**0.001 m**	**-0.007 m**

Loop including baselines 5-6-7 (being careful to preserve sign convention):

Baseline 5	32.134 m	51.175 m	94.198 m
Baseline 6	-29.000 m	-165.422 m	-265.719 m
Baseline 7	-3.136 m	114.242 m	171.527 m
Misclosures	**-0.002 m**	**-0.005 m**	**0.006 m**

All baselines have been included in the checks and all misclosures are acceptable. Therefore, it is legitimate to perform a least squares adjustment of the 7 baselines to determine the “best” adjusted position for points USPA, USPB, Pseudo, and Bromilow. Any adjustment should also provide information on the quality of the answers, i.e., “What is the standard deviation of the computed position?” - in both the geocentric (X/Y/Z) reference frame and in the local (east/north/up) reference frame. This paper uses 3 different weighting schemes and shows a comparison of the various answers.

Results

The unabridged paper contains many pages between the “Blunder Checks” (the previous section) and these “Results.” The complete paper is posted at www.zianet.com/globalcogo/nmsunet1.pdf. The portion omitted from this summary uses a lot of matrices but includes documentation of the entire computational process. It is intended for everyone to have access to the complete documented solution.

A summary of the original paper was presented and discussed at the NMPS Llano Estacado Chapter seminar in Ruidoso, NM on August 13, 2005. This abridged version is submitted to the NMPS Benchmarks and is intended to fulfill the promise made in the May 2005 issue of the Benchmarks to share subsequent results of the NMSU network adjustment project in a future issue of the Benchmarks.

There is small difference between this paper and the summary discussed in Ruidoso. As it turns out, the previous summary was defective because the adjusted inverse distance between USPA and USPB was 27.170 m +/- 0.001 m but the actual taped distance is 27.162 m +/- 0.002 m. Statistically, those answers appear to be incompatible. At the seminar, we discussed possible reasons for the discrepancy and noted that there was only one GPS occupation of USPA and USPB. Without redundant occupation at those two station, we have no way of knowing that the antennas were precisely over the intended marks. If the tribrachs were not well calibrated or if either or both set-ups were a bit sloppy, then both answers could be right. If both answers are right, then the problem is that they do not represent the same distance.

In order to resolve that discrepany, the USPA-USPB baseline was reobserved. The second GPS distance of that baseline (27.160 m) agrees much better with the taped distance. So, this paper is different from the one discussed in Ruidoso in that the original vector between USPA and USPB was replaced by the reobserved vector. The results shown here are based upon the re-observed vector USPA to USPB.

The following comparison reflects three different weighting decisions during adjustment of the NMSU GPS network based upon A-order HARN stations “Crucesair” and “Reilly.” As shown in Figure 1, four new points were surveyed and the network was computed using the three following options for selecting weights:

Option 1: All baseline components were weighted equally.

Option 2: Weights were assigned according to the standard deviation of each baseline component as reported by the baseline processor.

Option 3: The weight matrix was computed using the full covariance matrix of each baseline as reported by the baseline processor.

The geocentric X/Y/Z coordinates and their standard deviations were computed for each new point in the least squares adjustment. The geodetic coordinates and local standard deviations were then computed from the geocentric values. The comparison shows both the geocentric X/Y/Z coordinates along with their standard deviations and the geodetic coordinates along with the local reference frame (east/north/up) standard deviations. Option 3 values are highlighted with **bold** and recommended for subsequent use.

Geocentric Coordinates & Sigmas	Geodetic Coordinates & Local Sigmas
USPA – Option 1	deg min sec
X = -1,555,678.5843 m +/- 0.0050 m	φ = 32 16 23.00012 N +/- 0.0050 m (n)
Y = -5,169,961.4037 m +/- 0.0050 m	λ = 106 44 48.90828 W +/- 0.0050 m (e)
Z = 3,386,700.0922 m +/- 0.0050 m	h = 1,178.025 m +/- 0.0050 m (u)
USPA – Option 2	deg min sec
X = -1,555,678.5788 m +/- 0.0015 m	φ = 32 16 23.00020 N +/- 0.0027 m (n)
Y = -5,169,961.3966 m +/- 0.0029 m	λ = 106 44 48.90816 W +/- 0.0017 m (e)
Z = 3,386,700.0898 m +/- 0.0027 m	h = 1,178.016 m +/- 0.0028 m (u)
USPA – Option 3	**deg min sec**
X = -1,555,678.5792 m +/- 0.0015 m	**φ = 32 16 23.00018 N +/- 0.0027 m (n)**
Y = -5,169,961.3961 m +/- 0.0029 m	**λ = 106 44 48.90818 W +/- 0.0017 m (e)**
Z = 3,386,700.0889 m +/- 0.0026 m	**h = 1,178.015 m +/- 0.0027 m (u)**
USPB – Option 1	deg min sec
X = -1,555,663.6161 m +/- 0.0056 m	φ = 32 16 22.36248 N +/- 0.0056 m (n)
Y = -5,169,976.7629 m +/- 0.0056 m	λ = 106 44 48.19160 W +/- 0.0056 m (e)
Z = 3,386,683.4221 m +/- 0.0056 m	h = 1,177.912 m +/- 0.0056 m (u)
USPB – Option 2	deg min sec
X = -1,555,663.6123 m +/- 0.0019 m	φ = 32 16 22.36251 N +/- 0.0038 m (n)
Y = -5,169,976.7587 m +/- 0.0048 m	λ = 106 44 48.19151 W +/- 0.0023 m (e)
Z = 3,386,683.4202 m +/- 0.0034 m	h = 1,177.907 m +/- 0.0043 m (u)
USPB – Option 3	**deg min sec**
X = -1,555,663.6134 m +/- 0.0018 m	**φ = 32 16 22.36245 N +/- 0.0037 m (n)**
Y = -5,169,976.7610 m +/- 0.0047 m	**λ = 106 44 48.19152 W +/- 0.0022 m (e)**
Z = 3,386,683.4193 m +/- 0.0033 m	**h = 1,177.908 m +/- 0.0042 m (u)**
Pseudo – Option 1	deg min sec
X = -1,556,206.6167 m +/- 0.0050 m	φ = 32 16 45.74650 N +/- 0.0050 m (n)
Y = -5,169,400.7423 m +/- 0.0050 m	λ = 106 45 14.39979 W +/- 0.0050 m (e)
Z = 3,387,285.9885 m +/- 0.0050 m	h = 1,165.644 m +/- 0.0050 m (u)
Pseudo – Option 2	deg min sec
X = -1,556,206.6147 m +/- 0.0011 m	φ = 32 16 45.74653 N +/- 0.0021 m (n)
Y = -5,169,400.7396 m +/- 0.0021 m	λ = 106 45 14.39974 W +/- 0.0012 m (e)
Z = 3,387,285.9877 m +/- 0.0021 m	h = 1,165.641 m +/- 0.0021 m (u)
Pseudo – Option 3	**deg min sec**
X = -1,556,206.6150 m +/- 0.0011 m	**φ = 32 16 45.74652 N +/- 0.0020 m (n)**
Y = -5,169,400.7395 m +/- 0.0021 m	**λ = 106 45 14.39976 W +/- 0.0012 m (e)**
Z = 3,387,285.9873 m +/- 0.0020 m	**h = 1,165.641 m +/- 0.0020 m (u)**
Bromilow – Option 1	deg min sec
X = -1,556,209.7509 m +/- 0.0056 m	φ = 32 16 52.33408 N +/- 0.0056 m (n)
Y = -5,169,286.4971 m +/- 0.0056 m	λ = 106 45 15.77275 W +/- 0.0056 m (e)
Z = 3,387,457.5133 m +/- 0.0056 m	h = 1,165.525 m +/- 0.0056 m (u)

Bromilow – Option 2

				deg	min	sec	
X =	-1,556,209.7498 m	+/- 0.0015 m	φ =	32	16	52.33408 N	+/- 0.0022 m (n)
Y =	-5,169,286.4956 m	+/- 0.0028 m	λ =	106	45	15.77273 W	+/- 0.0016 m (e)
Z =	3,387,457.5119 m	+/- 0.0022 m	h =	1,165.523 m			+/- 0.0023 m (u)

Bromilow – Option 3

				deg	**min**	**sec**	
X =	**-1,556,209.7498 m**	**+/- 0.0014 m**	**φ =**	**32**	**16**	**52.33408 N**	**+/- 0.0022 m (n)**
Y =	**-5,169,286.4955 m**	**+/- 0.0024 m**	**λ =**	**106**	**45**	**15.77273 W**	**+/- 0.0015 m (e)**
Z =	**3,387,457.5120 m**	**+/- 0.0022 m**	**h =**	**1,165.523 m**			**+/- 0.0023 m (u)**

Commentary

The SUR 451 (Least Squares) class first computed this NMSU GPS network in the Spring of 2002. At that time we focused on what happened if we changed the magnitude of the weights. For weights we used σ_0^2/σ^2 (which is option 2 in this exercise), but we chose three different values for σ_0^2. First, we chose $\sigma_0^2 = 1.00$ (as in this exercise). Then we chose σ_0^2 such that the smallest weight was 1.00 and we chose σ_0^2 such that the largest weight was 1.00. Surprise, we obtained the same answers for all three solutions and we were happy with what we learned. For those who use the Star*Net least squares program for GPS vectors, choosing a “vector factor” is direclty related to choosing a value for σ_0^2. As one gets more into least squares, you learn that choosing σ_0^2 is really not arbitrary. But, starting off with an estimate of $\sigma_0^2 = 1.00$ (or 8.00 in the case of Star*Net) is a safe conservative approach.

In the Fall 2003 we used the same network, only this time we added more vectors and tied into an OPUS station located east of campus on the flood control dam by Dripping Springs Road. We used Star*Net to compute the network and had fun trying various combinations of vectors and weighting factors to see what it took to come up with an adjustment that passed the Chi Square test. That was also a good exercise but we never really got a good handle on computing the statistics (standard deviations) for the answers.

In the Spring 2005 there was a special projects GPS class (SUR 461) for a single student needing the class to graduate. Given the opportunity to work 1-on-1, we decided to recompute the NMSU GPS network utilizing the different weighting options described in this paper. Melvin Pyeatt (then a senior surveying student) keyed in the data and made numerous computer runs to come up with the three solutions for the network. Replacing vector #2 between USPA and USPB changed the results slightly. With that change, the inverse distance (mark to mark = $\sqrt{[\Delta X^2 + \Delta Y^2 + \Delta Z^2]}$) between USPA and USPB is 27.165 m +/- 4 mm and agrees well with the taped distance of 27.162 m +/- 2 mm.

Local Accuracy versus Network Accuarcy

For those who are interested, the underlying motive for pursuing various weighting options for the NMSU GPS network is to look at the issue of network accuracy versus local accuracy. Surveyors often have an intuitive feel about local accuarcy because we think in terms of “how good is this point with respect to my fixed control point or with respect to the adjacent traverse point?" In the case of GPS, the question may be how good is my OPUS point with respect to the CORS or how good are my rover points with respect to my OPUS station. I don't have specific answers for those questions, but I do suggest that we have the tools available that we can use to look at the issues.

The full paper which describes the NMSU network adjustment is posted at www.zianet.com/globalcogo/nmsunet1.pdf. The posted paper contains the full covariance matrix for the entire adjustment on page 19. That is the matrix from which the coordinate standard deviations listed in this paper are obtained. That covariance matrix also contains correlation information in the off-diagonal elements which can be used to compute the local accuracy of one point with respect to another.

The formal process for computing network accuracy and local accuracy is described in a paper, "Spatial Data Accurcay as Defined by the GSDM," published in the March 1999 issue of the ACSM Surveying and Land Information Systems journal, Vol. 59, Number 1, pp26-30. Equation 9 in that article uses pieces of that same network covariance matrix in the process of computing network accuracy, local accuracy, or P.O.B. datum accuracy. I shared some of the local/network accuracy results at the Ruidoso seminar but they are not included here or in the posted version of this paper. A separate techical paper showing how those details all fit together is being written and will be submitted to a national journal for peer review and possible publication. That is a lengthy process.

As surveyors and others rely more heavily on OPUS and CORS for their control, the issue of spatial data accuracy, network accuracy, and local accuracy will become more and more important. Stay tuned…