A Singular Value Inequality for Block Matrices

W. Govaerts*
Seminarie voor hogere analyse
Galglaan 2
B-9000 Gent, Belgium

and

J. D. Pryce
Computational Mathematics Group
RMCS Shrivenham
Swindon Wilts SN6 8LA, England

Submitted by Iain S. Duff

ABSTRACT

Let A be an (n, n) submatrix of a nonsingular $(n+m, n+m)$ matrix M, and let S be the inverse of the Schur complement of A in M. Let $p = \min(m, n)$. We obtain upper bounds on the p smallest singular values of A in terms of the corresponding ones in S, and vice versa.

1. INTRODUCTION

Singular values of matrices were not given much attention in the early textbooks on numerical linear algebra [2, 4, 6]. They are in the forefront, however, in more recent works [1], and the inclusion of the singular value decomposition is one of the most advertised features in numerical software packages like MATLAB and GAUSS.

The main result of this paper (the Theorem) suggested itself in a problem arising in numerical continuation theory (for general background see [5]).

---

*Senior Research Associate of the Belgian National Fund of Scientific Research N.F.W.O.

655 Avenue of the Americas, New York, NY 10010 0024-3795/89/$3.50

Typically one traces a curve $G(u, \lambda) = 0$, where λ is a parameter with a physical interpretation and u contains the n components of the discretized solution of a differential equation. In the continuation code one has to solve well-conditioned linear systems $Mz = h$ where M has the block structure

$$M = \underset{\quad n \quad\; m}{\begin{bmatrix} A & B \\ C & D \end{bmatrix}} \begin{matrix} n \\ m \end{matrix}$$

Here A is the Jacobian of G with respect to the variables in u only. It is typically a large matrix with special structure (e.g. banded or symmetric). A "black box" solver for systems like $Ax = y$ is therefore often available. Block elimination for $Mz = h$ is then preferable to full Gaussian elimination with M. Unfortunately, it breaks down if A is nearly singular, i.e. at turning points of the curve $G(u, \lambda) = 0$. These points usually have a physical interpretation like passing from a stable to an unstable solution.

We found that iterative refinement of block elimination is surprisingly successful at turning points, even when block elimination gives an error of more than 100 percent. The error analysis of this phenomenon ([3] and a future paper on the case $m > 1$) requires the Theorem, which seems to be unknown and interesting in its own right.

2. PRELIMINARIES

We formulate the results for the case of complex matrices; for the real case just replace "unitary" by "orthogonal" and adjoint matrices by transposed matrices throughout.

All norms are operator norms with respect to the Hilbert vector norm.

NOTATION. For any matrix A let $\sigma_i(A)$ denote the singular values of A for $i = 1, 2, \ldots$. If A is (m, n), then in the usual convention there are $p = \min(m, n)$ singular values $\sigma_1 \geqslant \cdots \geqslant \sigma_p \geqslant 0$. We extend this by padding out the σ_i with trailing zeros, defining $\sigma_i = 0$ for any $i > p$.This simplifies the statement of many results.

The following is a mixture of standard elementary facts (see [1] or another textbook) and standard deductions therefrom.

LEMMA.

(a) *Write* $\lambda_i(X)$ *for the eigenvalues of a nonnegative definite* (k,k) *matrix* X, *ordered* $\lambda_1(X) \geqslant \cdots \geqslant \lambda_k(X)$. *Then*

$$\sigma_i(A)^2 = \lambda_i(A^*A) \qquad (i=1,\dots,n),$$

$$\sigma_i(A)^2 = \lambda_i(AA^*) \qquad (i=1,\dots,m),$$

where the $\lambda_i(A^*A)$ *and* $\lambda_i(AA^*)$ *are padded out with trailing zeros if necessary.*

(b) $\sigma_i(A) = \sigma_i(A^*)$ *for all* i.

(c) *For any* $i \geqslant 1$, $\sigma_i(A) = \min\{\|A - \tilde{A}\| : \tilde{A}$ *has rank* $< i\}$.

(d) *The singular values of a matrix are unchanged by pre- or postmultiplication with a unitary matrix, by permuting or changing the sign of rows or columns, or by padding the matrix out with zero rows or columns.*

(e) *If* B *is a submatrix of* A, *then* $\sigma_i(B) \leqslant \sigma_i(A)$ *for all* i.

(f) *If* $B = XAY$ *then*

$$\sigma_i(B) \leqslant \|X\| \cdot \|Y\| \sigma_i(A) \qquad \textit{for all } i.$$

(g) *Let* A *be an* (m,n) *matrix and*

$$B = \underset{n}{\begin{bmatrix} I \\ A \end{bmatrix}}\begin{matrix} n \\ m \end{matrix} \qquad (I \textit{ is the identity matrix}).$$

Then $\sigma_i(B)^2 = 1 + \sigma_i(A)^2$ *for* $1 \leqslant i \leqslant n$.

Proof. (a) is standard and implies (b). (c) is standard as well and implies (d), (e), and (f) by straightforward arguments. To prove (g) remark that $B^*B = I + A^*A$. The n eigenvalues of B^*B are thus obtained by adding 1 to these of A^*A. Together with (a), this implies (g). ∎

3. RESULTS

PROPOSITION. *Consider a nonsingular block* $(n+n, n+n)$ *matrix*

$$M = \underset{n\quad n}{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}\begin{matrix} n \\ n \end{matrix}$$

with inverse

$$M^{-1} = \begin{bmatrix} P & Q \\ R & S \end{bmatrix} \begin{matrix} n \\ n \end{matrix}$$
$$\quad\; n \quad n$$

Then

$$\|M\|^{-2}\sigma_i(A) \leqslant \sigma_i(S) \leqslant \|M^{-1}\|^2 \sigma_i(A) \tag{1}$$

for all $i \geqslant 1$.

Proof. By the symmetry in the statement of the Proposition, it is sufficient to prove one of the inequalities in (1), e.g. the second. Without loss of generality we may assume that A is nonsingular. Indeed, all quantities in (1) are continuous functions of the coefficients of M, so that it is enough to prove (1) for an arbitrary small perturbation of M.

Put $A^{-1}B = V$. Then

$$[A^{-1} \quad 0]M = [I \quad V]. \tag{2}$$

From (2) and (e), (f) of the Lemma we get

$$\sigma_i(A^{-1}) \leqslant \|M^{-1}\| \sigma_i[I \quad V] \qquad \text{for all } i. \tag{3}$$

Now for $1 \leqslant i \leqslant n$

$$\begin{aligned}
(\sigma_i[I \quad V])^2 &= \left(\sigma_i\begin{bmatrix} I \\ V^* \end{bmatrix}\right)^2 && [\text{(b) of the Lemma}] \\
&= 1 + \sigma_i(V^*)^2 && [\text{(g) of the Lemma}] \\
&= 1 + \sigma_i(V)^2 && [\text{(b) of the Lemma}] \\
&= \left(\sigma_i\begin{bmatrix} I \\ V \end{bmatrix}\right)^2 && [\text{(g) of the Lemma}] \\
&= \left(\sigma_i\begin{bmatrix} -V \\ I \end{bmatrix}\right)^2 && [\text{(d) of the Lemma}].
\end{aligned}$$

Hence

$$\sigma_i[I \quad V] = \sigma_i\begin{bmatrix} -V \\ I \end{bmatrix} \qquad \text{for all } i. \tag{4}$$

But

$$M\begin{bmatrix} -V \\ I \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{bmatrix} -A^{-1}B \\ I \end{bmatrix} = \begin{bmatrix} 0 \\ S^{-1} \end{bmatrix}.$$

By (d) and (f) of the Lemma we get

$$\sigma_i\begin{bmatrix} -V \\ I \end{bmatrix} \leqslant \|M^{-1}\|\sigma_i(S^{-1}). \tag{5}$$

Combining (3), (4), and (5), we get

$$\sigma_i(A^{-1}) \leqslant \|M^{-1}\|^2\sigma_i(S^{-1}). \tag{6}$$

The singular values of the inverse of a (k, k) matrix are, for $i \leqslant k$, the reciprocals of the singular values of the matrix in reverse order; using this and (6), the Proposition now follows. ■

THEOREM. *Consider a nonsingular $(n+m, n+m)$ matrix with the block form*

$$M = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\begin{matrix} n \\ m \end{matrix}$$

$$\begin{matrix} n & m \end{matrix}$$

with inverse

$$M^{-1} = \begin{bmatrix} P & Q \\ R & S \end{bmatrix}\begin{matrix} n \\ m \end{matrix}$$

$$\begin{matrix} n & m \end{matrix}$$

Put $p = \min(n, m)$. Then

$$\|M\|^{-2}\sigma_{n-j}(A) \leqslant \sigma_{m-j}(S) \leqslant \|M^{-1}\|^2\sigma_{n-j}(A) \qquad \text{for} \quad 0 \leqslant j < p, \tag{7}$$

$$\|M\|^{-1} \leqslant \sigma_i(S) \leqslant \|M^{-1}\|$$

$$\text{if} \quad n < m \text{ and } 1 \leqslant i \leqslant m-n, \tag{8}$$

$$\|M^{-1}\|^{-1} \leqslant \sigma_i(A) \leqslant \|M\|$$

$$\text{if} \quad n > m \text{ and } 1 \leqslant i \leqslant n-m. \tag{9}$$

Proof. For $m = n$ this is precisely the Proposition. Again, there is a striking symmetry in the statement of the Theorem, and so it is sufficient to prove (7) and (8) in the case $n < m$. We pad out M to a matrix

$$\tilde{M} = \begin{bmatrix} \mu I & 0 & 0 \\ 0 & A & B \\ 0 & C & D \end{bmatrix} \begin{matrix} m-n \\ n \\ m \end{matrix}$$

$$\begin{matrix} m-n & n & m \end{matrix}$$

where $\mu = \|M\|$. Then

$$\tilde{M}^{-1} = \begin{bmatrix} \mu^{-1} I & 0 & 0 \\ 0 & P & Q \\ 0 & R & S \end{bmatrix} \begin{matrix} m-n \\ n \\ m \end{matrix}$$

$$\begin{matrix} m-n & n & m \end{matrix}$$

Applying the Proposition, we get

$$\|\tilde{M}\|^{-2}\sigma_i \begin{bmatrix} \mu I & 0 \\ 0 & A \end{bmatrix} \leqslant \sigma_i(S) \leqslant \|\tilde{M}^{-1}\|^2 \sigma_i \begin{bmatrix} \mu I & 0 \\ 0 & A \end{bmatrix}. \tag{10}$$

Since $\|A\| \leqslant \|M\| = \|\mu I\|$, we have

$$\sigma_i \begin{bmatrix} \mu I & 0 \\ 0 & A \end{bmatrix} = \|M\| \qquad \text{for} \quad 1 \leqslant i \leqslant m-n, \tag{11a}$$

$$= \sigma_{n-m+i}(A) \qquad \text{for} \quad m-n < i \leqslant m. \tag{11b}$$

Next,

$$\|\tilde{M}\| = \max(\mu, \|A\|) = \|M\|, \tag{12}$$

$$\|\tilde{M}^{-1}\| = \max(\mu^{-1}, \|M^{-1}\|) = \|M^{-1}\|. \tag{13}$$

Combining (10), (12), (13), and (11b), we obtain

$$\|M\|^{-2}\sigma_{n-m+i}(A) \leqslant \sigma_i(S) \leqslant \|M^{-1}\|^2 \sigma_{n-m+i}(A)$$

for $m - n < i \leqslant m$. Put $j = m - i$ to get (7).

Combining (10), (12), (13), and (11a), we obtain

$$\|M\|^{-2}\|M\| \leqslant \sigma_i(S) \leqslant \|M^{-1}\|^2\|M\|;$$

replacing the second inequality by the obvious bound $\sigma_i(S) \leqslant \|S\| \leqslant \|M^{-1}\|$, we get (8). ■

The result needed in the bordered matrix case ($m \ll n$) in [3] is

Corollary 1. *Let M be as in the Theorem and $m < n$. Then* $\|S\| \leqslant \|M^{-1}\|^2\sigma_{n-m+1}(A)$.

Proof. Take $j = m - 1$ in (7). ■

Corollary 1 quantifies the known fact that $S = 0$ iff A has rank deficiency m. As another striking application we mention

Corollary 2. *Let M be a unitary matrix written as*

$$M = \underset{\quad n \quad\; m}{\begin{bmatrix} A & B \\ C & D \end{bmatrix}}\begin{matrix} n \\ m \end{matrix} \qquad (n \leqslant m).$$

Then

$$\sigma_i(D) = \begin{cases} 1 & \textit{for} \quad 1 \leqslant i \leqslant m - n, \\ \sigma_{i-m+n}(A) & \textit{for} \quad m - n < i \leqslant m. \end{cases}$$

Proof. Apply the Theorem and note that

$$S = D^*,$$

so $\sigma_i(D) = \sigma_i(S)$ for all i by (b) of the Lemma. ■

Remarks. Corollary 2 may also be obtained by direct means (starting from $AA^* + BB^* = I_n$, $B^*B + D^*D = I_m$). In the real case it may be visualized as an orthogonal decomposition property.

Concerning the sharpness of (7), remark that equality on both sides for any j with $\sigma_{n-j}(A) \neq 0$ or $\sigma_{m-j}(S) \neq 0$ implies that M is a scalar multiple of a unitary matrix. One-sided bounds, however (such as we use in the applications), may be sharp for arbitrarily ill-conditioned M. Indeed, if M is a diagonal matrix $M = \mathrm{Diag}[\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_m]$ with $\alpha_1 \geqslant \alpha_2 \cdots \alpha_n \geqslant \beta_1 \geqslant \cdots \beta_m$, then (7) is equivalent to

$$\left(\alpha_1^{-1}\alpha_{n-j}\right)\alpha_1^{-1} \leqslant \beta_{j+1}^{-1} \leqslant \left(\beta_m^{-1}\alpha_{m-j}\right)\beta_m^{-1} \qquad [0 \leqslant j < \min(n, m)],$$

and the first (second) inequality is an equality if $\alpha_1 = \beta_{j+1}$ ($\alpha_{n-j} = \beta_m$).

If $B = C = 0$ (as in the diagonal case), then there is no relation whatsoever between A and S, and therefore the bounds in (7) cannot contain much information, although they may still be sharp. In our applications M is expected to be well-conditioned, so that the bounds cannot be very weak.

We remark also that for any nonzero γ the matrix

$$\overline{M} = \begin{bmatrix} A & \gamma B \\ \gamma^{-1} C & D \end{bmatrix}$$

has the same A and S as M, and so $\|\overline{M}\|$ and $\|\overline{M}^{-1}\|$ can replace $\|M\|$ and $\|M^{-1}\|$ in the formulae (7), (8), and (9), and the parameter γ used to tighten the bounds.

We thank the referees of this paper for some useful comments.

REFERENCES

1 M. Fiedler, *Special Matrices and Their Applications in Numerical Mathematics*, Martinus Nijhoff, Dordrecht, 1986.
2 L. Fox, *An Introduction to Numerical Linear Algebra*, Clarendon, Oxford, 1964.
3 W. Govaerts and J. D. Pryce, Block Elimination with One Iterative Refinement Solves Bordered Linear Systems Accurately, Report AM-88-01, School of Mathematics, Univ. of Bristol.
4 J. Stoer and R. Bulirsch, *Einführung in die Numerische Mathematik II*, Springer-Verlag, Berlin, 1973.
5 Werner C. Rheinboldt, *Numerical Analysis of Parametrized Nonlinear Equations*, Wiley, New York, 1986.
6 J. Wilkinson, *The Algebraic Eigenvalue Problem*, Clarendon, Oxford, 1965.

Received 5 July 1988; final manuscript accepted 6 December 1988