The Application of the Method of Least Squares to the Closing of Multiply-Connected Loops in Cave or Geological Surveys

By V. A. Schmidt * and J. H. Schelleng **

ABSTRACT

The method of least squares is applied to cave and geological surveys to obtain an algorithm that allows multiply-connected loops to be closed in a reasonable and analytical manner. In addition, constraints may be applied to any part of the survey. The method is well-suited to use with computer programs for the reduction of survey data.

Introduction

Cave and geological surveys are generally carried out with relatively imprecise instruments and under conditions of time and circumstance that make high accuracy difficult to achieve. An indication of this accuracy is obtained whenever the line or passage being surveyed doubles back on itself to form a loop. The extent to which the survey line, when plotted, closes on itself is a measure of the accuracy of the survey. It is common practice in cave surveys to use hand-held or tripod mounted Brunton compasses and steel or cloth tapes under less than ideal conditions of instrument location and reading comfort. As a result, closure errors of greater than 10 feet in loops only a thousand feet long are common, and closure errors of less than a foot in loops of almost any length are extremely rare.

The problem of closing these loops has long been a troublesome one, especially when multiple loops are involved. The mapper can take one of several tacks. He can simply plot his sightings as they are, and if the closure errors are reasonably small, he can ignore them and construct the map around the uncorrected sightings. If the closure error in the loop begins to approach the dimensions of a cave passage, this procedure must be abandoned. If only one loop is present, the closure error can readily be distributed throughout the loop to effect a closure. If, however, multiply-connected loops are present, the situation becomes more difficult to judge by eye, and the necessary distribution of errors throughout the survey can become extremely tedious if attempted by hand.

With the advent of computer processing of cave survey data, it became clear that some analytical method of closing arbitrarily complicated surveys was needed if the power and elegance of these machines was to be utilized fully. This paper describes an analytical method which is a straightforward application of one of the most powerful statistical tools available to the data gatherer—the method of least squares.

* Department of Earth and Planetary Sciences, University of Pittsburgh, Pittsburgh, Pa. 15213.

** 2000 Jamestown Rd., Alexandria, Va. 22308.

Theory

The method of least squares may be applied to any situation in which a larger number of measurements are made than are minimally necessary in order to uniquely determine the quantities being measured. As an illustration, let us assume that a number of measurements are made on some unknown quantity represented by r. The fundamental tenet of the method of least squares states (Young, 1962) that if the measurement errors are random and follow a Gaussian distribution, then the quantity r has a *most probable value* which we may call R and which is determined in such a way as to minimize the sum of the squares of the deviations of each measured value r_i from the most probable value R. If M independent measurements are made yielding measured values for r of $r_1, r_2, \ldots, r_i, \ldots, r_M$ with corresponding weights $w_1, w_2, \ldots, w_i, \ldots, w_M$, then R may be determined by minimizing the expression

$$\sum_{i=1}^{M} w_i d_i^2 = \sum_{i=1}^{M} w_i (r_i - R)^2 \qquad (1)$$

The application of this method to cave surveys is made by regarding the skeleton survey as a network of three-dimensional vectors, each vector representing a "sight" from one survey station to another. In general each vector represents three measurements: a distance, an azimuthal angle, and a vertical angle. In addition, the head and tail of each vector must be labelled with station names or numbers so that the network of vectors may be properly strung together to form the complete survey.

From this, we can see that the method of least squares may not be applied to a survey that contains no loops since only enough measurements have been made to uniquely determine the location of each survey station. (Once a base station has been established or defined, each sight locates one more station. If the traverse line never closes on itself, then each station has been given a location only once.) If, however, one or more loops are formed in the survey and the loops do not close perfectly, each station within the loops has been assigned more

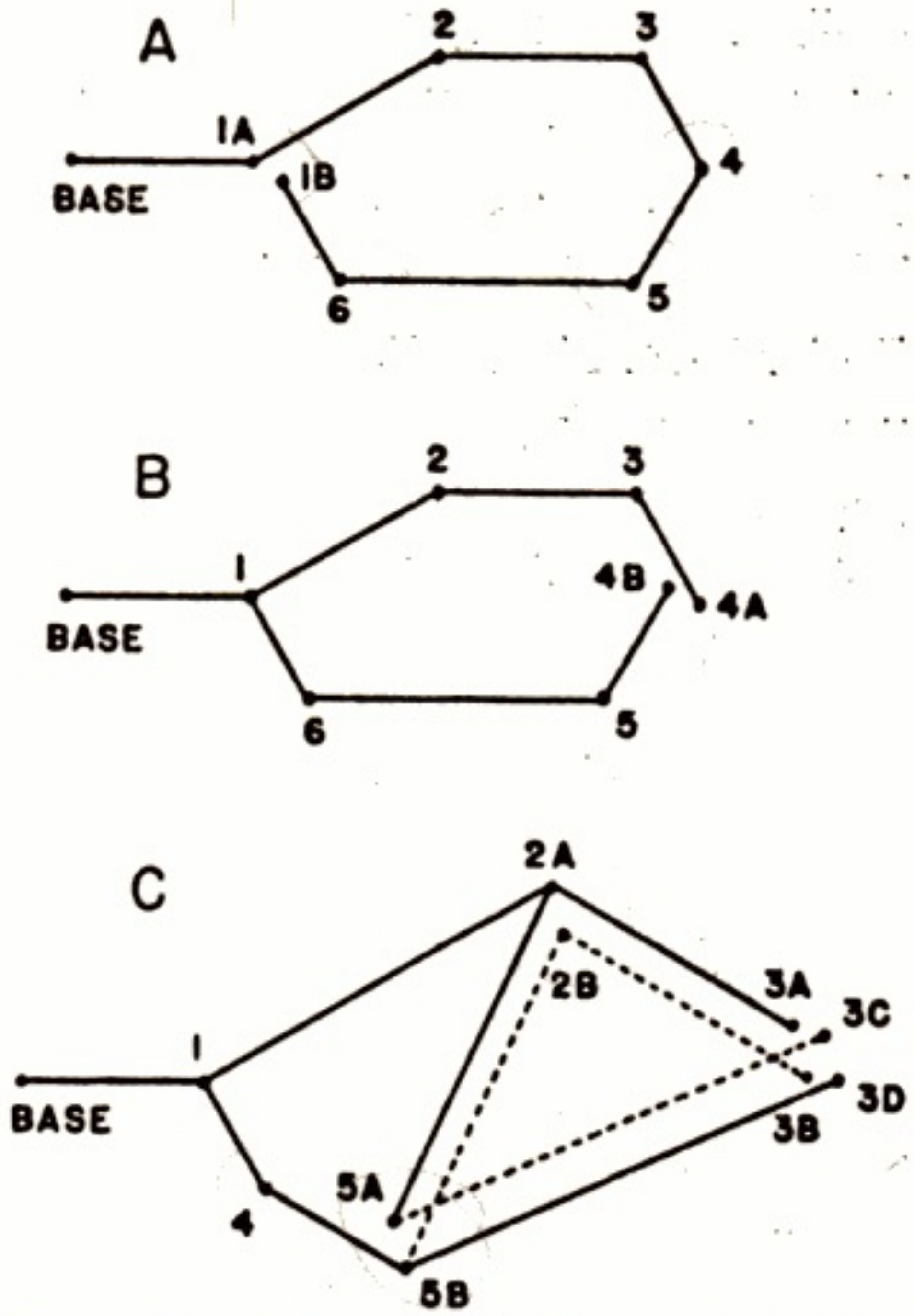

Figure 1. (a) and (b) demonstrate that the presence of a loop produces an ambiguity in the measured location of any station that may be reached from the base station via that loop. In (c), the addition of a second loop produces as many as four independent measured locations for station 3.

than one measured location. This is shown in Figs. 1a and 1b, where a survey of a simple measured loop has been plotted in two different ways in order to show that any station in the loop has been given two different measured locations as a result of measurement errors and the presence of the loop. Fig. 1c shows that adding an additional loop gives some stations within the loops *four* different and independent measured locations, since there are now four different ways of reaching these stations along measured paths without retracing. Each time another loop is added to any path connecting some station to the base station, that station's multiplicity is increased. Hence, in the presence of loops, at least some of the station locations are going to be overdetermined and the method of least squares may be applied.

The actual measurements that are involved in a survey are the sight-vectors that connect the stations. Let us call a measured sight vector connecting stations i and j, $\mathbf{r}_{ij}$. We may denote the most probable value of this vector by $\mathbf{R}_{ij}$, which should obey the vector relation

$$\mathbf{s}_i + \mathbf{R}_{ij} = \mathbf{s}_j \qquad (2)$$

where $\mathbf{s}_i$ and $\mathbf{s}_j$ are vectors giving the most probable locations for the stations i and j. The set of relations of the form of equation (2) for all sights (all measured pairs of i and j) defines the desired survey, with all loops properly closed. The measured and most probable vectors differ by the *deviation*

$$\mathbf{d}_{ij} = \mathbf{r}_{ij} - \mathbf{R}_{ij} \ , \qquad (3)$$

which is simply the probable error in each measurement. If w_{ij} is the weight assigned to each measured sight vector $\mathbf{r}_{ij}$, we may combine equations (1), (2), and (3) to get

$$\Sigma\, w_{ij}\, d_{ij}^{2} = \sum_{ij\ \mathrm{pairs}} w_{ij}\,(\mathbf{s}_i - \mathbf{s}_j + \mathbf{r}_{ij})^2 \qquad (4)$$

as the relation that must be minimized with respect to the station locations. This is accomplished in the usual way by successively taking the partial derivatives of equation (4) with respect to each station vector $\mathbf{s}_i$ and setting each derivative equal to zero. This results in N linear equations with N unknowns, where N is the number of stations. We now need only construct the N by N + 1 matrix representing these linear equations, as in the example given at the end of the paper, to put the problem in a standard form readily used by a computer library subroutine for solving linear simultaneous equations.

We will now derive an algorithm to construct the appropriate matrix from the survey data. Each term in equation (4) will contribute to only two partial derivatives: those with respect to $\mathbf{s}_i$ and $\mathbf{s}_j$. The contributions from each term will be $2\, w_{ij}\, (\mathbf{s}_i - \mathbf{s}_j + \mathbf{r}_{ij})$ to the derivatives with respect to $\mathbf{s}_i$ and $-2\, w_{ij}\, (\mathbf{s}_i - \mathbf{s}_j + \mathbf{r}_{ij})$ to the derivative with respect to $\mathbf{s}_j$. The factor of 2 is present in every contributed term and may be eliminated immediately. Bear in mind that the derivative with respect to $\mathbf{s}_i$ is represented by row i of the matrix and the coefficient of $\mathbf{s}_i$ in each derivative is represented by column i. Column N + 1 contains the negative of the constant terms that are not coefficients of an $\mathbf{s}$ vector. The negative sign arises because these constant terms are standardly placed to the right of the equals sign, as in the example. Hence each term of equation (4) contributes to row i (representing the partial derivative with respect to $\mathbf{s}_i$) $+w_{ij}$ to column i, $-w_{ij}$ to column j, and $-w_{ij}\mathbf{r}_{ij}$ to column N + 1. To row j it contributes $-w_{ij}$ to column i, $+w_{ij}$ to column j, and $+w_{ij}\mathbf{r}_{ij}$ to column N + 1. This procedure is repeated for each term in equation (4) to complete our construction of the matrix representing the N simultaneous equations. Note that since we are dealing with vectors, we have actually constructed three matrices, one for each cartesian coordinate.

We may restate the algorithm we have devised in a more directly applicable way. Let us assume we have M sights locating N stations in the survey. We first convert the measured sight vectors to cartesian coordinates. Then we construct three empty matrices (one for each of x, y, and z coordinates) with N rows and N + 1 columns. Next a weighting factor w_{ij} (see below) is assigned to each measured sight $\mathbf{r}_{ij}$, where the sight was made from station i to station j. For each sight we do the following:

1. Add $+w_{ij}$ to (i, i) and (j, j)
2. Add $-w_{ij}$ to (i, j) and (j, i)
3. Add $w_{ij}x_{ij}$ to (i, N + 1)
4. Add $-w_{ij}x_{ij}$ to (i, N + 1)

Carrying out the same procedure for the y and z matrices using y_{ij} and z_{ij} instead of x_{ij} completes the construction of the three matrices. The solution of the simultaneous equations that they represent yields the most probable locations for the N stations, which is the desired result.

Constraints

At least one constraint, the establishment of a base station, must be applied before a unique solution to the above is possible. In the simultaneous equations, this is accomplished simply by replacing the $\mathbf{s}$ vector for the constrained station by its assigned value and eliminating the partial derivative with respect to that station location vector, since it is no longer a variable. In the matrix formulation the following algorithm does the same thing.

To constrain station i to the value $s_i = C$:

1. Multiply column i by the constant C and subtract it from column N + 1.
2. Eliminate row i and column i from the matrix, reducing its order by one.

A dividend of this method is that the above algorithm may be applied successively to accommodate more than one constraint. Hence if a cave has more than one entrance and these entrances have been accurately located on a topo map or by a good transit survey, the underground survey may be constrained to conform to the known entrance locations.

Weighting

The simplest method of weighting assigns to each sight a weight that is inversely proportional to its absolute length, the tape distance ($|r_{ij}|$), giving

$$w_{ij} = \frac{1}{|r_{ij}|} \quad . \tag{5}$$

This method of weighting has been used with considerable success by the authors in a number of surveys. However, since the same weight is applied independently to each cartesian coordinate, this weighting scheme takes no account of the fact that the actual measurements are taken in spherical coordinates and that these measurements are not made with equal precision. In general, the tape measurement in a Brunton and tape survey does not contribute nearly as much error to a sight as the two angular measurements. It is not possible to take these considerations strictly into account so long as the least squares method is applied independently in each of the three Cartesian coordinates. Attempts to modify the weighting to take some account of what is going on in the other two dimensions (for example, by assigning a weight to each sight in the x dimension matrix that is inversely proportional to the projection of that sight on the yz plane, and similarly for the y and z matrices) have not proved to be worthwhile because there is still no way of coordinating the sense of the corrections made in each dimension.

Some Practical Considerations

At first sight the method outlined here might not seem practical for large surveys since each station adds to the number of simultaneous equations that must be solved. Most large computers have library subroutines that can deal efficiently with only 50 or 100 simultaneous equations, which would represent a rather small survey of an equal number of stations. However, two devices may be employed to increase the capacity of the method many times: (1) "Strings" of sights linking constrained stations, dead ends, and/or junctions of three or more sights can be constructed prior to least squares treatment and used in place of individual sights, eliminating all stations contained within strings. (2) Strings leading to unconstrained dead ends do not participate in the analysis and may be eliminated, to be reintroduced in the final plot of the map.

In this way the number of "stations" (and hence the number of simultaneous equations) can be reduced to the number of loop junctions plus the number of constrained stations. In all but the most complex maze surveys, this number should rarely exceed 50.

Using this approach, the analysis yields the most probable locations for the junction stations. The location of the intermediate stations on each string may be made by proportionally shifting each sight in a string according to its weight to fit the established junctions and constraints. Particular care must be exercised in assigning weights to strings, as our previous considerations apply only to individual sights. For strings of sights, equation (5) must become

$$w_{\text{string}} = \frac{1}{\Sigma\,|r_{ij}|} \quad , \tag{6}$$

where the sum is taken over all the sights in that string.

One of the authors (V.A.S.) has written computer programs that utilize the methods outlined here. They have proven to be practical and highly efficient, as evidenced

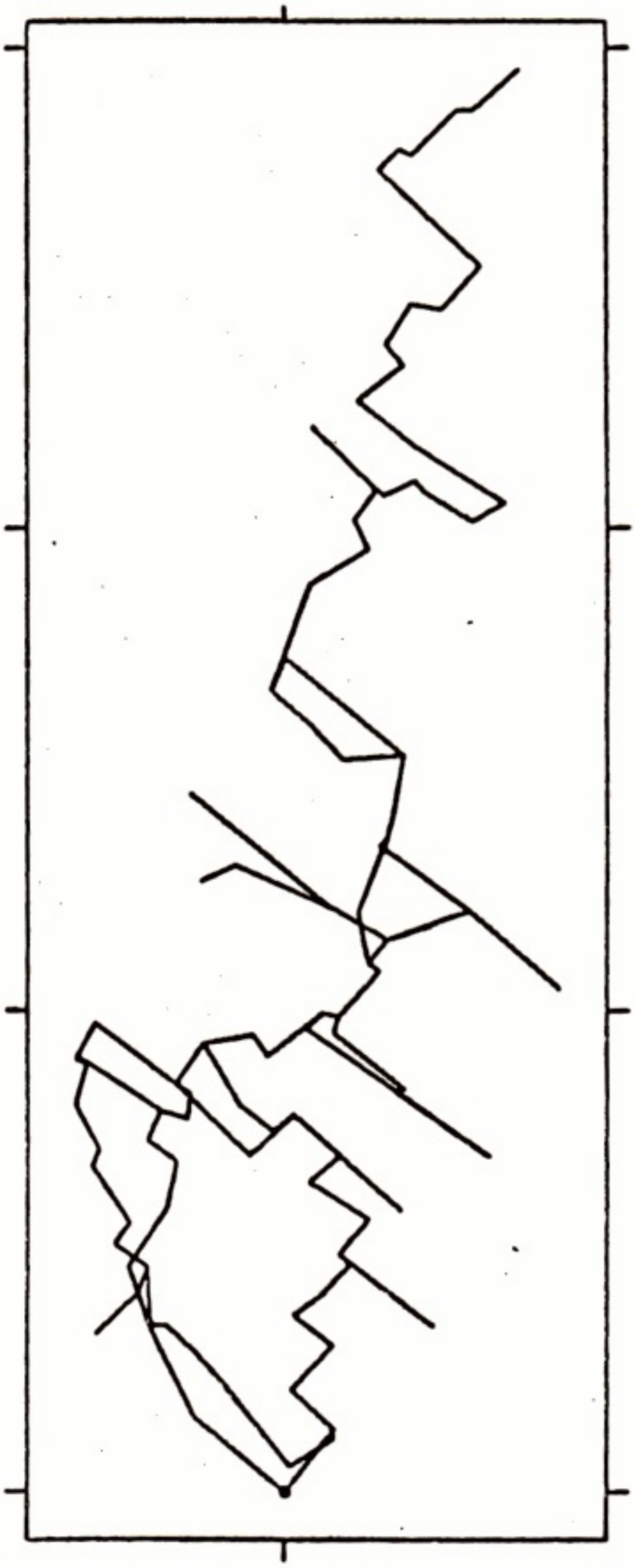

Figure 2. Bear Cave, Pennsylvania. This is the fully corrected output of a computer program utilizing the least squares method described in this paper. The tick marks along the border denote intervals of 200 feet. The diagram was drawn by a Cal Comp x-y plotter attached directly to the computer.

in Figs. 2 and 3, which are reproduced directly from Cal Comp plots drawn by the program itself. Bear Cave, Pa., shown in Fig. 2, is a survey of 2590 feet with 105 sights and 8 loop closures. The corrected survey was produced by an IBM 7090 computer in 14.8 seconds of execution time, exclusive of plotting time. The program will accept up to 1000 sights in up to 50 strings. The efficiency of this least squares method is due to its use of an analytical rather than an iterative approach, allowing the analysis to be made in a single pass.

A few words of caution should be extended concerning this method. First, application to a survey will not magically increase its accuracy to any great extent unless a large number of closures are present throughout the survey. For most rough surveys it should be regarded primarily as a sensible method for distributing errors to yield a self-consistent survey. Second, gross mistakes in a survey will be assimilated by this method along with normal measurement errors, possibly producing an inaccurate and misleading survey. Hence it is always advisable to look at an uncorrected plot as well as the corrected final result to see if any closures are badly out of line, indicating the need for a resurvey.

The least squares method outlined above may be applied, at least in principle, to any survey or measured network of vectors. In fact, a slightly modified form has been applied to high precision geodetic triangulation surveys for at least 60 years[1]. Until the advent of high-speed computers, however, the great amount of computation required by the method restricted its use to first-order triangulation nets for geodetic control purposes. It is seldom used for lower precision surveys, since even in these surveys, the normal closure errors are kept so small as to be infinitesimal on the scale used for the plot. It is precisely because the closure errors in cave or geological surveys appear

[1] For a detailed discussion of the application of the method of least squares to high-precision surveys, see Durgin and Sutcliffe (1927) and Adams (1915).

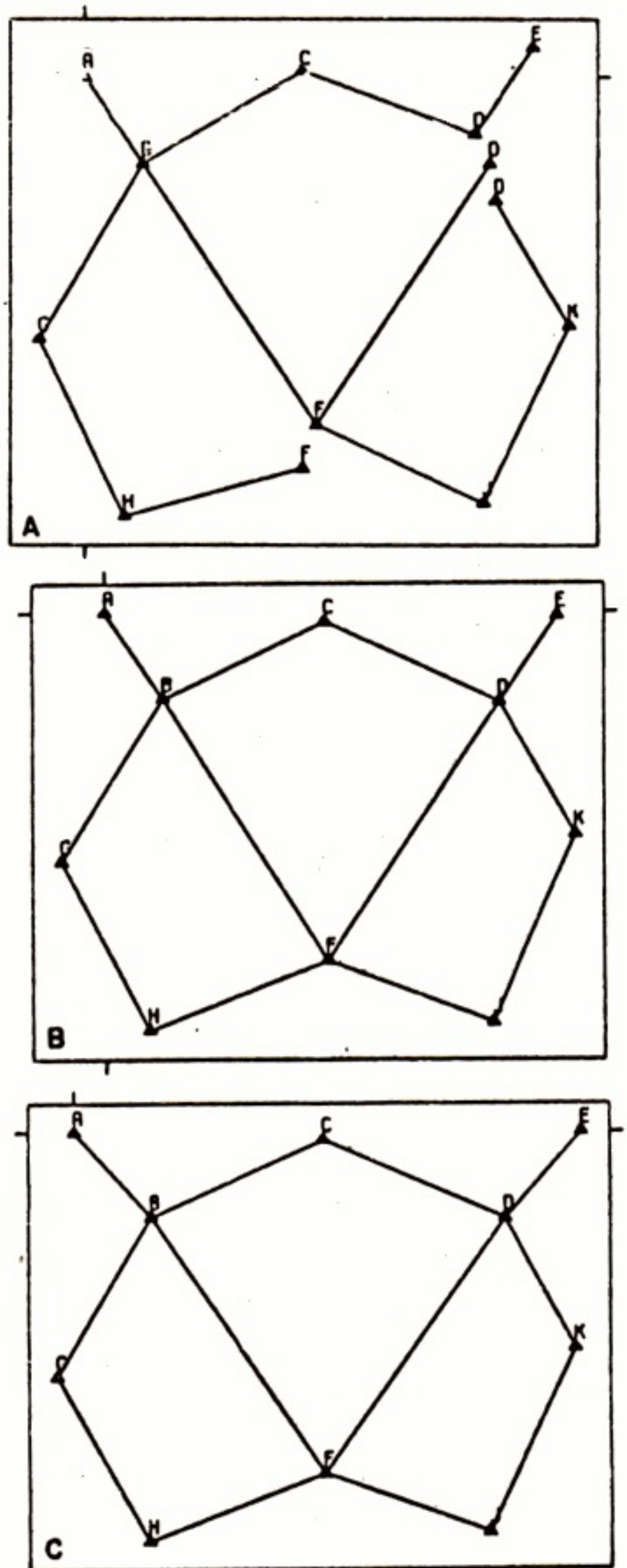

Figure 3 (to left). Pretzel Cave demonstrates the action of the least square method. (a) shows the uncorrected plot of the raw data. Note that the three loops fail to close. In (b) the least squares method has been applied to produce a self-consistent survey with errors distributed throughout the survey in a sensible manner. In (c) an additional constraint has been applied to station E, moving it farther to the right. The program has adapted the survey to this change with a minimum of distortion.

quite substantial at commonly used scales (50′/inch or 100′/inch) that the use of the method becomes worthwhile for these applications.

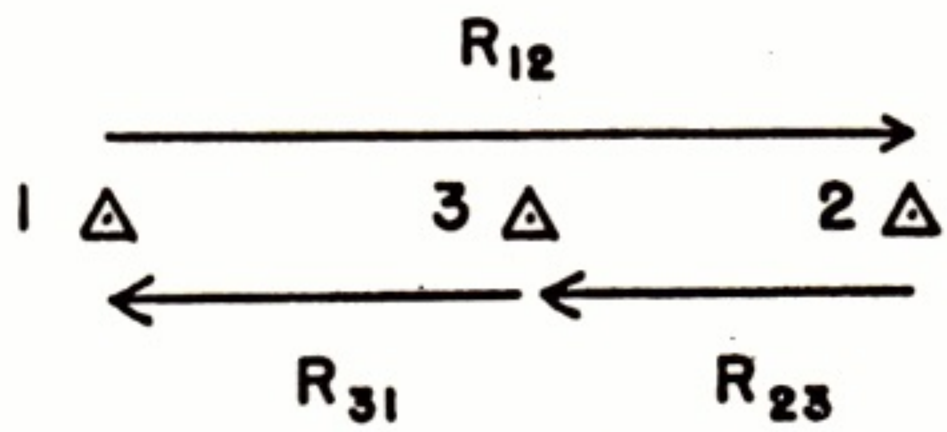

Figure 4. A simple one-dimensional loop survey for testing the algorithm derived in this paper.

Example

An extremely simple one-dimensional survey of three sights forming one loop (see Fig. 4) will serve to illustrate the important features of the least squares method. The sights are:

$$r_{12} = 48.0 \qquad w_{12} = 1/48.0$$
$$r_{23} = -20.5 \qquad w_{23} = 1/20.5$$
$$r_{31} = -26.7 \qquad w_{31} = 1/26.7$$

which produce a closure error of 0.8 feet. So simple an example could be dealt with by inspection, since in closing the sole loop we may distribute the error into two equal portions, one in the positive sight and the other in the two negative sights, yielding an adjusted position for station 2 at 47.6 feet. Nonetheless, we will apply the full least squares method for illustrative purposes.

Equation (4) becomes for this example

$$\Sigma\, w_{ij}\, d_{ij}^2 = w_{12}\,(s_1 - s_2 + r_{12})^2 + w_{23}\,(s_2 - s_3 + r_{23})^2 + w_{31}\,(s_3 - s_1 + r_{31})^2 \;. \tag{7}$$

There are three station locations, so we take the partial derivative with respect to each, yielding

$$\begin{aligned} \frac{1}{2}\,\frac{\delta}{\delta s_1}\,\Sigma\, w_{ij}\, d_{ij}^2 &= w_{12}\,(s_1 - s_2 + r_{12}) - w_{31}\,(s_3 - s_1 + r_{31}) = 0 \\ \frac{1}{2}\,\frac{\delta}{\delta s_2}\,\Sigma\, w_{ij}\, d_{ij}^2 &= -w_{12}\,(s_1 - s_2 + r_{12}) + w_{23}\,(s_2 - s_3 + r_{23}) = 0 \\ \frac{1}{2}\,\frac{\delta}{\delta s_3}\,\Sigma\, w_{ij}\, d_{ij}^2 &= -w_{23}\,(s_2 - s_3 + r_{23}) + w_{31}\,(s_3 - s_1 + r_{31}) = 0 \;. \end{aligned} \tag{8}$$

Rearranging terms, we get

$$\begin{aligned} (w_{12} + w_{31})\, s_1 \quad - w_{12}\, s_2 \quad - w_{31}\, s_3 &= -\, w_{12}\, r_{12} + w_{31}\, r_{31} \\ -\, w_{12}\, s_1 + (w_{12} + w_{23})\, s_2 \quad - w_{23}\, s_3 &= w_{12}\, r_{12} - w_{23}\, r_{23} \\ -\, w_{31}\, s_1 \quad - w_{23}\, s_2 + (w_{23} + w_{31})\, s_3 &= w_{23}\, r_{23} - w_{31}\, r_{31} \end{aligned} \tag{9}$$

which may be represented by the 3 by 4 matrix shown in Fig. 5. The reader may easily verify that the algorithm given earlier suffices to construct this same matrix directly. Assigning station 1 as the base station at the origin, we may let $s_1 = 0$ and eliminate the first of equations (9). This is equivalent to eliminating the first row and the first column in the matrix. The remaining two simultaneous equations are readily solved to yield

$s_2 = 47.6$ and $s_3 = 26.9$,

in accord with our earlier estimate.

$W_{12} + W_{31}$	$-W_{12}$	$-W_{31}$	$-W_{12}R_{12} + W_{31}\ R_{31}$
$-W_{12}$	$W_{12} + W_{23}$	$-W_{23}$	$+W_{12}R_{12} - W_{23}\ R_{23}$
$-W_{31}$	$-W_{23}$	$W_{23} + W_{31}$	$+W_{23}R_{23} - W_{31}\ R_{31}$

Figure 5. The matrix denoting the simultaneous equations to be solved for the survey given in Fig. 4.

ACKNOWLEDGEMENTS

The authors wish to express their appreciation to the Computer Centers of Carnegie-Mellon University and the University of Pittsburgh, who provided computer time and services that were invaluable in the development of this project. We are also indebted to Prof. J. M. Hutt of the Civil Engineering Department of the University of Pittsburgh for reviewing the manuscript. He also led us to the original applications of least squares to surveying methods outlined in the references cited in the footnote.

LITERATURE CITED

Adams, O. S.
1915 Application of the theory of least squares to the adjustment of triangulation. U. S. Coast and Geodetic Survey, Spec. Publ. 28.

Durgin, C. M., and W. D. Sutcliffe
1927 Manual of first order traverse. U. S. Coast and Geodetic Survey, Spec. Publ. 137, pp. 88-96.

Young, H. D.
1962 Statistical treatment of experimental data. McGraw-Hill Publ. Co.

Manuscript received by the editor October, 1969