A theorem on rank one modifications to a matrix and its inverse

By M. J. D. Powell*

Some algorithms need to compute the inverse of the matrix $(J + uv^T)$, where J is a non-singular $n \times n$ matrix, u and v are n-component column vectors, and where the inverse of J is available. This calculation can be carried out in many ways using only of order n^2 computer operations, but if the inverse of J is wrong, then the resultant expression for the inverse of $(J + uv^T)$ is usually wrong also. We identify an expression for $(J + uv^T)^{-1}$ that is useful because it tends to suppress the contribution from any errors in J^{-1}.

(Received March 1969)

1. Introduction

The calculation of $(J + uv^T)^{-1}$, where J is an $n \times n$ matrix and u and v are n-component column vectors, is important to many different fields of numerical computation, for instance the solution of non-linear algebraic equations (Broyden, 1965) and the inversion of a matrix by the method of modification (Householder, 1964). When the matrix J^{-1} is available, the formula

$$(J + uv^T)^{-1} = J^{-1} - \frac{J^{-1}uv^TJ^{-1}}{(1 + v^TJ^{-1}u)} \tag{1}$$

is often used, because its application requires only of order n^2 computer operations.

However in Broyden's algorithm the vector

$$\gamma = u(v^Tv) + Jv \tag{2}$$

is available instead of the vector u, in which case it is convenient to rewrite expression (1) so that γ replaces u. Thus we obtain the equation

$$(J + uv^T)^{-1} = J^{-1} + \frac{(v - J^{-1}\gamma)v^TJ^{-1}}{(v^TJ^{-1}\gamma)}. \tag{3}$$

The purpose of this paper is to point out that for numerical computation it is often preferable to use expression (3) instead of expression (1).

The advantage of the second formula is obtained in the usual case when, instead of the exact matrix J^{-1}, we have an approximation H, which differs from J^{-1} due to the errors of previous computation. In this case the two formulae

$$(J + uv^T)^{-1} \approx H - \frac{Huv^TH}{(1 + v^THu)} \tag{4}$$

and

$$(J + uv^T)^{-1} \approx H + \frac{(v - H\gamma)v^TH}{(v^TH\gamma)} \tag{5}$$

(γ is defined by equation 2) are not equivalent. Indeed we let the right-hand sides of equations (4) and (5) be H_* and H_+ respectively, and we deduce the identities

$$(J + uv^T) - H_*^{-1} = J - H^{-1} \tag{6}$$

and

$$(J + uv^T) - H_+^{-1} = (J - H^{-1})\left(I - \frac{vv^T}{(v^Tv)}\right). \tag{7}$$

Equation (7) shows that the advantage of using formula (5) is that there is some suppression of the errors in H.

In Section 2 we establish equation (7), and in Section 3 we give two numerical examples to show the advantages of preferring formula (5).

2. The theorem

Let J be any $n \times n$ matrix, and let H be any non-singular $n \times n$ matrix. Also let u and v be any n-component column vectors subject to the condition that the scalar $(v^TH\gamma)$ is non-zero, where

$$\gamma = u(v^Tv) + Jv. \tag{8}$$

Then the matrices J_+ and H_+, defined by the equations

$$J_+ = J + uv^T \tag{9}$$

and

$$H_+ = H + \frac{(v - H\gamma)v^TH}{(v^TH\gamma)}, \tag{10}$$

satisfy the equation

$$(J_+ - H_+^{-1}) = (J - H^{-1})\left(I - \frac{vv^T}{(v^Tv)}\right), \tag{11}$$

where I is the unit matrix.

Proof. By combining the identity

$$v^T\left(I - \frac{vv^T}{(v^Tv)}\right) = 0 \tag{12}$$

with the definitions (9) and (10), we obtain the equations

$$J_+\left(I - \frac{vv^T}{(v^Tv)}\right) = J\left(I - \frac{vv^T}{(v^Tv)}\right) \tag{13}$$

and

$$H_+H^{-1}\left(I - \frac{vv^T}{(v^Tv)}\right) = \left(I - \frac{vv^T}{(v^Tv)}\right). \tag{14}$$

Also, noting that $J_+v = \gamma$ and $H_+\gamma = v$, we deduce the identity

$$(H_+J_+ - I)v = 0. \tag{15}$$

Now equations (13), (14) and (15) give the relation

* *Theoretical Physics Division, Atomic Energy Research Establishment, Harwell*

$$(H_+J_+ - I) = (H_+J_+ - I)\Big(I - \frac{vv^T}{(v^Tv)}\Big)$$

$$= (H_+J - H_+H^{-1})\Big(I - \frac{vv^T}{(v^Tv)}\Big), \quad (16)$$

from which we can obtain the required equation (11) by multiplication by H_+^{-1}, provided that the matrix H_+ has an inverse.

To complete the proof we show that the conditions of the theorem imply that H_+ is non-singular, for then H_+^{-1} is well-defined.

We suppose that there is a non-zero vector x such that $H_+x = 0$, and show that consequently there is a contradiction. The definition (10) implies that the vector H_+x is a linear combination of Hx and $(v - H\gamma)$, so if it is zero there is, because of the non-singularity of H, a number λ such that

$$x = \lambda H^{-1}(v - H\gamma). \quad (17)$$

In this case direct calculation using equation (10) gives the result

$$H_+x = \lambda(v - H\gamma)(v^Tv)/(v^TH\gamma). \quad (18)$$

It follows that H_+x is zero only if λ or $(v - H\gamma)$ is zero, in which case expression (17) shows that x is zero also, which is the contradiction. Therefore H_+ is non-singular, and so the theorem is an immediate consequence of equation (16).

3. Discussion

Expression (11) shows that the difference $(J_+ - H_+^{-1})$ is equal to the difference $(J - H^{-1})$ multiplied by a symmetric projection matrix. Therefore we have the inequality

$$||J_+ - H_+^{-1}||_F \leqslant ||J - H^{-1}||_F, \quad (19)$$

where the subscript 'F' denotes the Frobenius norm. Moreover the inequality (19) is strict unless $(J - H^{-1})v = 0$, which is the feature of formula (5) that tends to suppress the effect of errors between H^{-1} and J. Indeed the fact that equation (15) is satisfied shows that we force a relation between H_+ and J_+ that would be obtained if H were equal to J^{-1}.

On the other hand H_*, the right-hand side of the alternative formula (4), is calculated to satisfy the equation

$$H_*^{-1} = H^{-1} + uv^T, \quad (20)$$

so we deduce the result

$$J_+ - H_*^{-1} = J - H^{-1}, \quad (21)$$

which is equation (6). This expression implies that any discrepancies between J and H^{-1} are neither suppressed nor amplified.

The relation (21) shows that the usual formula (4) is excellent for single calculations of $(J + uv^T)^{-1}$. However often one has to calculate a sequence of matrices $J^{(1)}, J^{(2)}, \ldots$ and their inverses $H^{(1)}, H^{(2)}, \ldots$, where for $k = 1, 2, \ldots$ the difference $(J^{(k+1)} - J^{(k)})$ is a matrix of rank one. In this case it is usual to apply formula (4) recursively. Here a numerical calculation will cause rounding errors at every stage of the process, so equation (21) and a consideration of probabilities indicate that gradually the discrepancies between $J^{(k)}$ and $H^{(k)-1}$ will tend to grow as k becomes large. However the theorem of Section 2 shows that this will not happen if formula (5) is preferred, provided that the successive vectors v do not have unfortunate tendencies towards linear dependence.

We illustrate typical behaviour by a numerical example. For this example we calculated three sequences of 10×10 matrices, $J^{(k)}$, $H_+^{(k)}$ and $H_*^{(k)}$, using single length arithmetic on an IBM 360/65 computer, which provides about six and a half decimals accuracy. For $k = 1, 2, \ldots$ we chose vectors $u^{(k)}$ and $v^{(k)}$ in a way to be described, then $J^{(k+1)}$ was defined by the equation

$$J^{(k+1)} = J^{(k)} + u^{(k)}v^{(k)T}, \quad (22)$$

$H_+^{(k+1)}$ was calculated by substituting $H_+^{(k)}$ and $J^{(k)}$ in equations (2) and (5), and $H_*^{(k+1)}$ was obtained by substituting $H = H_*^{(k)}$ in equation (4). To start the process we let $J^{(1)} = H_+^{(1)} = H_*^{(1)} = I$, the unit matrix. The successive vectors $u^{(k)}$ and $v^{(k)}$ were obtained by using random numbers in the following way: each component of $v^{(k)}$ was set equal to a random number from the distribution that is uniform over $[-1, 1]$, and also each component of a vector $g^{(k)}$ was set to a random number from the same distribution. Then we defined

$$u^{(k)} = (g^{(k)} - J^{(k)}v^{(k)})/(v^{(k)T}v^{(k)}), \quad (23)$$

and we tried calculating $J^{(k+1)}$, $H_+^{(k+1)}$ and $H_*^{(k+1)}$ in the way just specified. However if it happened that any element of any of these three matrices exceeded ten in modulus, then different random numbers were used to define $v^{(k)}$ and $g^{(k)}$, and if necessary this replacement continued, until all the elements of all the calculated matrices $J^{(k+1)}$, $H_+^{(k+1)}$ and $H_*^{(k+1)}$ were in the range $[-10, 10]$. Finally, before proceeding to the next value of k, we calculated the discrepancies

$$d_+^{(k)} = \max_{ij} |(J^{(k+1)}H_+^{(k+1)})_{ij} - \delta_{ij}| \quad (24)$$

and

$$d_*^{(k)} = \max_{ij} |(J^{(k+1)}H_*^{(k+1)})_{ij} - \delta_{ij}|, \quad (25)$$

where δ_{ij} is the Kronecker delta.

This calculation was repeated three times, and the resultant discrepancies are given in **Table 1**. We note that the table confirms that formula (5) is superior to formula (4) in controlling the rounding errors of the numerical calculation.

Table 1

Calculated values of $d_+^{(k)}$ and $d_*^{(k)}$ when $H^{(1)} = J^{(1)-1}$

k	$d_+^{(k)} \times 10^6$			$d_*^{(k)} \times 10^6$		
10	7	22	10	14	27	26
20	19	32	29	31	43	58
30	14	10	20	39	35	40
40	20	13	17	136	39	70
50	22	20	11	280	75	30
100	16	20	15	39	53	79
200	31	14	13	256	141	155
500	15	18	25	164	201	272

Note that in the example we constrain the size of the elements of $J^{(k)}$, $H_{+}^{(k)}$ and $H_{*}^{(k)}$. Without the constraints the values of $d_{+}^{(k)}$ and $d_{*}^{(k)}$ would fluctuate so much that a more extensive table would be needed to compare formulae (4) and (5). Then it is more preferable to use equation (5) because larger errors are suppressed. Also note that in the example we calculate γ from $u^{(k)}$ and $v^{(k)}$ (although, except for rounding errors, $\gamma = g^{(k)}$), in order to include all the rounding errors that usually occur in the evaluation of $H_{+}^{(k+1)}$.

Also we calculated a similar example to show the rate at which formula (5) suppresses errors. Here we followed the calculation just described, except that we introduced random, non-zero, off-diagonal elements into the initial matrices $H_{+}^{(1)}$ and $H_{*}^{(1)}$, but $J^{(1)}$ remained equal to I. Each off-diagonal element of $H_{+}^{(1)}$ was set to a number from the distribution that is uniform over $[-0\cdot 1,\ 0\cdot 1]$, and we set $H_{*}^{(1)} = H_{+}^{(1)}$. The resultant discrepancies are reported in **Table 2**. We see that formula (5) requires about 175 iterations to make the errors as small as those of **Table 1**, but in fact the number of iterations depends strongly on the directions of the successive vectors $v^{(k)}$. In particular the fewest number of iterations would occur if n consecutive values of $v^{(k)}$ were orthogonal, for then all the initial errors would be annihilated by n iterations, because equation (11) implies the identity

$$(J^{(k+1)} - \{H_{+}^{(k+1)}\}^{-1}) = (J^{(1)} - \{H_{+}^{(1)}\}^{-1}) \prod_{i=1}^{k} \left(I - \frac{v^{(k+1-i)}v^{(k+1-i)T}}{v^{(k+1-i)T}v^{(k+1-i)}}\right). \quad (26)$$

The discrepancies $d_{*}^{(k)}$ of **Table 1** are not very large, so often the loss of accuracy due to formula (4) is acceptable. Indeed formula (4) is often the better choice, because it does not require the elements of J. However formula (5) is natural to Broyden's (1965) algorithm, because v and γ are available but J is not. In Powell's (1968) method for non-linear equations both J and H are calculated, and here formula (5) is preferred.

We recommend that when H, J, v and γ are available, equation (5) should always be used, but when u is calculated instead of γ, other factors are important. Specifically a need for high accuracy supports equation (5), but the calculation of γ requires an extra vector by matrix multiplication, which increases the computing time by about one third.

For iterative processes that include updating the inverse of a matrix that is subject to a rank one modification, many readers may prefer the formula recommended in this paper, because it does suppress the accumulation of errors.

Table 2

Calculated values of $d_{+}^{(k)}$ and $d_{*}^{(k)}$ when $H^{(1)} \neq J^{(1)-1}$

k	$d_{+}^{(k)} \times 10^6$			$d_{*}^{(k)} \times 10^6$		
10	880,366	143,245	313,590	1,152,046	648,336	395,583
20	227,253	387,309	181,209	335,528	390,858	459,403
30	115,996	131,936	66,923	431,328	892,932	747,881
40	104,810	18,775	53,300	785,799	367,247	491,335
50	19,580	12,541	43,337	671,965	710,577	590,575
75	6,039	5,592	17,128	592,856	647,500	1,125,246
100	6,915	260	3,764	813,716	956,899	974,374
125	521	72	363	473,855	386,242	650,397
150	203	74	118	723,234	980,423	1,058,437
175	21	10	16	535,759	383,976	451,759
200	21	24	15	1,066,539	774,638	1,499,561

References

Broyden, C. G. (1965). A class of methods for solving non-linear simultaneous equations, *Math. Comp.*, Vol. 19, pp. 577–593.

Householder, A. S. (1964). *The theory of matrices in numerical analysis*, pp. 123–124, Blaisdell Publishing Co.: New York.

Powell, M. J. D. (1968). A FORTRAN subroutine for solving systems of non-linear algebraic equations, *Report No. R*–5947, A.E.R.E., Harwell.