SIAM J. NUMER. ANAL.
Vol. 3, No. 4, 1966
Printed in U.S.A.

ON ERROR BOUNDS FOR GENERALIZED INVERSES*

ADI BEN-ISRAEL†

Introduction. Solving the system of linear equations

$$Ax = b \tag{1}$$

on a computer, results in a solution $x + dx$ of a "perturbed" system

$$(A + dA)(x + dx) = b + db. \tag{2}$$

The errors involved in the numerical solution of (1) were studied by von Neumann and Goldstine [9], Turing [13], Householder [5] and [6], Lonseth [8], Wilkinson [14], [15] and [16], Bodewig [3] and others. The error analysis of (1) relates some measure of the error to the data A, b and to the method of solution, thus to the perturbations dA, db generated thereby.

A possible measure of the error is $\| dx \|$. Some popular bounds on $\| dx \|$ are:

$$\| dx \| \leqq \frac{\| A \| \, \| A^{-1} \| \, \| db \| \, \| x \|}{\| b \|} \quad \text{if} \quad dA = 0, \tag{3}$$

$$\| dx \| \leqq \frac{\| A^{-1} \| \, \| dA \| \, \| x \|}{1 - \| A^{-1} \| \, \| dA \|} \quad \text{if} \quad db = 0, \qquad \| A^{-1} \| \, \| dA \| < 1, \tag{4}$$

$$\| dx \| \leqq \frac{\| A^{-1} \|}{1 - \| A^{-1} \| \, \| dA \|} (\| A^{-1} \| \, \| dA \| \, \| b \| + \| db \|) \quad \text{if} \quad \| A^{-1} \| \, \| dA \| < 1. \tag{5}$$

These bounds, and some related results, are generalized below to the case where A is rectangular or singular and where (1) is solved by using the generalized inverse [10]. Corollaries 1, 2, 3 and 4 are well known, and are included here as immediate consequences of more general, and less useful, results.

Notations. Let E^n be the n-dimensional complex vector space with

$$(x, y) = \sum x_i \bar{y}_i, \qquad \| x \| = (x, x)^{1/2} \quad \text{for} \quad x, y \in E^n.$$

For an $m \times n$ complex matrix A let

A^* = the conjugate transpose of A,

$A^\dagger$ = the *generalized inverse* of A, [10],

* Received by the editors March 31, 1966, and in revised form April 25, 1966.

† Systems Engineering Department, University of Illinois, Chicago, Illinois. Presently at Department of Engineering Sciences, Northwestern University, Evanston, Illinois 60201. This research was supported by the National Science Foundation under Grant GP 5230.

$\| A \| = \max_i \lambda_i^{1/2}(A^*A)$, the *spectral norm* of A,[1]
$R(A) = \{y \in E^m : y = Ax, x \in E^n\}$, the *range space* of A,
$N(A) = \{x \in E^n : Ax = 0\}$, the *null space* of A.
A matrix P is a *projector* if $P = P^2 = P^*$, [7]. If P is a projector and $L = R(P)$ we write $P = \Pr(L)$.

The matrices A, B are arbitrary, and their dimension is apparent from the context in which they appear; thus $A + B$ means that A and B are of the same dimension, and in particular $I + A$ means that A is square.

The following results are needed in the sequel:

(6) $\| Ax \| \leqq \| A \| \| x \|$ for all $x \in E^n$ and A;

(7) $\| AB \| \leqq \| A \| \| B \|$ for all A, B, e.g., [7];

(8) $\| P \| = 1$ for every nonzero projector P;

(9) $\| x + y \| \geqq | \| x \| - \| y \| |$ for all $x, y \in E^n$;

(10) $\| A + B \| \leqq \| A \| + \| B \|$;

(11) $\| A + B \| \geqq | \| A \| - \| B \| |$;

(12) $AA^\dagger = \Pr(R(A))$, e.g., [2];

(13) $A^\dagger A = \Pr(R(A^*))$;

(14) $R(A^*) = N(A)^\perp$.

THEOREM 1. *If P is a projector and*

$$\| A \| < 1, \tag{15}$$

then

$$\operatorname{rank}(P + A) \geqq \operatorname{rank} P. \tag{16}$$

Proof. Let $p = \operatorname{rank} P$ and let $x_1, x_2, \cdots, x_p$ be a basis for $R(P)$. We prove (16) by proving the linear independence of the vectors $(P + A)x_i$, $i = 1, \cdots, p$. If there are scalars $\alpha_1, \cdots, \alpha_p$, not all of which are zero, such that

$$\sum \alpha_i(P + A)x_i = 0, \tag{17}$$

then

$$\begin{aligned} 0 &= \| \sum \alpha_i(P + A)x_i \| = \| (P + A) \sum \alpha_i x_i \| \\ &\geqq \| P \sum \alpha_i x_i \| - \| A \| \| \sum \alpha_i x_i \| \qquad \text{by (9) and (6)} \end{aligned}$$

[1] For a discussion of matrix norms see [7], especially p. 44 where the spectral norm is denoted by lub_S (A).

$$\begin{aligned}(18)\qquad &\geqq \| \textstyle\sum \alpha_i x_i \| - \| A \| \| \textstyle\sum \alpha_i x_i \| && \text{since} \quad Px_i = x_i \\ &\geqq (1 - \| A \|) \| \textstyle\sum \alpha_i x_i \| \\ &> 0 && \text{by (15) and} \quad \textstyle\sum \alpha_i x_i \neq 0.\end{aligned}$$

COROLLARY 1. *If* $\| A \| < 1$, *then* $(I + A)$ *is nonsingular.*

THEOREM 2. *If* $\| A \| < 1$ *and* L *is a subspace such that*

$$L \supset R(A), \tag{19}$$

then

$$\| (\mathrm{Pr}\,(L) + A)^{\dagger} \| \leqq \frac{1}{1 - \| A \|}. \tag{20}$$

Proof. Using (13) we get

$$\begin{aligned}\mathrm{Pr}\,(R(\mathrm{Pr}\,(L) + A^*)) &= \mathrm{Pr}\,(R(\mathrm{Pr}\,(L) + A)^*) \\ (21)\qquad &= (\mathrm{Pr}\,(L) + A)^{\dagger}(\mathrm{Pr}\,(L) + A) \\ &= (\mathrm{Pr}\,(L) + A)^{\dagger}\mathrm{Pr}\,(L) + (\mathrm{Pr}\,(L) + A)^{\dagger}A.\end{aligned}$$

Now from (19) and Theorem 1 it follows that

$$R(\mathrm{Pr}\,(L) + A) = L; \tag{22}$$

therefore $(AA^{\dagger} + A)^{\dagger}$ is the zero operator on $L^{\perp}$, [2], and we conclude that

$$(\mathrm{Pr}\,(L) + A)^{\dagger}\mathrm{Pr}\,(L) = (\mathrm{Pr}\,(L) + A)^{\dagger}. \tag{23}$$

From (21) and (23) we get

$$\mathrm{Pr}\,(R(\mathrm{Pr}\,(L) + A^*)) = (\mathrm{Pr}\,(L) + A)^{\dagger}(I + A), \tag{24}$$

and by taking norms and using (8) and (11),

$$1 = \| \mathrm{Pr}\,(R(\mathrm{Pr}\,(L) + A^*)) \| \geqq \| (\mathrm{Pr}\,(L) + A)^{\dagger} \| (1 - \| A \|), \tag{25}$$

which proves (20).

COROLLARY 2. *If* $\| A \| < 1$, *then*

$$\| (I + A)^{-1} \| \leqq \frac{1}{1 - \| A \|}. \tag{26}$$

THEOREM 3. *If* $\| A \| < 1$ *and* L *is a subspace such that* $L \supset R(A)$, *then*

$$\| (\mathrm{Pr}\,(L) + A)^{\dagger} - \mathrm{Pr}\,(L + R(A^*)) \| \leqq \frac{\| A \|}{1 - \| A \|}. \tag{27}$$

Proof. From (24) we get

$$\mathrm{Pr}\,(R(\mathrm{Pr}\,(L) + A^*)) - (\mathrm{Pr}\,(L) + A)^{\dagger} = (\mathrm{Pr}\,(L) + A)^{\dagger}A, \tag{28}$$

but

$$R(\Pr(L) + A^*) = L + R(A^*) \tag{29}$$

and (27) follows by using (28), (29), (7) and (20).

COROLLARY 3. *If* $\| A \| < 1$, *then*

$$\| (I + A)^{-1} - I \| \leqq \frac{\| A \|}{1 - \| A \|}. \tag{30}$$

LEMMA 1. *If* A, B *satisfy*

$$AA^\dagger B = B, \tag{31}$$

$$A^\dagger AB^* = B^*, \tag{32}$$

$$\| A^\dagger B \| < 1, \tag{33}$$

then

$$(A + B)^\dagger = (I + A^\dagger B)^{-1}A^\dagger. \tag{34}$$

Proof. Using (31) we write

$$A + B = A(I + A^\dagger B). \tag{35}$$

By Corollary 1 and (33) the matrix $(I + A^\dagger B)$ is nonsingular. To prove (34) amounts therefore to proving that

$$(AC)^\dagger = C^\dagger A^\dagger, \tag{36}$$

where

$$C = I + A^\dagger B. \tag{37}$$

Greville [4] showed that (36) holds if, and only if, the following two conditions are satisfied:

$$A^\dagger ACC^*A^* = CC^*A^* \tag{38}$$

and

$$CC^\dagger A^*AC = A^*AC. \tag{39}$$

Now, (39) follows from the nonsingularity of C, and (38) follows from

$$\begin{aligned} CC^*A^* &= (I + A^\dagger B)(I + A^\dagger B)^*A^* \\ &= (I + A^\dagger B)(I + B^*A^{\dagger *})A^* \\ &= (I + A^\dagger B)(A^* + B^*(AA^\dagger)^*) \\ &= (I + A^\dagger B)(A^* + (AA^\dagger B)^*) \qquad (40) \\ &= (I + A^\dagger B)(A^* + B^*) \qquad \text{by (31)} \end{aligned}$$

$$= (I + A^{\dagger}B)(A^* + A^{\dagger}AB^*) \qquad \text{by (32)}$$
$$= A^* + A^{\dagger}(BA^* + AB^* + BA^{\dagger}AB^*),$$

and from $A^{\dagger}AA^* = A^*$ by (13), and $A^{\dagger}AA^{\dagger} = A^{\dagger}$.

THEOREM 4. *If A, B satisfy* (31), (32) *and* (33), *then*

$$(A + B)^{\dagger} - A^{\dagger} = \sum_{k=1}^{\infty} (-1)^k (A^{\dagger}B)^k A^{\dagger} \tag{41}$$

and

$$\| (A + B)^{\dagger} - A^{\dagger} \| \leqq \frac{\| A^{\dagger}B \| \, \| A^{\dagger} \|}{1 - \| A^{\dagger}B \|}. \tag{42}$$

If (31) *and* (32) *hold, but* (33) *is replaced by*

$$\| A^{\dagger} \| \, \| B \| < 1, \tag{43}$$

then

$$\| (A + B)^{\dagger} - A^{\dagger} \| \leqq \frac{\| A' \|^2 \| B \|}{1 - \| A^{\dagger} \| \, \| B \|}. \tag{44}$$

Proof. From Lemma 1 it follows that

$$(A + B)^{\dagger} - A^{\dagger} = [(I + A^{\dagger}B)^{-1} - I]A^{\dagger}. \tag{45}$$

From (33) follows the expansion

$$(I + A^{\dagger}B)^{-1} = \sum_{k=0}^{\infty} (-1)^k (A^{\dagger}B)^k, \tag{46}$$

and the last two equations yield (41). The inequality (42) follows from (45) by using Corollary 3. The condition (43) (which is stronger than (33)) then implies (44) (which is weaker than (42) but does not involve $\| A^{\dagger}B \|$).

COROLLARY 4. *If A is nonsingular and*

$$\| A^{-1}B \| < 1, \tag{47}$$

then $(A + B)$ is nonsingular and

$$(A + B)^{-1} - A^{-1} = \sum_{k=1}^{\infty} (-1)^k (A^{-1}B)^k A^{-1}, \tag{48}$$

and

$$\| (A + B)^{-1} - A^{-1} \| \leqq \frac{\| A^{-1}B \| \, \| A^{-1} \|}{1 - \| A^{-1} B \|}. \tag{49}$$

If

(50) $$\| A^{-1} \| \| B \| < 1,$$

then

(51) $$\| (A + B)^{-1} - A^{-1} \| \leqq \frac{\| A^{-1} \|^2 \| B \|}{1 - \| A^{-1} \| \| B \|}.$$

Recall now that the "best approximate solution" (least squares solution of least norm) of (1) is [11]

(52) $$x = A^{\dagger}b.$$

We study now the sensitivity of x in (52) to variations in the data A, b.

THEOREM 5. *If* (1) *is solvable, x satisfies* (52) *and*

(53) $$(x + dx) = A^{\dagger}(b + db)$$

then

(54) $$\frac{\| dx \|}{\| x \|} \leqq \frac{\| A \| \| A^{\dagger} \| \| db \|}{\| b \|}.$$

Proof. From (53) and (52) we have

(55) $$dx = A^{\dagger}db,$$

so that

(56) $$\| dx \| \leqq \| A^{\dagger} \| \| db \|.$$

Since (1) is solvable and x satisfies (52), we conclude that x is a solution of (1) and hence

(57) $$\| b \| \leqq \| A \| \| x \|,$$

and (54) follows from (56) and (57).

THEOREM 6. *If* (1) *is solvable, and if x satisfies* (52) *and dx is defined by*

(58) $$x + dx = (A + dA)^{\dagger}b,$$

where the matrix dA satisfies

(59) $$AA^{\dagger}\, dA = dA,$$

(60) $$A^{\dagger}A(dA)^* = (dA)^*$$

and

(61) $$\| A^{\dagger} \| \| dA \| < 1,$$

then

(62) $$\frac{\| dx \|}{\| x \|} \leqq \frac{\| A^{\dagger} \| \| dA \|}{1 - \| A^{\dagger} \| \| dA \|}.$$

Proof. Using (52) and (58) we get

$$dx = [(A + dA)^\dagger - A^\dagger]b. \tag{63}$$

From (41) with $B = dA$, and x being a solution of (1), we get

$$dx = \sum_{k=1}^{\infty} (-1)^k (A^\dagger\, dA)^k A^\dagger A x, \tag{64}$$

and (62) follows by taking norms and using (6), (7), (10), (13) and (8).

THEOREM 7. *If x satisfies* (52), *and if dx is defined by*

$$(x + dx) = (A + dA)^\dagger (b + db), \tag{65}$$

where dA satisfies (59), (60) *and* (61), *then*

$$\| dx \| \leqq \frac{\| A^\dagger \|}{1 - \| A^\dagger \| \| dA \|} (\| dA \| \| A^\dagger \| \| b \| + \| db \|). \tag{66}$$

Proof.

$$dx = ((A + dA)^\dagger - A^\dagger)b + (A + dA)^\dagger\, db. \tag{67}$$

Using Theorem 4 with $B = dA$ we get

$$\| [(A + dA)^\dagger - A^\dagger]\, b \| \leqq \frac{\| A^\dagger \|^2 \| dA \| \| b \|}{1 - \| A^\dagger \| \| dA \|}. \tag{68}$$

Similarly, from Lemma 1 and Corollary 2,

$$\begin{aligned} \| (A + dA)^\dagger\, db \| &\leqq \| (I + A^\dagger\, dA)^{-1} \| \| A^\dagger \| \| db \| \\ &\leqq \frac{\| A^\dagger \| \| db \|}{1 - \| A^\dagger \| \| dA \|}, \end{aligned} \tag{69}$$

and (66) follows from (67), (68) and (69).

Remarks. (a) Theorems 5, 6 and 7 generalize, to the singular case, the bounds (3), (4) and (5) respectively. As in the nonsingular case, e.g., [12, p. 420], in order to use (66) in the error analysis of (52) some bounds or estimates for $\| A^\dagger \|$, $\| dA \|$ and $\| db \|$ are required. These computations, for some of the present day generalized inversion methods, will be produced elsewhere.

(b) The bound (54) (as well as (62) and (66) in the case where the measure of error is not $\| dx \|$ but the residual $\| A(x + dx) - b \| = \| A\, dx \|$) suggests the definition of $\| A \| \| A^\dagger \|$ as the spectral condition number, e.g., [7, p. 81], for singular or rectangular matrices.

(c) The conditions (31) and (32) are equivalent to the condition

$$AA^\dagger B A^\dagger A = B, \tag{70}$$

which is equivalent to the solvability [10] of

$$A X A = B. \tag{71}$$ [10]

Thus the perturbations studied in the Theorems 4, 6 and 7 are (using the notations of [1]) elements of $R(A, A)$.

Using the direct sum decomposition of the space of $m \times n$ complex matrices [1] it is easy to extend our results to perturbations having components outside $R(A, A)$.

REFERENCES

[1] A. Ben-Israel, *On direct sum decompositions of Hestenes algebras*, Israel J. Math., 2 (1964), pp. 50–54.

[2] A. Ben-Israel and A. Charnes, *Contributions to the theory of generalized inverses*, J. Soc. Indust. Appl. Math., 11 (1963), pp. 667–699.

[3] E. Bodewig, *Matrix Calculus*, 2nd ed., North-Holland, Amsterdam, 1959.

[4] T. N. E. Greville, *Note on the generalized inverse of a matrix product*, MRC Report #623, The University of Wisconsin, Madison, 1966.

[5] A. S. Householder, *The approximate solution of matrix problems*, J. Assoc. Comput. Mach., 5 (1958), pp. 204–243.

[6] ———, *Generated error in rotational tridiagonalization*, J. Assoc. Comput. Mach., 5 (1958), pp. 335–338.

[7] ———, *The Theory of Matrices in Numerical Analysis*, Blaisdell, New York, 1964.

[8] A. T. Lonseth, *The propagation of error in linear problems*, Trans. Amer. Math. Soc., 62 (1947), pp. 193–212.

[9] J. von Neumann and H. H. Goldstine, *Numerical inverting of matrices of high order*, Bull. Amer. Math. Soc., 53 (1947), pp. 1021–1099.

[10] R. Penrose, *A generalized inverse for matrices*, Proc. Cambridge Philos. Soc., 51 (1955), pp. 406–413.

[11] ———, *On best approximate solutions of linear matrix equations*, Proc. Cambridge Philos. Soc., 52 (1956), pp. 17–19.

[12] A. Ralston, *A First Course in Numerical Analysis*, McGraw-Hill, New York, 1965.

[13] A. M. Turing, *Rounding-off errors in matrix processes*, Quart. J. Mech. Appl. Math., 1 (1948), pp. 287–308.

[14] J. H. Wilkinson, *Rounding errors in algebraic processes*, Information Processing, UNESCO, Paris, 1960.

[15] ———, *Error analysis of direct methods of matrix inversion*, J. Assoc. Comput. Mach., 8 (1961), pp. 281–330.

[16] ———, *Rounding Errors in Algebraic Processes*, Prentice-Hall, Englewood Cliffs, New Jersey, 1963.