Numerical Linear Algebra in Statistical Computing

Higham, Nicholas J. and Stewart, G. W.

1987

MIMS EPrint: **2008.113**

Manchester Institute for Mathematical Sciences
School of Mathematics

The University of Manchester

ISSN 1749-9097

2 Numerical Linear Algebra in Statistical Computing

N. J. HIGHAM and G. W. STEWART

ABSTRACT

Some of the factors to be considered when applying the techniques of numerical linear algebra to statistical problems are discussed with reference to three particular examples: the use of the normal equations in regression problems; the use of perturbation theory to assess the effects of errors in regression matrices; and the phenomenon of benign degeneracy, in which the numerical problem becomes more difficult even as the associated statistical problem becomes easier.

1. INTRODUCTION

Although statistics contains a wealth of problems for the practitioner of numerical linear algebra, their solution is not as straightforward as it might at first seem. Some of the verities of our field appear in a curious light when we attempt to adapt them to the realities of the statistical world. In this paper we shall give three examples, each pertaining to the classical linear regression model

$$\underline{y} = X\underline{b} + \underline{e}, \tag{1.1}$$

where

$$\underline{y} \in \mathbb{R}^n, \quad X \in \mathbb{R}^{n\times p}, \quad \underline{b} \in \mathbb{R}^p, \quad n \geqslant p,$$

and the random vector $\underline{e} \in \mathbb{R}^n$ is normally distributed according to

$$\underline{e} \sim N(\underline{0}, \sigma^2 I).$$

X is referred to as the regression matrix and $\underline{b}$ the vector of regression coefficients.

In Section 2 we appraise the role of the normal equations in regression problems and offer some explanations as to why the normal equations method has been used very satisfactorily by statisticians for a long time, despite its shortcomings when compared to the orthogonalisation methods preferred by numerical analysts.

In Section 3 we consider the use of perturbation theory to assess the effects of errors in the regression matrix on the regression coefficients. Standard perturbation results tend to be too crude in the context of statistical problems and their sensitivity to the scaling of the problem is unsatisfactory. We indicate how finer bounds can be obtained and show that certain "collinearity coefficients" can provide useful information about the sensitivity of a regression problem.

It is not uncommon in statistics to find the phenomenon of benign degeneracy, in which the numerical problem becomes more difficult even as the associated statistical problem becomes easier. This phenomenon is examined in Section 4, using the Fisher discriminant for illustration.

Throughout this paper we will assume that the regression matrix X in (1.1) has full rank. Pertinent discussions concerning rank deficient regression problems may be found in Stewart (1984) and Hammarling (1985).

We shall use $\|\cdot\|$ to denote the vector 2-norm,

$$\|\underline{x}\| = (\underline{x}^T\underline{x})^{\frac{1}{2}},$$

or the induced matrix norm,

$$\|X\| = \max_{\|\underline{x}\|=1} \|X\underline{x}\| = \rho(X^TX)^{\frac{1}{2}},$$

where ρ denotes the spectral radius (the largest eigenvalue in modulus). Some feel for the size of $\|X\|$ may be obtained from the relation

$$\|X\| \leq \left(\sum_i \sum_j x_{ij}^2\right)^{\frac{1}{2}} \leq \sqrt{p}\|X\| .$$

We shall also make use of the matrix condition number, defined by

$$\kappa(X) = \|X\| \ \|X^+\|$$

where X^+ is the pseudo-inverse of X.

2. THE NORMAL EQUATIONS

For the regression problem (1.1) the least squares estimate of the regression coefficients is the unique vector $\hat{\underline{b}}$ satisfying

$$\|\underline{y}-X\hat{\underline{b}}\| = \min_{\underline{b}} \|\underline{y}-X\underline{b}\| . \tag{2.1}$$

There are two methods commonly used for solving the least squares problem (2.1). The first is based on the readily derived normal equations, which were known to Gauss,

$$A\hat{\underline{b}} = \underline{c} , \tag{2.2a}$$

where

$$A = X^TX , \ \underline{c} = X^T\underline{y} . \tag{2.2b}$$

Since X has full rank, the cross-product matrix $A = X^TX$ is symmetric positive definite. Hence the normal equations may be solved by computing a Choleski decomposition

$$A = T^TT ,$$

where T is upper triangular with positive diagonal elements (Golub and Van Loan, 1983, p.88), and performing a forward substitution followed by a backward substitution.

The second popular approach is to make use of a QR factorisation of the matrix X,

$$X = Q\begin{bmatrix} R \\ 0 \end{bmatrix},$$

where $Q \in \mathbb{R}^{n\times n}$ is orthogonal and $R \in \mathbb{R}^{p\times p}$ is upper triangular. Since rank (R) = rank (X), and

$$\|\underline{y}-X\underline{b}\| = \left\|Q^T\left(\underline{y}-Q\begin{bmatrix}R\\0\end{bmatrix}\underline{b}\right)\right\| = \left\|\begin{bmatrix}Q_1^T\\Q_2^T\end{bmatrix}\underline{y}-\begin{bmatrix}R\\0\end{bmatrix}\underline{b}\right\|,$$

where $Q=[Q_1,Q_2]$, the least squares estimate $\hat{\underline{b}}$ is obtained by solving the nonsingular triangular system

$$R\hat{\underline{b}} = Q_1^T\underline{y}. \qquad (2.3)$$

The QR factorisation may be computed in several ways which we now summarise; for further details see Golub and Van Loan (1983, Chapters 3 and 6). The preferred approach for general problems is orthogonal reduction of X to triangular form using Householder transformations, as first described in detail by Golub (1965). The reduction takes the form

$$H_s H_{s-1}\cdots H_1 X = \begin{bmatrix}R\\0\end{bmatrix}, \qquad s=\min\{n-1,p\},$$

where the Householder matrix H_k satisfies

$$H_k = I-2\underline{u}_k\underline{u}_k^T, \qquad \|\underline{u}_k\| = 1,$$

and where the first $k-1$ components of $\underline{u}_k$ are zero. The last $n-k+1$ components of $\underline{u}_k$ are chosen so that pre-multiplication by H_k creates zeros below the diagonal in the kth column of the partially triangularised matrix $H_{k-1}\cdots H_1X$.

An alternative elimination technique is one based on Givens rotations. A Givens rotation is an orthogonal matrix which differs from the identity matrix only in one submatrix of order 2, which takes the form

$$\begin{array}{c} \\ i \\ k \end{array}\begin{array}{c} \begin{array}{cc} i & k \end{array} \\ \begin{bmatrix} c & s \\ -s & c \end{bmatrix} \end{array}, \qquad c^2+s^2 = 1.$$

The reduction of X to upper triangular form may be accomplished by pre-multiplying X by a sequence of Givens rotations, each of which is chosen so as to introduce one new zero into the lower triangular part of the matrix product. Although up to twice as expensive as the Householder approach for full matrices, the

Givens QR reduction has two redeeming features. First, zeros are introduced in a selective fashion, so that a suitably tailored implementation can be more efficient on sparse or structured problems. Second, the rows of X can be processed one at a time, which is desirable if X is too large to fit into main storage, or if X is generated row-wise, as is often the case in statistical computations.

Other techniques of interest for historical reasons, though of less practical importance nowadays, are the Gram-Schmidt and modified Gram-Schmidt algorithms.

A relationship between the normal equations and the QR factorisation methods can be seen by noting that

$$X^TX = R^TQ^TQR = R^TR = R^T\,\mathrm{diag}\,(\mathrm{sign}(r_{ii}))^2R,$$

so that R is the Choleski factor of X^TX up to scaling of each row by ±1. It is interesting to note that equation (2.3) can be derived by substituting $X=Q_1R$ into the normal equations (2.2) and pre-multiplying by R^{-T}.

The normal equations method is almost universally used by statisticians while the QR factorisation method is almost universally recommended by numerical analysts. On the surface the numerical analysts would seem to have the better of it. In the first place the QR equations have a favourable backward error analysis which the normal equations do not. For the QR factorisation method using Householder transformations it can be shown (Stewart, 1973, p.240) that the solution $\bar{\underline{b}}$ computed in floating point arithmetic with rounding unit ε_M is the true least squares estimate for the perturbed regression equation

$$\underline{y}+\underline{f} = (X+E)\underline{b}+\underline{e}, \qquad (2.4)$$

where the perturbations $\underline{f}$ and E are bounded by

$$\|E\| \leq \phi_1(n,p)\varepsilon_M\|X\|,$$

$$\|\underline{f}\| \leq \phi_2(n,p)\varepsilon_M\|\underline{y}\|,$$

where ϕ_1 and ϕ_2 are low degree polynomials in n and p. Thus the QR factorisation method solves a "nearby" regression problem, and if the computed solution is unsatisfactory, the blame can be placed on the provider of the problem rather than the numerical method.

From the backward error analysis for the Choleski decomposition (Golub and Van Loan, 1983, p.89) it follows that the computed solution $\tilde{\underline{b}}$ for the normal equations method satisfies

$$(A+G)\tilde{\underline{b}} = \underline{c}\,, \tag{2.5}$$

where

$$\|G\| \leqslant c_p\, \varepsilon_M\, \|A\|\,,$$

c_p being a small constant depending on p. Here, for simplicity, we have assumed that the normal equations are formed exactly.

It is possible to translate this backward error result into one of the form of (2.4), but inevitably the perturbation G is magnified in the process. To see this, assume $A+G$ is symmetric positive definite and consider the equation

$$(X+H)^T(X+H) = A+G\,,$$

where the smallest of the many solutions H is to be determined (the one that minimises $\sum_i \sum_j h_{ij}^2$, say). By passing to the singular value decomposition we may assume that X has the form

$$X = \begin{bmatrix} \Psi \\ 0 \end{bmatrix},$$

where

$$\Psi = \operatorname{diag}(\psi_1, \cdots, \psi_p)\,, \qquad \psi_1 \geqslant \cdots \geqslant \psi_p > 0\,.$$

Partitioning H conformally as

$$H = \begin{bmatrix} H_1 \\ H_2 \end{bmatrix},$$

the equation to solve becomes

$$\Psi^2 + \Psi^T H_1 + H_1^T \Psi + H^T H = \Psi^2 + G\,.$$

Ignoring second order terms and writing the equation in scalar form, we get

$$\psi_i h_{ij} + h_{ji}\psi_j = g_{ij}\,, \tag{2.6}$$

where $H_1 = (h_{ij})$, $G = (g_{ij})$. Since we require the smallest solution we minimise $h_{ij}^2 + h_{ji}^2$ subject to (2.6). Remembering that G is symmetric, the solution is easily seen to be

$$h_{ij} = \frac{\psi_i g_{ij}}{\psi_i^2 + \psi_j^2}\,,$$

which for $i=j=p$ reduces to

$$h_{pp} = \frac{g_{pp}}{2\psi_p}\,.$$

Thus, to first order, part of the error is magnified by a factor proportional to ψ_p^{-1}. Since $\|G\|$ is proportional to $\|X\|^2$, this means that $\|H\|/\|X\|$ will be proportional to $\kappa(X) = \psi_1/\psi_p$. Clearly, then, the QR method is superior from the point of view of backwards stability.

Turning to the forward error, $\hat{\underline{b}} - \tilde{\underline{b}}$, a bound for the normal equations may be obtained directly from (2.5) on using standard perturbation theory for square linear systems (Golub and Van Loan, 1983, p.27). We have

$$\frac{\|\hat{\underline{b}} - \tilde{\underline{b}}\|}{\|\hat{\underline{b}}\|} \leqslant \kappa(A)\, \frac{\|G\|}{\|A\|} + O\left(\frac{\|G\|}{\|A\|}\right)^2, \tag{2.7}$$

where

$$\kappa(A) = \kappa(X^T X) = \kappa(X)^2\,. \tag{2.8}$$

A forward error bound for the QR method can be obtained by making use of standard least squares perturbation theory. From Golub and Wilkinson (1966), provided $X+E$ has full rank,

$$\frac{\|\hat{\underline{b}} - \underline{\bar{b}}\|}{\|\hat{\underline{b}}\|} \leqslant \varepsilon\kappa(X)\left(1 + \frac{\|\underline{y}\|}{\|X\|\,\|\hat{\underline{b}}\|}\right) + \varepsilon\kappa(X)^2\, \frac{\|\hat{\underline{r}}\|}{\|X\|\,\|\hat{\underline{b}}\|} + O(\varepsilon^2), \tag{2.9}$$

where

$$\varepsilon = \max\left\{\frac{\|E\|}{\|X\|}, \frac{\|\underline{f}\|}{\|\underline{y}\|}\right\},$$

$$\hat{\underline{r}} = \underline{y} - X\hat{\underline{b}}\,.$$

Comparing (2.7) and (2.9) we see that while both bounds contain the term $\kappa(X)^2$, in (2.9) a small residual vector mitigates the effect of this term. Hence the bounds suggest that the QR method will produce more accurate solutions than the normal equations method for ill-conditioned problems that have a small residual.

A consequence of the condition squaring effect (2.8) is the fact that while the condition $\kappa(X)\,\varepsilon_M < 1$ is sufficient to ensure that the QR procedure does not break down with a singular computed R-factor, one must impose the much stronger condition $\kappa(X)^2\,\varepsilon_M < 1$ to guarantee that the normal equations method runs to completion. Indeed, merely forming the normal equations may cause valuable information to be lost when X is ill-conditioned, unless the evaluation is done and the results are stored in extended precision arithmetic.

In view of the above comparisons of error and stability properties it is natural to ask why statisticians continue to use the normal equations and why they are generally satisfied with the results. We identify several reasons.

In practical statistical problems the residual vector is usually not very small, so the comparison of the forward error bounds is not strongly in favour of the QR method. Moreover, in many problems the elements of the regression matrix are contaminated by errors of measurement, which are large compared with the rounding errors contemplated by the numerical analyst. If the normal equations are formed and solved in a reasonable precision, the effects of rounding errors will be insignificant compared with the effects of measurement errors. In other words, the problem becomes statistically intractable before it becomes numerically intractable.

To make this assertion precise, let ε_D denote the normwise relative error in the regression matrix. Then, as in the perturbation result (2.9), the data errors alone can induce a

perturbation in the regression vector $\hat{\underline{b}}$ of order $\kappa(X)^s\,\varepsilon_D$, where, roughly, $s = 1$ or 2 according as the residual vector is very small or not. Solution via the normal equations introduces a relative error of order $\kappa(X)^2\,\varepsilon_M$, by (2.7). Thus as long as

$$\kappa(X)^{2-s}\varepsilon_M \leqslant \varepsilon_D\ ,$$

the rounding errors in the normal equations method will play an insignificant role compared with the errors in the regression matrix. For example, if $\varepsilon_M = 10^{-7}$ (as, for example, in IEEE standard arithmetic), $\varepsilon_D = 10^{-3}$ and $\kappa(X) = 10^2$, then $\hat{\underline{b}}$ will have at best one correct figure, because of errors in the data, yet the normal equations method will provide three or more correct figures to the machine problem. However, modifying the example slightly so that $\varepsilon_D = 10^{-5}$ and $\kappa(X) = 10^4$ shows that the normal equations method can fail to solve a meaningful problem.

The conclusion is that if one works to high precision (relative to the accuracy of the data) and takes certain elementary precautions (such as computing estimates of the condition number $\kappa(X)$ (Cline, Moler, Stewart and Wilkinson, 1979)), then one can safely use the normal equations. On the other hand, if one is constructing transportable software which must run on machines with a 32-bit floating point word, then one should use the QR factorisation.

An additional feature which works in favour of the normal equations for statistical problems is that in regression models with a constant term (so that $x_{i1} = 1$ for all i), statisticians often "centre their data" by subtracting the means from the columns of X (Graybill, 1976, p.252; Seber, 1977, p.330). This transformation leads to better conditioned normal equations of order one less (Golub and Styan, 1974). To see this, write

$$X = [\underline{e}\ ,\ X_2]$$

where $\underline{e} = [1, 1, \cdots, 1]^T$. Subtracting the means from the columns of X_2 is equivalent to forming

$$\tilde{X} = X\begin{bmatrix} 1 & -\bar{\underline{x}}^T \\ \underline{0} & I_{p-1} \end{bmatrix} = [\underline{e}, X_2 - \underline{e}\bar{\underline{x}}^T]$$

where $\bar{\underline{x}}^T = n^{-1}\underline{e}^T X_2$. The new cross-product matrix is

$$\tilde{X}^T\tilde{X} = \begin{bmatrix} n & \underline{0}^T \\ \underline{0} & X_2^T X_2 - n\bar{\underline{x}}\bar{\underline{x}}^T \end{bmatrix},$$

which gives reduced normal equations of order $p-1$ with the coefficient matrix

$$\bar{A} = X_2^T X_2 - n\bar{\underline{x}}\bar{\underline{x}}^T .$$

One can show that if A has Choleski factor

$$R_p = \begin{bmatrix} \sqrt{n} & \underline{r}^T \\ \underline{0} & R_{p-1} \end{bmatrix},$$

then $\bar{A}$ has Choleski factor R_{p-1}. It follows that

$$\kappa(\bar{A}) = \kappa(R_{p-1})^2 \leq \kappa(R_p)^2 = \kappa(A) .$$

This potential improvement of the condition can be expected to lead to more accurate computed solutions provided that $\bar{A}$ and the corresponding right-hand side vector are computed using extra precision, or perhaps even in standard precision using formulae of the type (Seber, 1977, pp.331, 333; Chan, Golub and LeVeque, 1983)

$$\bar{a}_{ij} = \sum_{k=1}^{n} (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j) ,$$

where $X_2 = (x_{ij})$, $\bar{\underline{x}} = (\bar{x}_i)$.

There is one case in which the normal equations are undoubtedly to be preferred, even when one must resort to high precision. These problems arise in unbalanced analysis of variance and the analysis of categorical data, where the regression matrix is large and sparse but the normal equations are relatively small and dense. Here, formation of the normal

equations will be far more efficient than computation of the QR factorisation, because of intermediate fill-in in the course of computing the R-factor. For example, if $p = 5$ and if the rows of X are being processed one at a time, then one may find matrices of the form

$$R_5 = \begin{bmatrix} X & X & X & X & X \\ 0 & X & X & X & X \\ 0 & 0 & X & X & X \\ 0 & 0 & 0 & X & X \\ 0 & 0 & 0 & 0 & X \end{bmatrix} \qquad \underline{x}_6\underline{x}_6^T = \begin{bmatrix} X & 0 & 0 & X & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ X & 0 & 0 & X & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\underline{x}_6^T = (X \;\; 0 \;\; 0 \;\; X \;\; 0).$$

In the QR factorisation by Givens rotations the rotation in the (1,6) plane which zeros the (1,6) element has the undesirable effect of "filling in" the rest of the row, which makes subsequent treatment of this row as expensive as if it were a full vector. In contrast, the contribution of the new row $\underline{x}_6^T$ to

$$A = \sum_{i=1}^{n} \underline{x}_i\underline{x}_i^T$$

is inexpensive to compute, as it has only four nonzero elements.

3. PERTURBATION THEORY

We have observed that regression matrices often have errors in their elements. It is natural to attempt to use perturbation theory to assess the effects of these errors on the regression coefficients. The results of such an attempt are generally disappointing, for two reasons. First, the use of triangular and submultiplicative inequalities in the course of deriving the bounds reduces their sharpness. This is not a major concern to numerical analysts, whose errors are typically very small and can suffer some magnification without ill effect. The statistician, on the other hand, with his larger errors must fear that

the weakness of the bounds will cause him to declare a good problem intractable.

The second reason why perturbation theory gives disappointing results is that often it is impossible to arrive at a suitable scaling of the problem. For the norm of a regression matrix to represent the sizes of its columns, the columns must be scaled so that they are roughly equal in norm. The same is true of the error in the regression matrix. On the other hand, for a bound on the norm of the vector of regression coefficients to say something meaningful about all the coefficients, the problem must be scaled so that the coefficients are roughly equal. This three way balancing act will in general be unsolvable because any permissible scaling of the regression matrix will scale the error matrix identically and the regression coefficients inversely, as can be seen in the relation (for nonsingular $S \in \mathbb{R}^{p\times p}$)

$$\|(X+E)\,\underline{b}-\underline{y}\| = \|(XS+ES)S^{-1}\,\underline{b}-\underline{y}\| . \qquad (3.1)$$

One cure is to produce finer bounds in terms of individual coefficients and columns. A way of doing this is as follows. Suppose X is perturbed in its ith column by a vector $\underline{f}$, so that the error matrix E is given by

$$E = \underline{f}\underline{e}_i^T ,$$

where $\underline{e}_i$ is the ith column of the $p\times p$ identity matrix, and let $\bar{\underline{b}}$ be the corresponding vector of perturbed regression coefficients. Assume that $X+E$ has full rank. Using the expansion

$$(X+E)^+ = X^+ - X^+EX^+ + (X^TX)^{-1}E^T(I-XX^+) + O(\|E\|^2)$$

we have

$$\begin{aligned}\bar{\underline{b}} &= (X+E)^+\underline{y} = \hat{\underline{b}} - X^+E\hat{\underline{b}} + (X^TX)^{-1}E^T(I-XX^+)(\underline{y}-X\hat{\underline{b}}) + O(\|E\|^2)\\ &= \hat{\underline{b}} - X^+\underline{f}\underline{e}_i^T\hat{\underline{b}} + (X^TX)^{-1}\underline{e}_i\underline{f}^T(I-XX^+)\hat{\underline{r}} + O(\|\underline{f}\|^2)\\ &= \hat{\underline{b}} - X^+\underline{f}_1\hat{b}_i + (X^TX)^{-1}\underline{e}_i\underline{f}_2^T\hat{\underline{r}} + O(\|\underline{f}\|^2) ,\end{aligned}$$

where $\underline{f}_1$ and $\underline{f}_2$ are the projections of $\underline{f}$ onto the range of X and its orthogonal complement respectively. On pre-multiplying by $\underline{e}_j^T$, and using the bound

$$|\underline{e}_j^T(X^TX)^{-1}\underline{e}_i| = |\underline{e}_j^TX^+X^{+T}\underline{e}_i| \leq \|X^{+T}\underline{e}_j\|\,\|X^{+T}\underline{e}_i\| ,$$

we obtain

$$|\bar{b}_j-\hat{b}_j| \leq \|\underline{x}_j^{(+)}\|\,|\hat{b}_i|\,\|\underline{f}_1\| + \|\underline{x}_j^{(+)}\|\,\|\underline{x}_i^{(+)}\|\,\|\hat{\underline{r}}\|\,\|\underline{f}_2\| + O(\|\underline{f}\|^2)$$
$$j = 1,\cdots,p, \qquad (3.2)$$

where $\underline{x}_k^{(+)}$ denotes the transpose of the kth *row* of X^+. The interpretation of this bound is clearly much less dependent on the scaling of the problem than is the case for the bound (2.9). In fact, a diagonal scaling $S=\mathrm{diag}(s_i)$ of the form in (3.1) leaves (3.2) unchanged.

The above analysis leads naturally to the introduction of the *collinearity coefficients*

$$\kappa_i = \|\underline{x}_i\|\,\|\underline{x}_i^{(+)}\| , \qquad i = 1,\cdots,p .$$

In addition to playing a key role in the perturbation bound (3.2) the collinearity coefficients have at least two other important properties. First, the reciprocal of κ_i is the smallest relative perturbation in the ith column of X that makes X exactly collinear (that is, rank deficient). This can be shown by using the QR factorisation

$$X = Q\begin{bmatrix}R\\0\end{bmatrix} = Q\begin{bmatrix}R_{11} & \underline{r}\\ \underline{0}^T & r_{pp}\\ 0 & \underline{0}\end{bmatrix}, \qquad R_{11}\in\mathbb{R}^{(p-1)\times(p-1)},$$

where we can assume without loss of generality that the column of interest is the last. Clearly, a perturbation

$$\underline{h} = -Q\begin{bmatrix}\underline{0}\\ r_{pp}\\ \underline{0}\end{bmatrix},$$

to the last column of X makes X collinear, and $\|\underline{h}\| = |r_{pp}|$.

Also it is easily seen that if

$$X + \underline{h}\underline{e}_p^T = Q \begin{bmatrix} R_{11} & \underline{r}+\underline{f} \\ \underline{0}^T & r_{pp}+\rho \\ 0 & \underline{g} \end{bmatrix}, \qquad Q^T\underline{h} = \begin{bmatrix} \underline{f} \\ \rho \\ \underline{g} \end{bmatrix},$$

is collinear, then $\rho = -r_{pp}$ so that $\|\underline{h}\| \geqslant |r_{pp}|$. Thus $|r_{pp}|/\|\underline{x}_p\|$ is the size of the smallest relative perturbation to the last column of X that makes X collinear. But

$$X^+ = [R^+ \; 0]\, Q^+ = \begin{bmatrix} R_{11}^{-1} & -(r_{pp}R_{11})^{-1}\underline{r} & 0 \\ \underline{0}^T & r_{pp}^{-1} & \underline{0}^T \end{bmatrix} Q^T$$

so that $\underline{e}_p^T X^+ = r_{pp}^{-1}\underline{e}_p^T Q^T$, and hence

$$|r_{pp}| = \|\underline{x}_p^{(+)}\|^{-1}.$$

Random errors in the regression matrix X tend to cause a systematic reduction in the size of the regression coefficients when X is ill-conditioned, since such errors tend to increase the size of small or zero singular values. Another use of the collinearity coefficients is to measure the extent of this bias in the regression coefficients; the analysis is fairly lengthy and appears in Stewart (1986).

A desirable property of the numbers κ_i is that they are invariant to diagonal scalings of the columns of X, unlike the standard condition number $\kappa(X)$. Some other interesting properties of the collinearity coefficients are discussed in Stewart (1986).

We mention in passing that Fletcher (1985) attempts to overcome the two drawbacks discussed at the start of this section by using a probabilistic perturbation analysis. This approach may be of particular interest in the context of statistical computations.

4. BENIGN DEGENERACY

The phenomenon of benign degeneracy is best illustrated by an example. The Fisher discriminant function (Graybill, 1976, Section 12.5) is a method for deciding whether a sample vector $\underline{x}$, known to be drawn from one of two populations distributed according to $N(\underline{\mu}_1, \Sigma)$ and $N(\underline{\mu}_2, \Sigma)$ respectively, belongs to the first or to the second. The Fisher discriminant classifies $\underline{x}$ as belonging to the $N(\underline{\mu}_1, \Sigma)$ population if

$$\underline{t}^T\underline{x} > \tfrac{1}{2}\underline{t}^T(\underline{\mu}_1+\underline{\mu}_2),$$

where

$$\underline{t} = \Sigma^{-1}(\underline{\mu}_1-\underline{\mu}_2),$$

and to the $N(\underline{\mu}_2, \Sigma)$ population otherwise. As the dispersion matrix Σ approaches singularity the problem of computing the Fisher discriminant becomes increasingly ill-conditioned.

However, the numerical problem does not correspond to an intrinsic statistical problem. To see this, let Σ have the spectral decomposition

$$\Sigma = Q^T D Q, \quad Q^T Q = I, \quad D = \operatorname{diag}(d_i),$$

and transform to the new coordinate system

$$\underline{x}' = Q\underline{x} - \tfrac{1}{2}Q(\underline{\mu}_1 + \underline{\mu}_2),$$

in which the two populations are distributed according to $N(\underline{\mu}_1', D)$ and $N(\underline{\mu}_2', D)$ respectively, where $\underline{\mu}_1' = -\underline{\mu}_2' = \tfrac{1}{2}Q(\underline{\mu}_1-\underline{\mu}_2)$. The Fisher discriminant declares $\underline{x}'$ to belong to the first population if

$$(\underline{\mu}_1'-\underline{\mu}_2')^T D^{-1}\underline{x}' > 0.$$

In the new coordinate system the singularity of the dispersion matrix corresponds to one or more of the components having zero variance, and in the Fisher discriminant the inverse weights these components infinitely. In other words, if a component has zero variance, then it is sufficient to look at that component alone to determine to which population a sample vector

belongs — as is clear intuitively. (If there are several components with zero variance then any single one of these may be considered.)

The question for the numerical analyst is how to evaluate the Fisher discriminant. One possibility is to apply the above-mentioned transformation to diagonalise the problem, after which it is obvious what to do. An alternative is to use the original formula, no matter how ill-conditioned the dispersion matrix. In the unlikely event that one is required to divide by zero, the zero is replaced by a suitable small number. Whether this seemingly risky procedure works is an open question!

REFERENCES

Chan, T.F., Golub, G.H. and LeVeque, R.J. (1983), "Algorithms for computing the sample variance: analysis and recommendations", *Amer. Statist.*, *37*, pp.242-247.

Cline, A.K., Moler, C.B., Stewart, G.W. and Wilkinson, J.H. (1979), "An estimate for the condition number of a matrix", *SIAM J. Numer. Anal.*, *16*, pp.368-375.

Fletcher, R. (1985), "Expected conditioning", *IMA J. Numer. Anal.*, *5*, pp.247-273.

Golub, G.H. (1965), "Numerical methods for solving linear least squares problems", *Numer. Math.*, *7*, pp.206-216.

Golub, G.H. and Styan, G.P.H. (1974), "Some aspects of numerical computations for linear models", *Interface — Proceedings of Computer Science and Statistics*, 7th Annual Symposium on the Interface 1973, pp.189-192.

Golub, G.H. and Van Loan, C.F. (1983), *Matrix Computations*, Johns Hopkins University Press (Baltimore, Maryland).

Golub, G.H. and Wilkinson, J.H. (1966), "Note on the iterative refinement of least squares solution", *Numer. Math.*, *9*, pp.139-148.

Graybill, F.A. (1976), *Theory and Application of the Linear Model*, Duxbury Press (North Scituate, Mass.).

Hammarling, S.J. (1985), "The singular value decomposition in multivariate statistics", *ACM SIGNUM Newsletter*, *20(3)*, pp.2-25.

Seber, G.A.F. (1977), *Linear Regression Analysis*, John Wiley (New York).

Stewart, G.W. (1973), *Introduction to Matrix Computations*, Academic Press (New York).

Stewart, G.W., (1984), "Rank degeneracy", *SIAM J. Sci. Statist. Comput.*, *5*, pp.403-413.

Stewart, G.W. (1986), "Scale invariant measures of the effects of near collinearity", manuscript (submitted for publication).

N.J. Higham
Department of Mathematics
University of Manchester
Manchester M13 9PL
England

G.W. Stewart
Department of Computer Science
University of Maryland
College Park
Maryland 20742
U.S.A.