X. SOME COMPUTATIONAL DEVICES

BY HAROLD HOTELLING

In solving normal equations it should be understood that it is highly desirable first to obtain the inverse matrix. This fact is often overlooked, but the necessity of computing standard errors and the frequent desirability of other uses of the inverse matrix – for example, adding to or taking from the set of predictors – are cogent arguments for adopting a standing rule that no large system of linear equations should ever be solved excepting by first finding the inverse of the matrix of their coefficients.

Direct calculation by methods such as those of Doolittle and Dwyer requires labor of the order of p^3, where p is the number of rows or of unknowns. Thus inverting a matrix of 50 rows requires only about one-eighth as much work as inverting a matrix of 100 rows. This consideration contributes interest to methods of inverting a matrix by partitioning it and inverting submatrices. A method of doing this has been set forth[1], involving four inversions of matrices of half the order of the given one. An improvement given by Waugh involves only two inversions of submatrices. In each case there are also multiplications and additions of submatrices, but these are more straightforward operations than inversion, even though the labor of multiplying two matrices of order p is also of the order of p^3. In partitioning a matrix for this purpose there is an advantage in dividing the rows into *equally numerous* groups, since when the sum of two positive numbers is fixed, the sum of their cubes is a minimum when they are equal.

In special cases there are further advantages in partitioning. Thus there may be whole blocks of zeros, or there may be triangles of zeros that make it easy to invert particular submatrices.

Another important method in matrix inversion is iteration. Several iterative methods are discussed in the paper cited [Hotelling, 1943 – 1]. The method recommended when a fairly good approximation C_0 to A^{-1} has been reached is to use

$$C_{m+1} = C_m(2 - A\,C_m)$$

---

[1] [Hotelling, 1943 - A]. Further results on matrix calculation are given in [Hotelling, 1943 - B and 1949] and [Ullman, 1944].

repeatedly to compute successively C_1, C_2, ..., while considering at each state the matrix of errors

$$D_m = 1 - A\,C_m .$$

The *norm* of a real matrix is the square root of the sum of the squares of the elements. It is an extremely useful means of setting an upper bound for the errors. Let the norm of D_0, denoted by $N(D_0)$, be k. Then if $k < 1$, we have the following upper bound for the errors in the approximations:

$$N(C_m - A^{-1}) \leq N(C_0)\,\frac{k^{2^m}}{1-k} .$$

This formula shows two advantages of this particular method of iteration. One is that the limit of error is expressed only in terms of known numbers, without involving things calculated with a degree of accuracy not previously determined. The other advantage is that this limit of error decreases with great speed as m increases, because of the exponential of the exponential appearing in it. If k happens to be greater than unity, the method will sometimes converge nevertheless, and a new $k < 1$ will emerge at a later stage. If the method diverges, a better first approximation must be found in some other way.

Both iteration and partitioning are useful in least-squares problems with constraints. These problems are of two main kinds. The first is typified in surveying, where the Euclidean value for the sum of the angles of a triangle provides a set of side conditions that modify the solution.

The second kind is represented by the regression treatment of the analysis of variance with disproportionate class frequencies. Here the side conditions are introduced solely for convenience, to make definite the selection of a particular solution among an infinity of possible solutions. When this fact is realized, the arbitrary character of the side conditions becomes evident, and it will be recognized that essentially the same final results will be reached even if the usual simple conditions are considerably altered.

It is customary to take side conditions asserting that the simple sums of certain of the regression coefficients are zero. The labor of calculation is cut down if these are altered by introducing the marginal total frequencies as coefficients.

The whole system of normal equations and side conditions in such cases is treated in a new paper by the author with the help of partitioned matrices, which clarify many of the relationships. The matrix of the whole system has an inverse, of which a particular submatrix plays a part in this theory closely analogous to that of the inverse of the matrix of the normal equations in the ordinary nonsingular case. This analogy includes both the computational and the probability aspects. Furthermore, the matrix of the whole system can be inverted with the help of the iterative method set forth above. This will be advantageous whenever a good first approximation is available. A suitable source to which we may look for such a first approximation is the familiar treatment of proportionate frequencies.