26. Automatic grid balancing in geophysical survey

J.G.B. Haigh
Department of Mathematics, University of Bradford, Bradford, West Yorkshire BD7 1DP, U.K.

26.1 Introduction

In geophysical survey for archaeological purposes, it is customary to take readings at a large number of stations, distributed over the area of interest as the vertices of a regular square lattice. It is often convenient to divide the region of survey into a number of square grids, and to record each grid separately. The size of grid normally adopted by surveyors from Bradford is 20×20 readings, but other workers prefer to use grids of 30×30 or of 10×10 readings.

There are several good reasons for dividing the survey in this manner. When making earth resistance measurements with a twin-probe array, the mobile probes are connected to the fixed probes through a cable. As the cable is of finite length, the fixed probes must be moved from time to time in order that the survey may continue, and the survey area must be divided into blocks, each of which is accessible from a single location of the fixed probes. When a flux-gate magnetometer is in use, it must be rebalanced and recalibrated occasionally, as a precaution against instrumental drift; the completion of a grid provides a convenient opportunity to recalibrate the instrument.

The use of separate grids generally allows the progress of the survey to be controlled more effectively, with checks on both locational accuracy and on the quality of the results. If instrumental problems occur, they are likely to be confined to a few clearly defined grids, which can then be resurveyed and the new readings can readily replace the corresponding grids in the original survey.

Some surveyors now use instruments mounted on wheeled transport, making it more convenient to take readings in long strips rather than in square grids. The quality of results from such surveys does not yet appear to be as good as that from more traditional methods, using hand-held instruments and working in small grids. Hence the use of square grids seems likely to continue for the foreseeable future.

When the data from a gridded survey are interpreted with the aid of a sensitive graphical display, the boundaries between grids are often clearly apparent as distinct contrasts within the display. The contrasts are usually caused by differences in instrumental calibration between adjoining grids. In magnetometer survey the zero-reading may not have been set to exactly the same value at each recalibration; in twin-probe resistance survey it may not have been possible to establish the same background resistance after the fixed probes had been relocated.

Modern display techniques are capable of making the contrasts glaringly obvious. Fig. 26.1 is a laser printer interpretation of a VGA display in sixteen levels, illustrating a typical magnetometer survey; the contrast between grids has been deliberately exaggerated for illustration purposes. It should be pointed out that similar discrepancies are likely to occur when a survey is carried out in long strips rather then in grids, but they may not be revealed so clearly by standard display techniques.

In order to produce a pleasing display, emphasising archaeological features rather than discrepancies in calibration, it is necessary to balance the grids through zero-point adjustment. An adjustment is made by adding an appropriate constant to each of the readings in one grid. For a small survey, perhaps of a dozen grids or so, a good balance can be usually achieved by empirical estimation of the required adjustments. Nowadays, however, surveys of several hectares are quite common, and data sets may contain hundreds of grids, making it almost impossible to achieve a good balance by empirical judgment. Adjusting a grid to match its neighbour on one edge may upset the balance along the other edges; adjusting the other neighbours may cause imbalances elsewhere, leading to a lengthy cycle which fails to produce a satisfactory result.

The purpose of this paper is to discuss whether it is possible to define an optimum state of balance, in which the adjustments are set to give a minimum mismatch on all the internal edges of the gridded survey. Assuming that such an optimum is possible, is there an algorithm through which it can be achieved? And can the algorithm be offered in a form which is useful to the general archaeological surveyor?

The answer to these questions involves an approach to the problem in two distinct stages. First a quantitative value for the mismatch has to be defined at each of the internal edges. Secondly an adjustment has to be determined for each grid, so that the mismatches are reduced to their optimum values.

26.2 Definition of edge mismatch

Suppose that the survey is made in square grids of $N \times N$ readings, and that there are n grids in the overall survey. Suppose also that grid i and grid j are neighbours, with an internal edge in common. Then grid i has up to N readings along the common edge, and similarly grid j has up to N readings on the other side of the edge. The estimation of the mismatch between the grids involves those readings that actually form pairs on opposite sides of the edge. Let there be M such pairs, where $M<N$, with readings $x_1,\ldots, x_M$ in grid i, and corresponding readings $y_1,\ldots, y_M$ in grid j. These lead to M estimates

$$\delta_k = x_k - y_k\,, \qquad k = 1,\ldots, M,$$

of the mismatch between grid i and grid j.

From these M estimates the mean can be calculated, which is taken to be the actual mismatch between grid i and grid j:

$$d_{ij} = \left(\sum_{k=1}^{M} \delta_k\right) / M .$$

In this definition d_{ij} is supposed to be the mismatch going from grid i to grid j; the mismatch in the reverse direction would be $d_{ji}=-d_{ij}$.

A mismatch which is based on a large number of estimates with comparatively little variation can be regarded as more reliable than one which is based on only a few estimates with large variations. It is therefore necessary to introduce a weighting factor w_{ij}, which gives an indication of the reliability of the mismatch:

$$w_{ij} = M^2 / \left(\sum_{k=1}^{M} (d_{ij} - \delta_k)^2\right).$$

The weighting factor is effectively the inverse square of the standard error, and will be introduced into the calculations of the grid adjustments.

In order to increase the reliability of the results, two minor modifications have been introduced into the definition of d_{ij} and w_{ij}. The first is that any values of the difference δ_k lying more that 2.5 standard deviations from the mean d_{ij} are regarded as outliers. They are removed from the list, the value of M is reduced accordingly, and the values of d_{ij} and w_{ij} are recalculated. This modification prevents any spurious readings, as occasionally occur in geophysical survey, or any sharply changing features, having an undue effect on the final balance.

The second modification is to take account of any overall trend in the geophysical readings in the direction perpendicular to the edge. Instead of taking a simple difference between readings across the grid edge, the linear trend to an edge reading in grid i from the next reading inwards is extrapolated to the mid-point between the grids. A similar extrapolation is made from the two corresponding points in grid j, and the difference is taken between the two extrapolated values. This should be a better estimate than the simple difference, but some surveyors have been found to create an artificial trend by mishandling the instrument at the edge of the grid. The most reliable value for δ_k seems to be a linear combination of the two methods of estimation.

26.3 Applying the grid adjustments

Suppose that a given grid i is situated in the overall survey, with another grid i_1 to the east, and that the mismatch from grid i to grid i_1, as defined in section 26.2, is d_{i1}. Likewise, suppose that grid i_2 is situated to the north with mismatch d_{i2}, grid i_3 to the west with mismatch d_{i3}, and grid i_4 to the south with mismatch d_{i4}. In a general balancing scheme, suppose that an adjustment x_i is applied to grid i, with corresponding adjustments x_{i1} to grid i_1, x_{i2} to grid i_2, x_{i3} to grid i_3, and x_{i4} to grid i_4. When these changes are taken into account, edge j of grid i has an adjusted mismatch:

$$d'_{ij} = d_{ij} + x_i - x_{ij} .$$

Now from the weighted sum of squares S of the mismatches:

$$S = \sum_{ij} w_{ij}(d'_{ij})^2 ,$$

or

$$S = \sum_{ij} w_{ij}(d_{ij} + x_i - x_{ij})^2 .$$

In the summation on the right-hand side, every internal boundary between grids is represented twice, once in a sense from grid i to grid i_j, and once in the reverse sense from grid i_j to grid i. Hence differentiating S with respect to x_i gives

$$\frac{\partial S}{\partial x_i} = 4\sum_{j=1}^{4} w_{ij}(d_{ij} + x_i - x_{ij}) .$$

In order to obtain the optimal values for the adjustments x_i, on the basis of a weighted least-squares criterion, the sum of squares S must be minimised. The minimum is attained when all the derivatives with respect to the parameters x_i are zero, giving

$$i = 1,\ldots,n. \qquad (1)$$

Equation (1) assumes that all four neighbours of grid i exist, and that a non-zero mismatch d_{ij} is defined for each. Should that not be the case for any direction, then the terms for that direction are omitted from equation (1). Since there are n grids in the overall survey, there are n adjustments x_i, and also n simultaneous equations within equation (1).

The system of n equations in n unknowns may be written as a matrix equation

$$Ax = b , \qquad (2)$$

and may in principle be solved to give the required adjustments x_i, making up the vector x. The difficulty is that the matrix A is singular, as can easily be seen by observing that the elements in any row sum to zero. A least-squares problem should have a solution, but the singularity of A indicates that the solution is not unique.

In fact the nature of the underlying problem indicates that the solution cannot be unique. If a set of adjustments is calculated to achieve a balance between grids, and the same constant is added to every adjustment, the result will still be balanced, since every reading has been raised by the same constant. Even though a solution is known to exist, however, the singularity of A means that it cannot be obtained by a straightforward application of Gaussian elimination.

26.4 Methods of solution for the singular equations

A feasible method to solve the simultaneous equations (1) is to find a grid k which has four neighbours (if no such grid exists, then one with only three neighbours will do), to set $x_k=0$, and to eliminate the k-th equation. The remaining (n-1) equations should be non-singular, and solution should be possible. The method chosen initially was *Gauss-Seidel iteration*, a simple technique which works well when the equations

Figure 26.1: A sixteen grey-level display of magnetometer data, showing the effects of lack of balancing between grids. The boundaries of each 20×20 metre grid are clearly visible, largely disguising the archaeological features.

are sparse, as is the case here with at most five non-zero coefficients per equation, and diagonally dominant. Although Gauss-Seidel iteration worked quite well for simple cases, it became laborious when the number of grids was large, and sometimes failed altogether.

The difficulty with large data sets arose from the fact that the diagonal dominance is not sufficiently strong, leading to extremely slow convergence of the iterative method. The cases of failure occurred when setting adjustment $x_k=0$ was not sufficient to remove the singularity. In these cases the *nullity* or *rank deficiency* of the matrix A is greater than 1, indicating that the survey is divided into a number of disjoint portions, which are not connected to each other through any internal edge. In fact the nullity is equal to the number of disjoint portions within the survey; a completely connected survey gives a nullity of 1, one divided into two disjoint portions a nullity of 2, and so on. To obtain a solution, as many adjustments must be set to zero as the nullity.

This raises the problem of how to determine the nullity of A. One possible approach is to solve the simultaneous equations (1) using Gaussian elimination with full pivoting. This technique is not usually recommended for simple problems, but on a singular matrix its effect would be systematically to eliminate elements below the diagonal, until eventually a square array of zeros is obtained in the bottom right-hand corner of the matrix. The size of the zero array indicates the nullity of A. Having set the corresponding unknowns x_i to zero, it is possible to calculate the remaining unknowns by back substitution, and thus reach a solution to the original problem.

There is a numerical difficulty in attempting this calculation, for in numerical analysis exact zeros hardly ever occur, but a value which is smaller than the accuracy of the calculation may be assumed to be equivalent to zero. Consequently careful attention has to be given to setting the criteria which determine whether the zero submatrix has been obtained.

Before attempting to program the calculation, the author sought advice on its feasibility, and on any precautions that should be taken. An inspection of the manuals revealed that the Gaussian elimination subroutines within the NAG library do not provide a solution to a singular problem. Golub and Van Loan (1989) mention Gaussian elimination with full pivoting briefly in Chapter 3, opining that it is only of interest in cases where the nullity has to be determined, as in the present problem. If a serious attempt is to be made at solving singular problems, they recommend that *singular value decomposition* (SVD) should be used, which they discuss at greater length in Chapter 8.

The basis of SVD is that any real matrix A can be decomposed into the form

$$A = UDV^{T} ,$$

where U and V are orthogonal matrices, and D is a diagonal matrix with elements $\sigma_1, \sigma_2, \ldots, \sigma_n$:

$$D = [\mathrm{diag}(\sigma_i)] .$$

In the present problem, it may be assumed that the matrix A is square. If A is singular, then at least one of the elements σ_i will be zero, and the number of such zero elements will be equal to the nullity of A. It is also possible to check whether A is effectively singular in numerical terms, by selecting the largest magnitude among the elements σ_i and dividing the remaining elements by it. Any element for which the result is of the order of the computing precision is effectively zero.

The solution to the matrix equation (2) may be expressed in the form

$$x = A^{-1}b ,$$

where the generalised inverse A^{-1} is defined by

$$A^{-1} = V\,[\mathrm{diag}(\sigma_i^{-1})]\,U^{T} .$$

If case A is effectively singular (or *ill-conditioned*), σ_i^{-1} is set to zero whenever the element σ_i is effectively zero.

The advantages of SVD over the natural extension of Gaussian elimination lie in the unambiguous determination of the nullity, in offering a solution which is optimal according to well defined criteria, and in the ability to assess the arbitrary terms which may be included in the general solution. The disadvantages lie in the complexity of the concepts, and in the length of the calculation involved, although the concepts are quite neatly summarised in Press *et al*. (1989), who also offer a detailed routine for numerical solution. Since it was not clear that any archaeological advantages would accrue from the extra computational effort, it was decided to proceed with the simpler technique based upon Gaussian elimination with full pivoting.

26.5 Practical results

The routine necessary to solve equation (1) was incorporated into the author's existing program for the interpretation of geophysical data. The data for each grid of 20×20 readings are held in a separate file, and the various data files are related to each other through a report file, which specifies their relative positions together with any adjustment necessary to achieve an overall balance. The new routine automatically writes the optimal adjustments into the report file; if required, a general bias can be applied to the adjustments to achieve an average reading specified by the user.

Once the amended report file has been produced, it can be used to set up a new display of the data, on which the contrast between mismatched grids should be reduced to the minimum. Fig. 26.2 shows the effect of the new calculation on the data of Fig. 26.1. Every visible grid boundary has been removed from the display, and the grey scales have been adjusted to emphasise archaeological features uniformly throughout the survey.

Similarly good results have been obtained for virtually every data set to which the technique has been applied. Difficulties arise only where the edge of a very sharply defined feature coincides with a grid boundary. It is then necessary to define the mismatch between the adjoining grids with especial care, to avoid matching the top of the feature to the background level in the adjoining grid. The presence of such sharply varying features suggests that the survey has been technically undersampled, but the combination of matching techniques discussed in section 26.2 — weighted values, removal of outliers, and extrapolation across the edge — copes with most situations.

The routines were given a very practical test when they were used by the geophysical survey team attached to the Newstead Research Project during the 1990 season. The field workers were instructed not to attempt any fine calibration of the instruments during fieldwork, but to rely on the new routine to achieve a balance once the results had been transferred to the computer. This scheme worked extremely well throughout the season, and obviated the need for inexperienced workers to spend much time in adjusting instruments.

The only problem from Newstead came towards the end of the season, when data files for one extended area of survey had accumulated steadily. Automatic balancing was applied on a daily basis to the accumulating data, but took longer and longer to run, until eventually it failed, having exceeded the memory available for the calculation. This problem illustrates two difficulties of the technique when applied to large data sets: first it requires a large amount of memory, since the whole of the $n \times n$ matrix A has to be stored; and secondly it may be very slow, since the calculation time is proportional to n^3. The failure occurred because the author, unaware that very large surveys were to be undertaken, had only made provision for 150 grids. The accumulated survey finally covered some seven hectares, and contained about 180 grids!

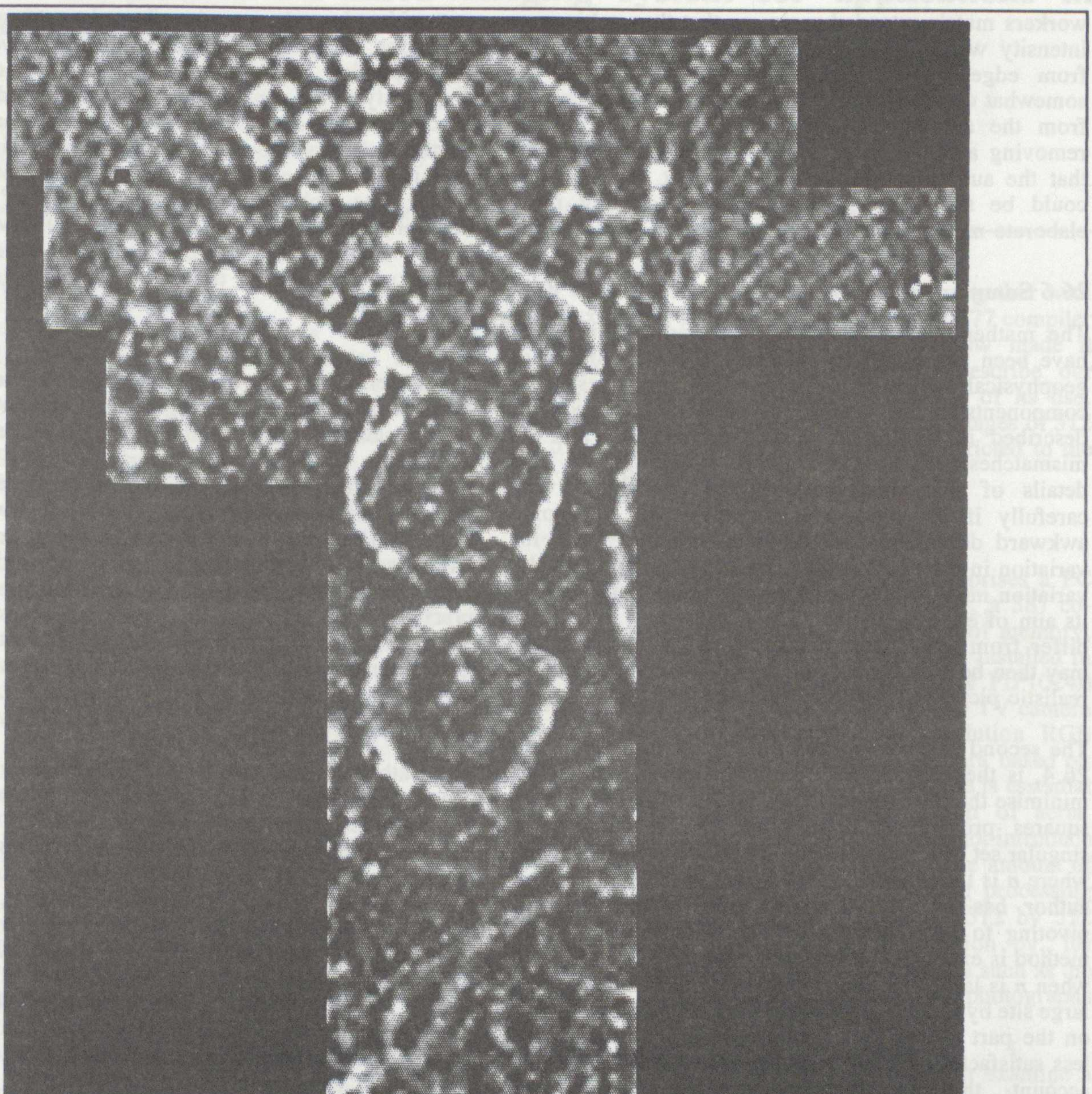

Figure 26.2: A display of the same data as Fig. 26.1, after application of the automatic balancing routine. The grid boundaries have virtually disappeared, and the archaeological features are clearly visible at uniformly high contrast.

It might be possible to circumvent the problems of size and speed by taking advantage of the fact, mentioned at the beginning of section 26.4, that the simultaneous equations (1) are sparse. Sparse equations may be solved by special techniques which reduce both the memory requirement and the calculation time. It is not entirely clear, however, that such techniques can be adapted to the solution of singular sets of equations. The author is currently investigating the matter, but for the time being it appears that practical considerations must limit the technique to surveys of not much more than 150 grids.

Another difficulty may be encountered when a survey divides into several disjoint portions, so that the nullity of matrix A is greater than 1. The grids in any one portion are then correctly balanced, but no information is available to define a balance between the separate portions. Consequently it may not be possible to produce a pleasing display of all the portions simultaneously, and the benefit of the balancing technique will be lost. One solution is to regard the disjoint portions as separate surveys, and not to bring them together for display until the balancing operation has been completed. There may still be a problem when a site is divided by a feature, such as a hedge or a track, over which readings cannot be taken; if the grids are located so that the features cuts through the middle of at least one grid, rather than dividing adjacent grids, the site will remain connected for balancing purposes.

When data have been subjected to automatic balancing, it is sometimes remarked that the display does not have a uniform appearance, but that some areas appear lighter than average, whereas others appear darker. This effect is illustrated in Fig. 26.2, where the lower portion tends to be darker than the upper. This variation may reflect the actual physical condition of the site, since there is evidence that the magnetometer readings in the lower portion tend to be negative on average, while those in the upper tend to be positive. When using their own judgment to balance grids, many workers match extended areas, so that they achieve an intensity which is generally more uniform than that from edge matching. Consequently, they may be somewhat disconcerted by the less uniform appearance from the automatic method but, given the aim of removing artificial edge effects, there is no evidence that the automatic results are incorrect, or that they could be substantially improved by adopting more elaborate matching procedures.

26.6 Summary and conclusions

The mathematical techniques described in this paper have been successful in balancing a large number of geophysical surveys. There are two essential components to the calculations. The first component, described in section 26.2, is the estimation of the mismatches between the edges of adjacent grids. The details of this stage have to be considered very carefully if the method is to work well for some awkward data sets. Clearly, there is some room for variation in the exact application of the details, but the variation must be limited if the technique is to achieve its aim of edge balancing. The results may sometimes differ from what users expect from balancing, but it may then be that the automatic method gives the more realistic picture of the underlying geophysics.

The second component, described in sections 26.3 and 26.4, is the estimation of the adjustments required to minimise the edge mismatches, on the basis of the least squares procedure. This entails the solution of a singular set of n simultaneous equations in n unknowns, where n is the number of data grids in the survey. The author has adapted Gaussian elimination with full pivoting to take account of the singularity, but the method is expensive in terms of both memory and time when n is large. On the other hand, the balancing of a large site by judgment would require considerable effort on the part of the user, and would probably produce less satisfactory results; when user effort is taken into account, the new method may be regarded as acceptably economical.

Various alternative methods are available for solving simultaneous equations. Methods for sparse systems could reduce the expense of memory and time, but it is not evident how they can be adapted to sets of singular equations. Singular value decomposition would give more precise information about the nature of the singularity of the simultaneous equations, and might be used to achieve a better balance between disjoint portions of the survey. Unfortunately, SVD is likely to be far more expensive than the method adopted by the author. The problems of both expense and of balancing disjoint portions can be alleviated by dividing the survey into smaller sections during the preliminary stages of analysis of the survey, and bringing the sections together only for the final stage of presentation.

The experience of the Newstead project shows that the benefits of the new balancing method are not confined to the computing laboratory, but also extend to fieldwork. Because satisfactory displays can be produced from data which have been produced without regard to balance, there is no need for fieldworkers to spend much time in instrumental calibration, beyond ensuring that all equipment is in proper working order. The problem discussed in this paper arose from the practical difficulties of organising extensive field surveys, and it is gratifying that proper mathematical analysis has produced a worthwhile solution.

Acknowledgements

The author is grateful to Paul Cheetham, who first brought the problem back to mind by pointing out its implications for archaeological geophysics in the field, and who subsequently tested the author's solution during the 1990 season of the Newstead Research Project.

References

GOLUB, G.H. & C.F. VAN LOAN 1989. *Matrix Computations*, 2nd Edition, Baltimore, John Hopkins University Press.

PRESS, W.H., B.P. FLANNERY, S.A. TEUKOLSKY & W.T. VETTERLING 1989. *Numerical Recipes, the Art of Scientific Computing (FORTRAN version)*, Cambridge, Cambridge University Press. (Pascal and C versions are also published.)