SIMULTANEOUS MULTIPLE LOOP CLOSURES FOR CAVE SURVEYS: A COMPUTER PROGRAM

NEIL I. SMITH*
Cave Exploration Group (S.A.) Inc.

Abstract

The mathematics of a new least-squares procedure for cave survey loop closures is presented. Previously-published methods are briefly reviewed and it is suggested that the new method is superior, in both single- and multiple-loop situations. Brief description is given of a FORTRAN program for survey reduction incorporating this algorithm.

INTRODUCTION

The loop closure problem will be familiar to all cave surveyors. However careful we may be in making measurements, it is most unlikely that a survey station reachable by more than one path will be positioned unambiguously to within the plotting accuracy. Before the final plot is drawn it is therefore necessary to "adjust" the survey.

Closed loops (with a good closure scheme) actually improve the positional accuracy of the survey in the same sense that the average of repeated measurements of a single quantity is more reliable than one estimate. Of course, if the closure error is large we would not (or certainly should not) attempt to perform the closure - the data needs to be checked for such gross mistakes as reversed bearings, etc., and if all else fails a re-survey may be necessary.

Methods in common use for resolving closures include simple adjustment by eye, and separate distribution of X, Y and Z errors in proportion to leg lengths. Ellis (1976) suggests distributing errors equally between legs since this is the simplest method and, he claims, as likely to produce reasonable results as any other. If there are several interconnected loops, however, the simple methods require that closures be adjusted sequentially. This has the disadvantage that the early closures may force errors onto the remaining loops over and above those due to mere measurement inaccuracies. Precisely for this reason, most surveyors choose to adjust the "best" closure first.

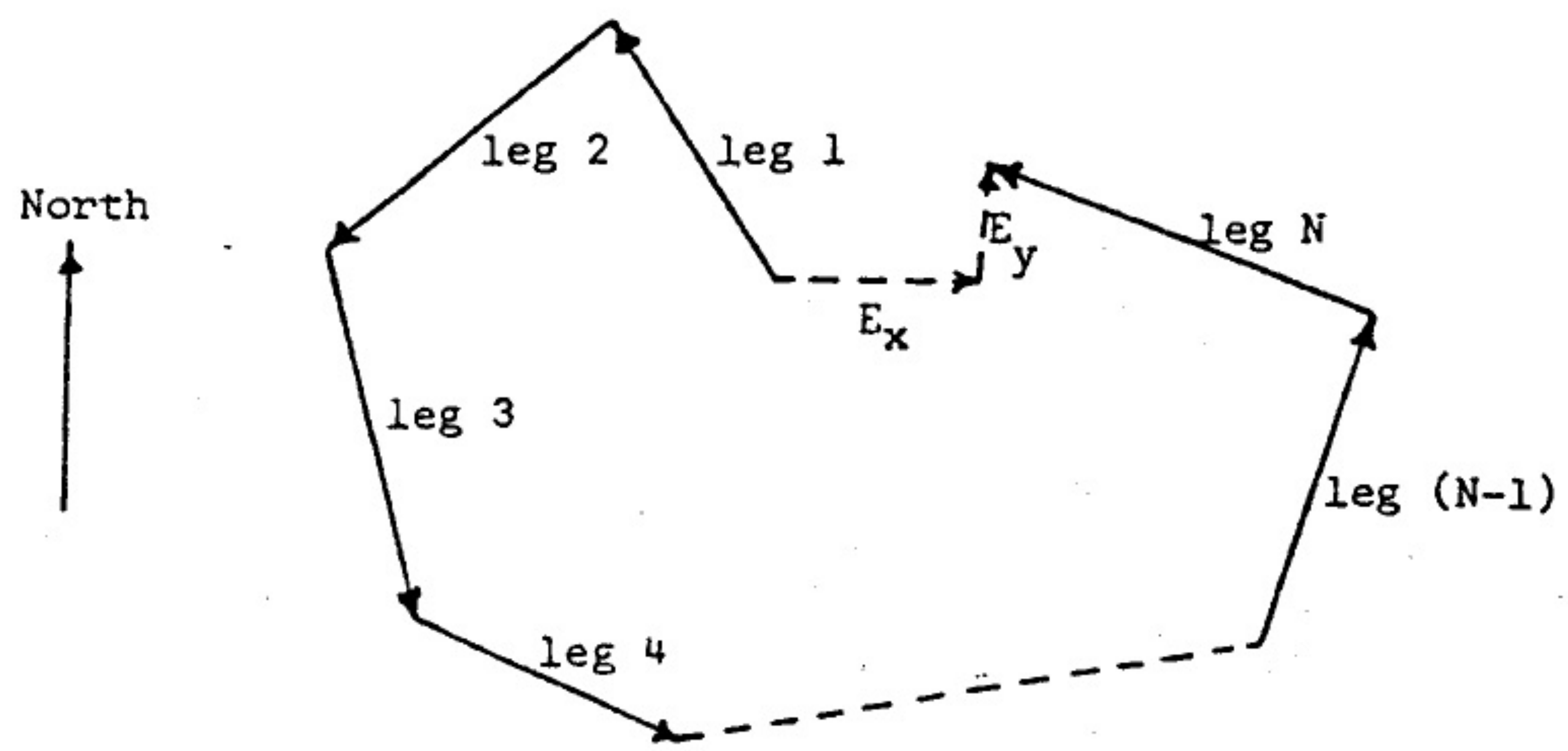
The purpose of this paper is to show that the closure problem may be put on a firm mathematical basis. The problems attending sequential closure are avoided because the opportunity exists for adjusting any number of loops simultaneously. At least two papers (Schmidt & Schelleng, 1970; Luckwill, 1970) have treated the simultaneous closure problem. I claim some superiority for the new method, for reasons to be discussed later. Many cave surveyors are already using computers for data reduction - why not get the most out of them?

MATHEMATICS OF LOOP CLOSURE

For simplicity, we consider at first a survey traverse of N legs forming a single closed loop, as shown in plan view in Fig. 1. The generalisation to

SMITH – SIMULTANEOUS MULTIPLE LOOP CLOSURES

Figure 1. Example of uncorrected survey plot.



other situations is straight forward and is explained later. Due to unavoidable errors in the measurements, the last point does not coincide with the first; instead there is an error which in general has an easterly component E_x, a northerly component E_y, and (although not shown in Fig. 1) a vertical component E_z. The values of the error components may be readily calculated without plotting. Let us assume that for each leg we have measured the length L, the bearing B (clockwise from north, in degrees) and the slope or elevation S (in degrees, relative to the horizontal).

Distinguishing between the legs by a subscript i, the easterly (X), northerly (Y) and vertical (Z) components of each leg are given by straight forward trigonometry as:

$$X_i = L_i \sin B_i \cos S_i \qquad \ldots\ldots (1)$$

$$Y_i = L_i \cos B_i \cos S_i \qquad \ldots\ldots (2)$$

$$Z_i = L_i \sin S_i \qquad \ldots\ldots (3)$$

Negative values indicate, of course, west, south or down. (More usual mathematical notation would employ r, θ and ϕ, with the azimuth θ being taken anticlockwise from east, and θ and ϕ would be measured in radians. However, the equations are just as readily derivable in terms of the "geographical" angles.)

Provided that all legs are taken in the same sense around the loop, the error components are simply the sums of the corresponding leg components, for example:

$$E_x = \sum_{i=1}^{N} L_i \sin B_i \cos S_i \qquad \ldots\ldots (4)$$

our aim in "adjusting" the survey is to introduce small changes l_i, b_i and s_i some or all of the L_i, B_i and S_i in such a way that the error components are reduced to zero. The smaller the changes that are required, the more appealing the adjustment will be to the surveyor. The problem is mathematically tractable if we choose these adjustments in the way which minimises a "weighted least-squares objective function" F, defined as:

$$F = \sum_{i=1}^{N} (w_{li} l_i^2 + w_{bi} b_i^2 + w_{si} s_i^2) \qquad \ldots . (5)$$

The w's are weighting factors which may be freely chosen to reflect the realities of the survey. For example, we may consider an error in bearing of 1° to be as likely as a 0.1 m error in length, and make $w_{li} = 100\ w_{bi}$ so that each would contribute equally to the value of F. The appearance of the subscript i implies that different weights may be given to different legs, a useful feature if some of the surveying was done accurately with (say) a miner's dial and some only with hand compasses. Some measurements may be considered "perfectly" accurate if desired - that is, some l, b or s values may be defined as zero and simply omitted from equation (5).

Considering now any one leg (dropping the subscript i), the increments l, b and s will introduce corresponding changes x, y and z to the easting, northing and height differences. Provided that the increments are small enough, approximate formulae for these changes may be obtained by partial differentiation of equations (1), (2) and (3), that is

$$x = \frac{\partial X}{\partial L} l + \frac{\partial X}{\partial B} b + \frac{\partial X}{\partial S} s \qquad \ldots . (6)$$

and similarly for y and z. Explicitly,

$$x = l \sin B \cos S + brL \cos B \cos S - srL \sin B \sin S \qquad . . (7)$$

$$y = l \cos B \cos S - brL \sin B \cos S - srL \cos B \sin S \qquad . . (8)$$

$$z = l \sin S + srL \cos S \qquad \ldots . (9)$$

The factor r, equal to $\pi/180$, appears because b and s are in degrees. Geometrical demonstrations of these formulae are possible, and illustrate that they are more accurate the smaller the increments. For length changes only (b = s = 0) the formulae are exact.

To "close the loop" we want the net effect of the changes x, y and z summed around the loop to be equal and opposite to the original error components E_x, E_y and E_z, while simultaneously making F as small as possible. That is, we have an "equality constrained" optimisation problem, the three constraint equations being

$$\sum_{i=1}^{N} x_i + E_x = 0 \qquad \ldots (10)$$

$$\sum_{i=1}^{N} y_i + E_y = 0 \qquad \ldots (11)$$

$$\sum_{i=1}^{N} z_i + E_z = 0 \qquad \ldots (12)$$

SMITH – SIMULTANEOUS MULTIPLE LOOP CLOSURES

To solve such a problem we define a "Lagrange function" G (see any general text on optimisation theory, for example, Gottfried & Weisman, 1973), involving the objective function F, the constraints, and one new variable p for each constraint, that is,

$$G = F + p_x\left(\sum_{i=1}^{N} x_i + E_x\right) + p_y\left(\sum_{i=1}^{N} y_i + E_y\right) + p_z\left(\sum_{i=1}^{N} z_i + E_z\right) \quad \ldots \quad (13)$$

The solution is obtained by simultaneously equating to zero the partial derivatives of G with respect to all the variables (the original l's, b's and s's, and the p's). Substituting for F from equation (5) and for x_i, y_i and z_i from equations (7), (8) and (9), and performing the differentiation, we get:

$$\frac{\partial G}{\partial l_i} = 0 = 2\,w_{li}l_i + p_x \sin B_i \cos S_i + p_y \cos B_i \cos S_i + p_z \sin S_i \quad . \quad (14)$$

$$\frac{\partial G}{\partial b_i} = 0 = 2\,w_{bi}b_i + p_x rL_i \cos B_i \cos S_i - p_y rL_i \sin B_i \cos S_i \quad \ldots \quad (15)$$

$$\frac{\partial G}{\partial s_i} = 0 = 2\,w_{si}s_i - p_x rL_i \sin B_i \sin S_i - p_y rL_i \cos B_i \sin S_i + p_z rL_i \cos S_i \quad \ldots \quad (16)$$

The partial derivatives with respect to the p's yield the original constraint equations (10), (11) and (12).

We thus have a set of 3N + 3 simultaneous linear equations to solve. This would certainly be an unpleasant task by hand calculator. At first sight it would seem to tax the capabilities of most computers also - if there were 100 legs in the loop, over 90 000 memory locations would be needed just to store the matrix of coefficients. However, the situation is not as bad as that. The coefficient matrix is symmetric and (better still) is sparse. A set of very simple transformations reduce the problem to that of solving three simultaneous equations, followed by a straightforward back-substitution to produce the l's, b's and s's. In fact, hand caculation would be feasible for a single loop, even with 100 legs.

Because the equations (7), (8) and (9) are not exact, this process does not solve the closure problem exactly, but will come extremely close. An exact closure (with a very slightly sub-optimal value of F) may be obtained by a second application of the process, this time allowing only leg lengths to vary (which they will do by miniscule amounts).

GENERALISATIONS

So far we have considered a single survey loop, allowed changes in all the measured variables, and required the loop to close in all three dimensions. Very simple modifications take care of other situations. First, if we regard some measurements as much more reliable than others we can adjust the closure

SMITH – SIMULTANEOUS MULTIPLE LOOP CLOSURES

Figure 2. Illustrating simultaneous loop closure.

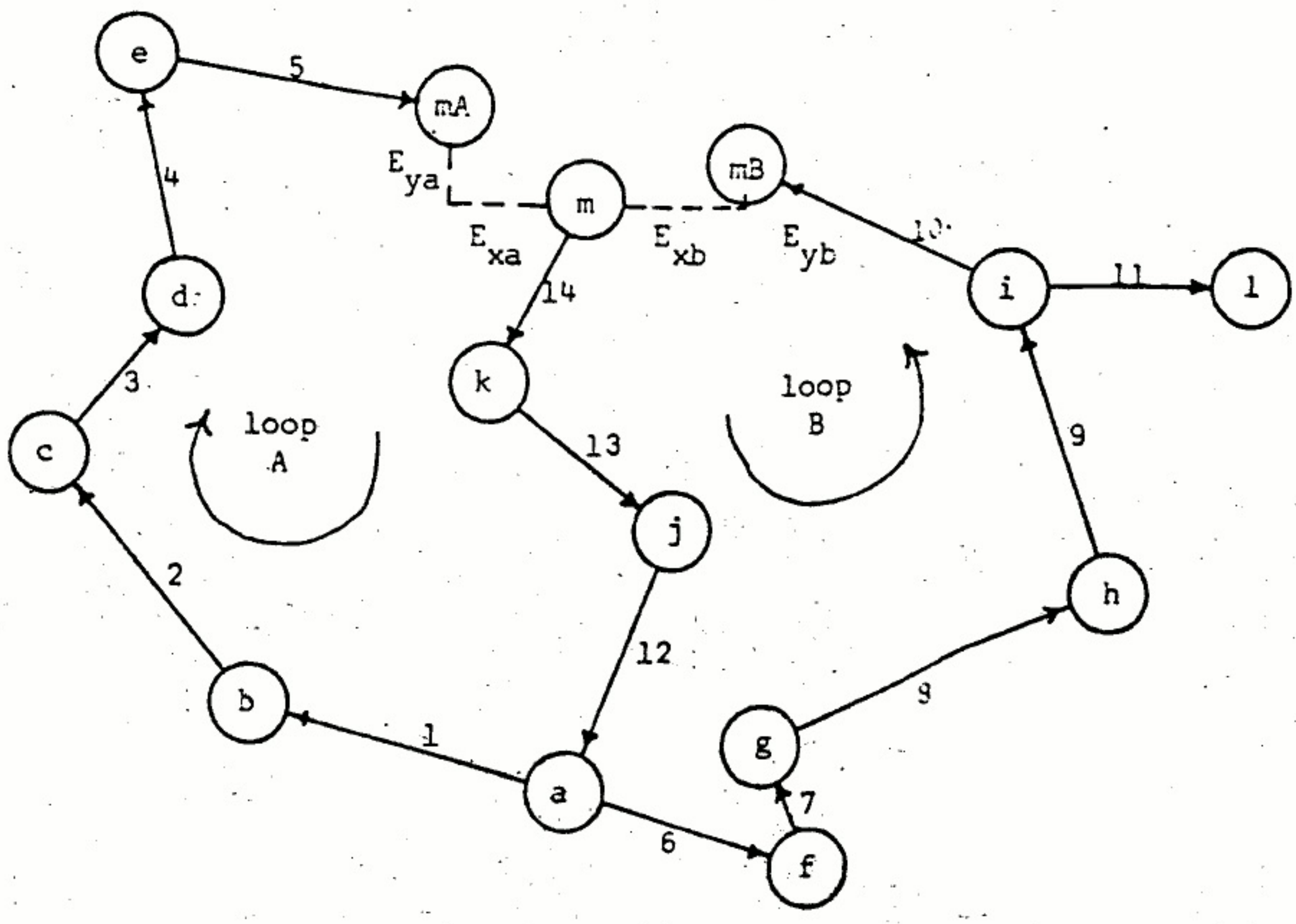
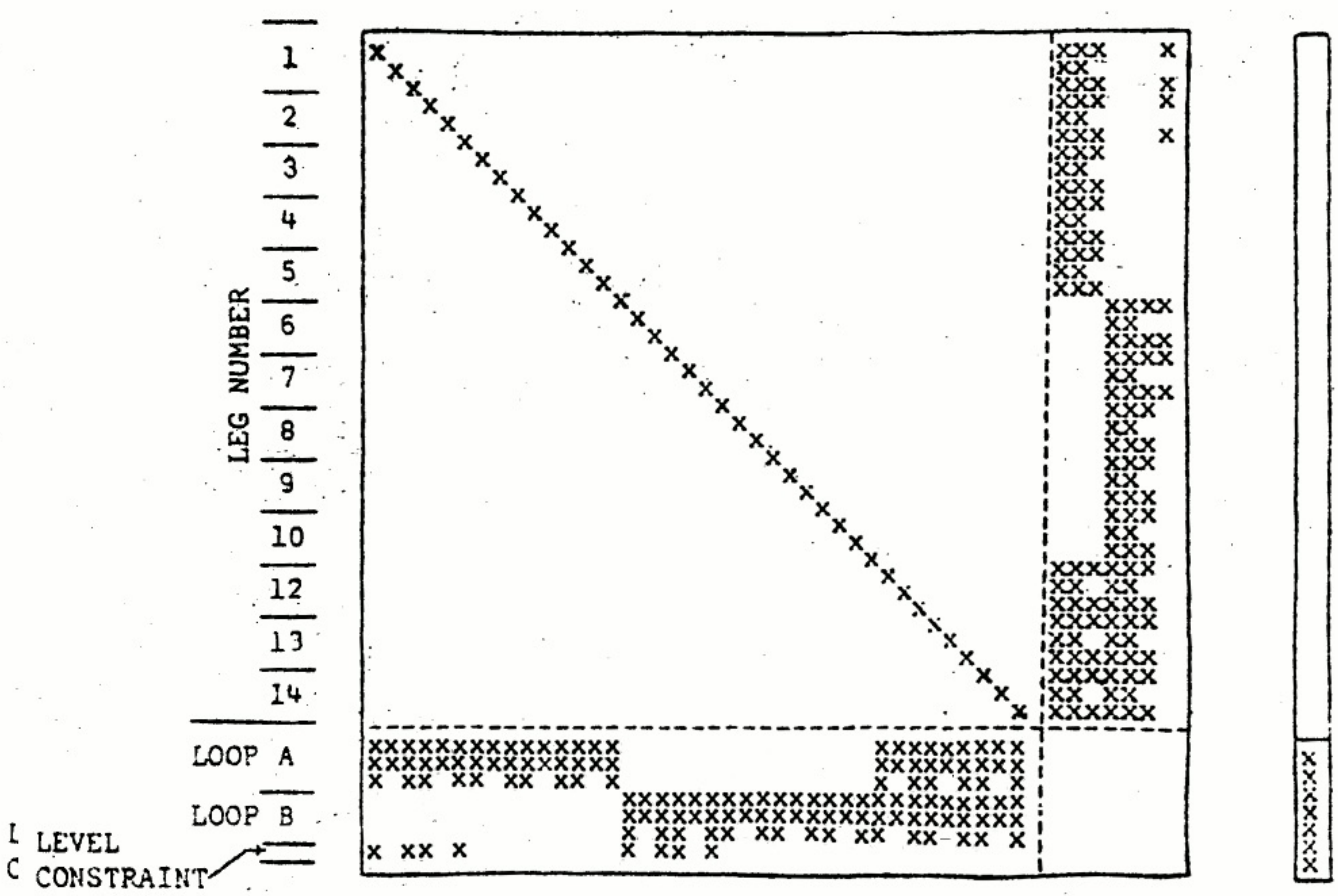


Figure 3. Example of Coefficient Matrix and R.H. Side

This applies to the two-loop example of Fig. 2.
x Denotes non-zero entry.



by varying only the latter - this leads to equations of the same form but fewer of them. Second, we may require closure in only one or two dimensions, for example, two stations may be spatially separated but both on the water table - there would be a z constraint (equation (12)) but no x and y constraints; or an electromagnetic ("RDF") measurement may constrain a station in x and y but not in z.

The generalisation to multiple loops is also straightforward and is illustrated in Fig. 2, where legs are numbered and stations identified by letters. Stations mA and mB are physically the same as station m. We still wish to minimise F as defined by equation (5), where the summation is over the entire survey and quite unrelated to the number of loops. Of course, if a leg forms part of no loop, we have no reason to adjust it at all, and so we do not include it. We now have six constraints because we wish to cancel simultaneously the errors E_{xa}, E_{ya}, E_{za}, E_{xb}, E_{yb} and E_{zb}. A change to (say) leg 10 would have no effect on the closure of "loop A" - thus the summation in the constraint equations (10), (11) and (12) is not over the whole survey but only around the relevant loop. For example, one constraint would be:

$$x_{14} + x_{13} + x_{12} + x_1 + x_2 + x_3 + x_4 + x_5 + E_{xa} = 0 \qquad \ldots (17)$$

Only legs 12, 13 and 14 are common to both loops and so appear in the equations for both E_{xa} and E_{xb}.

As a further illustration, suppose that we know in addition that stations g and c are at the same level although the raw survey would place c 0.3 metre higher. We would then have a seventh constraint, namely,

$$-z_7 - z_6 + z_1 + z_2 + 0.3 = 0 \qquad \ldots (18)$$

(the minus signs occurring because we are looking at legs 6 and 7 in the reverse sense). Fig. 3 shows the coefficient matrix for this example. Simple transformations reduce the problem to the solution of seven simultaneous equations. If we were to consider instead of "loop A" and "loop B" just one of them and the entire exterior loop, the equations would look different, but the solution would be the same.

COMPARISON WITH OTHER METHODS

Schmidt & Schelleng (1970) have published a least-squares method for simultaneous closures. They minimise not the adjustments to lengths, bearings and slopes (the actual measurements made) but the changes to the Cartesian components (X, Y and Z) of the survey legs. The authors admit that they do not have the control they would like over bearings, for example, some of which can end up in the adjusted version to be uncomfortably different from their measured values.

The method requires the solution of k simultaneous equations (k = number of stations) for each coordinate, and the authors employ the idea of a "string" of legs to reduce the coefficient storage requirements. Luckwill (1970) describes (although not fully) what is potentially a better method, because it requires only one equation for each loop. If weighting factors are introduced, Luckwill's method yields exactly the same results as Schmidt and Schelleng's but with much less effort - it solves a related "dual" problem which is easier because there are always fewer loops than stations.

The principal advantage claimed for the method presented here is the explicit formulation in terms of the actual measured quantities - length, bearing and slope. During the adjustment process none of these values can wander further from their measured values than absolutely necessary. The small price to pay is some elementary trigonometry, and the solution of three equations for each loop (once) rather than one equation for each loop (three times). The computer is not worried.

A COMPUTER PROGRAM

A complete survey data reduction program has been written (in FORTRAN) incorporating the above. It has turned out to be quite long (23 pages of listing) not because of the mathematics but because of requirements for generality in entering data and in setting up the equations for all possible constraint and weight combinations.

In summary, the first action of the program is to read in the coordinates of any stations which were finalised on earlier runs, if any. New raw survey data is then entered in the form "station from", "station to", length, bearing, slope and instrument height. Alternatively, the program can accept stadia data or any mix of the two types. Compass and clinometer calibration data, etc., is entered on special cards and applies to all raw data following until overridden by a new value indicating, perhaps, a different instrument. An important feature is that stations reachable by multiple paths must be given more than one name (as with station m in Fig. 2). To the computer, then, there are no closed loops and no ambiguities in station positions, which are calculated and printed out to allow the surveyor to decide whether immediate closure is justified or whether the data needs re-assessment.

The program then enters a "command" mode. Commands include equating stations to each other (that is, closing loops) in one, two, or three dimentions, and forcing station coordinates to assume certain values. There is no ambiguity in choosing the legs involved in the closure: to the program there are no loops and so only one path. It just "happens" that two "different" stations wind up with the same coordinates. The command structure allows closures to be all simultaneous, all sequential, or any intermediate combination. Another useful command produces a print-out of coordinates for a map at any scale. Plotting would be another possibility but has not been included at this stage. Commands may be stacked on the input file after the data, but if a remote terminal is available the user may enter them interactively.

REFERENCES

Ellis, B.M. et al. (1976). *Surveying Caves*. British Cave Research Association: p. 18.

Gottfried, B.S. & Weisman, J. (1973). *Introduction to optimization theory*. Prentice-Hall, Englewood Cliffs. p. 44-45.

Luckwill, M. (1970). Lecture notes for the C.R.G. symposium on cave survey. *Trans. Cave Res. Group Gt Britain* 12(3):227-229.

Schmidt, V.A. & Schelleng, J.H. (1970). The application of the method of least squares to the closing of multiply-connected loops in cave or geological surveys. *Bull. Nat. Speleol. Soc.* 32(3):51-58.

I offer the following in answer to your question about transformations":

Consider first the case of three-dimensional closure of a loop of only three legs (i.e. N=3). Application of equations (14), (15) and (16) of the paper to each of the three legs gives us nine equations. In addition there are three constraint equations (one for each dimension); these are equations (10), (11) and (12) of the paper. These <u>twelve</u> simultaneous linear equations in twelve unknowns (l_i, b_i and s_i for i=1,2,3 and the new variables p_x, p_y and p_z) are written out in matrix-vector form in figure A (with this letter). I have used a shorthand notation for some elements of the matrix, namely

$$a_{i1} = \sin B_i \cos S_i$$

$$a_{i2} = \cos B_i \cos S_i$$

$$a_{i3} = \sin S_i$$

$$a_{i4} = r\ L_i \cos B_i \cos S_i$$

$$a_{i5} = -\ r\ L_i \sin B_i \cos S_i$$

$$a_{i6} = -\ r\ L_i \sin B_i \sin S_i$$

$$a_{i7} = -\ r\ L_i \cos B_i \sin S_i$$

$$a_{i8} = r\ L_i \cos S_i$$

for i = 1,2,3.

All these values are known, being functions of the measured distances, bearings, and slopes.

Although the matrix has 12 x 12 elements (144), it is symmetric and many of the elements are zero. In fact, there are only 33 independent elements (11N elements for general N, 3N diagonal elements and 8N elements in the upper right (or lower left) partition). While we could in principle simply solve this set of 12 equations directly, it is not efficient. Furthermore, when N is increased to a realistic figure the matrix becomes too big even for most computers, e.g. there are 91809 elements for a loop of 100 legs. All this is by way of introduction - I guess you followed thus far from my paper anyhow.

The "transformations" to which I refer are those we used at school for solving sets of two or three simultaneous equations by hand, and are based on the fact that any constant multiple of one equation may be subtracted from any other equation without affecting the solution. Refer again to figure A. The equation I have labelled (10) has the coefficient of l_1 (namely a_{11}) anihilated by subtracting a fraction $\frac{a_{11}}{2\ w_{11}}$ of equation (1) ; none of the other non-zero coefficients are affected but the coefficients of p_x , p_y and p_z (previously zero) become $\frac{-a_{11}^2}{2\ w_{11}}$, $\frac{-a_{11}\ a_{12}}{2\ w_{11}}$, and $\frac{-a_{11}\ a_{13}}{2\ w_{11}}$ respectively. Still looking at equation (10), the next coefficient a_{14} may then be anihilated by subtracting a fraction $\frac{a_{14}}{2\ w_{b1}}$ of equation (2), <u>without affecting the</u>

coefficient of l_1 which we made zero on the last step. Of course, the coefficients of p_x, p_y and p_z are again changed. Continuing this process, equation (10) is modified so that the coefficients of all the l, b, and s variables are made zero. We are left with a linear equation in only 3 unknowns, i.e.

$$M_{11}\,p_x + M_{12}\,p_y + M_{13}\,p_z = -E_x .$$

where

$$M_{11} = -\tfrac{1}{2}\sum_{i=1}^{N}\left(\frac{a_{i1}^2}{w_{li}} + \frac{a_{i4}^2}{w_{bi}} + \frac{a_{i6}^2}{w_{si}}\right)$$

$$M_{12} = -\tfrac{1}{2}\sum_{i=1}^{N}\left(\frac{a_{i1}\,a_{i2}}{w_{li}} + \frac{a_{i4}\,a_{i5}}{w_{bi}} + \frac{a_{i6}\,a_{i7}}{w_{si}}\right)$$

$$M_{13} = -\tfrac{1}{2}\sum_{i=1}^{N}\left(\frac{a_{i1}\,a_{i3}}{w_{li}} + \frac{a_{i6}\,a_{i8}}{w_{si}}\right)$$

In a similar manner, equations (11) and (12) are modified, leaving just 3 equations in 3 unknowns p_x, p_y and p_z, viz.,

$$\begin{bmatrix} M_{11} & M_{12} & M_{13} \\ M_{12} & M_{22} & M_{23} \\ M_{13} & M_{23} & M_{33} \end{bmatrix} \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix} = \begin{bmatrix} -E_x \\ -E_y \\ -E_z \end{bmatrix}$$

Because this set of equations is also symmetric, only 6 (not 9) elements need to be calculated and stored, and a computer library subroutine tailored for solving symmetric systems may be used. A little calculation will demonstrate to you the symmetry of this system, and also produce formulae for the rest of the M's, namely

$$M_{22} = -\tfrac{1}{2}\sum_{i=1}^{N}\left(\frac{a_{i2}^2}{w_{li}} + \frac{a_{i5}^2}{w_{bi}} + \frac{a_{i7}^2}{w_{si}}\right)$$

$$M_{23} = -\tfrac{1}{2}\sum_{i=1}^{N}\left(\frac{a_{i2}\,a_{i3}}{w_{li}} + \frac{a_{i7}\,a_{i8}}{w_{si}}\right)$$

$$M_{33} = -\tfrac{1}{2}\sum_{i=1}^{N}\left(\frac{a_{i3}^2}{w_{li}} + \frac{a_{i8}^2}{w_{si}}\right)$$

In terms of the original 12 by 12 matrix, we can think of the 3 by 9 lower left partition being "emptied" by the transformation process while the 3 by 3 lower right partition is "built up".

Once the values of p_x , p_y and p_z are found, all of the original unknowns may be determined one-at-a-time by back-substitution. For example, l_1 is found from equation (1) :

$$l_1 = - (a_{11} p_x + a_{12} p_y + a_{13} p_z) / 2 w_{11}$$

It will now be apparent to you that the original equations are solved exactly (apart from numeric roundoff) by this process. There is no need to lump terms together as you suggest. The nice thing is that there are only three simultaneous equations to solve regardless of the value of N. Of course, more labour is involved for higher N because more a_{ij} values must be calculated and there are more steps in the summations to generatet the M_{ij}.

Now, looking at figure 3 of my ASF paper, the coefficient matrix in the case of multiple closures is of very similar form and exactly the same sort of transformations are used. In this example there are seven constraints (2 loops to be closed in 3 dimensions, plus one additional closure in z only). Thus we are left with seven simultaneous equations to solve before back-substituting to get the original unknowns. When some legs don't affect particular constraints, we get some zeros in the off-diagonal partitions of the matrix.

Interestingly, Schmidt and Schelleng could have used the same sort of transformations and so avoided having to combine "strings" of survey shots to reduce the size of their matrix. I think that the Schmidt formulation obscures this somewhat because it uses actual station coordinates as variables rather than the X, Y or Z components of legs. However (as you observe), the method of Luckwill with the addition of weights gives identical results to Schmidt's method. Transformations of the type I have described are precisely what Luckwill used to formulate his "normal equations" (his paper doesn't explain this, unfortunately). In his example there are three normal equations because there are three loops; the number of legs in each loop doesn't matter.

What I don't like about Luckwill's method (or Schmidt's) is that closure is performed separately in each Cartesian coordinate. I feel that this sort of adjustment could seriously distort some parts of a cave map. For example, consider a loop which closes perfectly in the easterly direction but poorly in the northing, and includes two legs of equal length, AB bearing 090 degrees and BC bearing 180 degrees. (See figure B herewith). When the loop is closed there will be no easterly adjustments and the northerly adjustments to AB and BC will be equal (with Schmidt's inverse length weighting). Thus AB is distorted almost entirely in bearing whereas BC is changed in length (only). The different treatment results entirely from the difference in true bearing of the legs. This is hardly realistic since the expected magnitudes of measurementx errors will be independent of bearing. Furthermore, in the usual cave survey situation lengths are measured more accurately than bearings. The change in length forced on leg BC is quite undesirable - the northerly misclose is much more likely to be due to compass errors in those legs with substantial easterly components. All this is of course quite well known; Schmidt and Schelleng refer to it (p54, bottom left). I suggest that my method provides the best way of dealing with this problem.

$$
\begin{array}{c}
(1)\\(2)\\(3)\\(4)\\(5)\\(6)\\(7)\\(8)\\(9)\\(10)\\(11)\\(12)
\end{array}
\left[\begin{array}{ccccccccc|ccc}
2w_{l_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{11} & a_{12} & a_{13}\\
0 & 2w_{b_1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & a_{14} & a_{15} & 0\\
0 & 0 & 2w_{s_1} & 0 & 0 & 0 & 0 & 0 & 0 & a_{16} & a_{17} & a_{18}\\
0 & 0 & 0 & 2w_{l_2} & 0 & 0 & 0 & 0 & 0 & a_{21} & a_{22} & a_{23}\\
0 & 0 & 0 & 0 & 2w_{b_2} & 0 & 0 & 0 & 0 & a_{24} & a_{25} & 0\\
0 & 0 & 0 & 0 & 0 & 2w_{s_2} & 0 & 0 & 0 & a_{26} & a_{27} & a_{28}\\
0 & 0 & 0 & 0 & 0 & 0 & 2w_{l_3} & 0 & 0 & a_{31} & a_{32} & a_{33}\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 2w_{b_3} & 0 & a_{34} & a_{35} & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2w_{s_3} & a_{36} & a_{37} & a_{38}\\
\hline
a_{11} & a_{14} & a_{16} & a_{21} & a_{24} & a_{26} & a_{31} & a_{34} & a_{36} & 0 & 0 & 0\\
a_{12} & a_{15} & a_{17} & a_{22} & a_{25} & a_{27} & a_{32} & a_{35} & a_{37} & 0 & 0 & 0\\
a_{13} & 0 & a_{18} & a_{23} & 0 & a_{28} & a_{33} & 0 & a_{38} & 0 & 0 & 0
\end{array}\right]
\left[\begin{array}{c}
l_1\\ b_1\\ s_1\\ l_2\\ b_2\\ s_2\\ l_3\\ b_3\\ s_3\\ p_x\\ p_Y\\ p_z
\end{array}\right]
=
\left[\begin{array}{c}
0\\0\\0\\0\\0\\0\\0\\0\\0\\-E_x\\-E_Y\\-E_Z
\end{array}\right]
$$

FIGURE A.

Original set of Simultaneous Linear Equations, being for a 3-dimensional closure with $N = 3$.

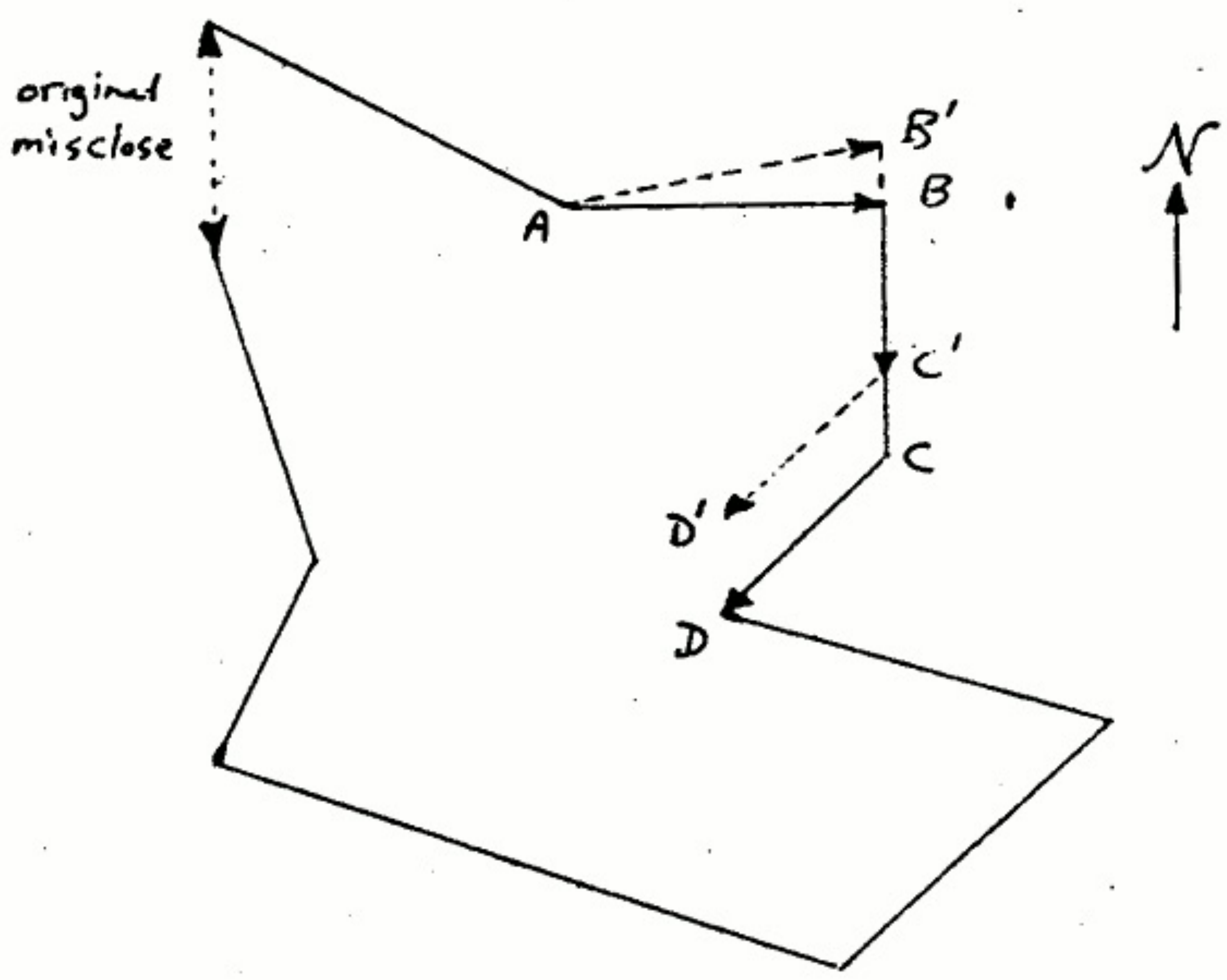


FIGURE B.

Neil I Smith
5 Marine Parade
Seacliff SA 5049

11 March 1983