

AD-A065 625

MARYLAND UNIV COLLEGE PARK COMPUTER SCIENCE CENTER  
A NOTE ON THE PERTURBATIONS OF SINGULAR VALUES. (U)

F/G 12/1

UNCLASSIFIED

DEC 78 G W STEWART  
TR-720

N00014-76-C-0391

NL

| OF |  
AD  
A065625

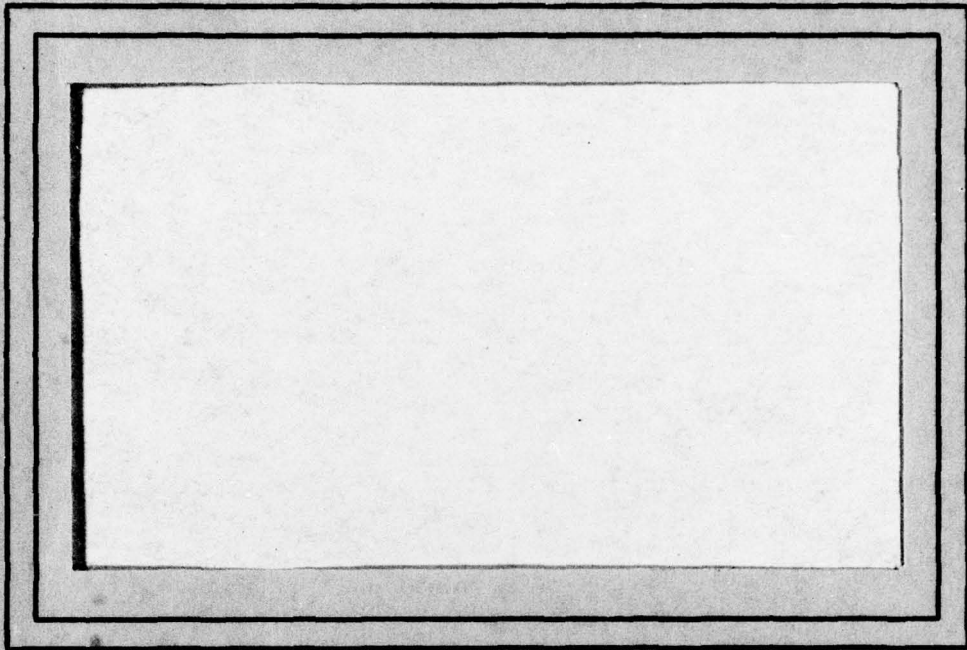


END  
DATE  
FILMED  
5-79  
DDC

LEVEL

①  
NW

AD A0 65625



DDC FILE COPY



DDC  
RECEIVED  
MAR 14 1979  
LIBRARY

*[Handwritten signature]*

UNIVERSITY OF MARYLAND  
COMPUTER SCIENCE CENTER

COLLEGE PARK, MARYLAND  
20742

DISTRIBUTION STATEMENT A  
Approved for public release;  
Distribution Unlimited

79 02 21 058

⑨ Technical Report TR-720  
⑭  
⑪ December 1978  
⑫ 8p.

⑥ A Note  
on the Perturbations of  
Singular Values  
⑩ G. W. Stewart\*

Abstract

↓  
In this note a variant of the classical perturbation theorem for singular values is given. The bound explain why perturbations will tend to increase rather than decrease singular values of the same order of magnitude as the perturbation.  
↑

DDC  
RECEIVED  
MAR 14 1979  
RESOLVED  
A

DISTRIBUTION STATEMENT A  
Approved for public release  
Distribution Unlimited

\*This work was supported in part by the Office of Naval Research under Contract N00014-76-C-0391.  
⑮

page - A - 79 02 21 05 8  
403 018 Tgm

A Note  
on the Perturbation of  
Singular Values

G. W. Stewart

Dedicated to A. S. Householder on his  
seventy-fifth birthday

ACCESSION NO.	
DTIC	White Section <input checked="" type="checkbox"/>
DDC	Self Section <input type="checkbox"/>
DDA/SOURCE	<input type="checkbox"/>
JUSTIFICATION	<i>White on file</i>
BY	
DISTRIBUTION/AVAILABILITY CODES	
Dist.	USUAL, GEN. or SPECIAL
A	

In this note we shall be concerned with a sharpening of the usual perturbation bounds for the singular values of a general rectangular matrix. Specifically let  $X$  be an  $n \times p$  matrix. Then the singular values

$$(1) \quad \sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p$$

of  $X$  may be defined as the nonnegative square roots of the eigenvalues of  $X^T X$ . If  $\tilde{X} = X + E$  is a perturbation of  $X$  and the singular values  $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \dots \geq \tilde{\sigma}_p$  of  $\tilde{X}^T \tilde{X}$  are ordered so that

$$(2) \quad \tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq \dots \geq \tilde{\sigma}_p,$$

then

$$(3) \quad |\sigma_i - \tilde{\sigma}_i| \leq \|E\| \quad (i = 1, 2, \dots, p),$$

where  $\|E\|$  denotes the spectral norm of  $E$  (for definitions and proofs see, e.g., [2]).

Although the perturbation bound (3) is satisfactory in most applications, it does not give a complete description of how the singular values, especially

the small ones, actually behave under perturbations in  $X$ . The following theorem provides a somewhat clearer picture.

Theorem. Let the singular values of  $X$  be ordered as in (1) and those of  $\tilde{X} = X + E$  as in (2). Let  $P$  denote the orthogonal projection onto the column space of  $X$ . Then

$$(4) \quad \tilde{\sigma}_i^2 = (\sigma_i + \xi_i)^2 + \eta_i^2 \quad (i = 1, 2, \dots, p)$$

where

$$(5) \quad |\xi_i| \leq \|PE\|$$

and

$$\inf [(I - P)E] \leq \eta_i \leq \|(I - P)E\|.$$

Here

$$\inf(E) = \inf_{\|x\|=1} \|Ex\|.$$

Proof. We use the classical min-max characterization

$$(6) \quad \sigma_i^2 = \min_{\dim(\mathcal{X})=p-i+1} \max_{\substack{x \in \mathcal{X} \\ \|x\|=1}} x^T (X^T X) x.$$

Let  $\mathcal{X}$  be a subspace for which equality is attained in (6). Then from the min-max characterization of  $\tilde{\sigma}_i^2$ , we have

$$\begin{aligned} \tilde{\sigma}_i^2 &\leq \max_{\substack{x \in \mathcal{X} \\ \|x\|=1}} x^T [(X+E)^T (X+E)] x \\ &= \max_{\substack{x \in \mathcal{X} \\ \|x\|=1}} x^T [X^T X + E^T P X + X^T P E + E^T P^2 E + E^T (I-P)^2 E] x, \end{aligned}$$

where we have made extensive use of the usual properties of projections in passing from the first to the second form of the bound. Now since

$$\sigma_i^2 = \max_{\substack{x \in \mathcal{X} \\ \|x\|=1}} x^T (X^T X) x = \max_{\substack{x \in \mathcal{X} \\ \|x\|=1}} \|Xx\|^2,$$

it follows that

$$\begin{aligned} \tilde{\sigma}_i^2 &\leq \sigma_i^2 + 2\sigma_i \|PE\| + \|PE\|^2 + \|(I-P)E\|^2 \\ (7) \quad &= (\sigma_i + \|PE\|)^2 + \|(I-P)E\|^2 \end{aligned}$$

For a lower bound we use the dual characterization

$$(8) \quad \sigma_i^2 = \max_{\dim(\mathcal{X})=1} \min_{\substack{x \in \mathcal{X} \\ \|x\|=1}} x^T (X^T X) x.$$

Again let  $\mathcal{X}$  be a subspace for which equality is attained in (8).

Proceeding as above, we get

$$\begin{aligned} \tilde{\sigma}_i^2 &\geq \min_{\substack{x \in \mathcal{X} \\ \|x\|=1}} x^T [X^T X + E^T P X + X^T P E + E^T P^2 E] x + \min_{\substack{x \in \mathcal{X} \\ \|x\|=1}} x^T E^T (I-P)^2 E x \\ (9) \quad &\geq \min_{\substack{x \in \mathcal{X} \\ \|x\|=1}} x^T [X^T X + E^T P X + X^T P E + E^T P^2 E] x + \inf [(I-P)E]^2. \end{aligned}$$

Let  $x$  be a vector for which the minimum in the last expression of (9) is attained. Calling this minimum  $\mu$ , we have

$$\begin{aligned}\mu &= \|Xx\|^2 + 2(x^T X^T)(PEx) + \|PEx\|^2 \\ &\geq (\|Xx\| - \|PEx\|)^2.\end{aligned}$$

Since  $\|Xx\| \geq \sigma_i^2$ , we have

$$(10) \quad \mu \geq \begin{cases} (\sigma_i - \|PE\|)^2 & \text{if } \sigma_i \geq \|PE\| \\ 0 & \text{if } \sigma_i \leq \|PE\| \end{cases}.$$

The theorem now follows on combining (7), (9), and (10). ■

We make three observations on this theorem. First, there is a trivial variant in which  $PE$  is replaced by  $ER$ , where  $R$  is the orthogonal projection onto the row space of  $X$ .

Second, when  $\sigma_i$  is reasonably larger than  $\|E\|$ , say  $\sigma_i > 5\|E\|$ , the first term in the bound (4) dominates and we have

$$\tilde{\sigma}_i \approx \sigma_i + \xi_i,$$

where  $\xi_i$  satisfies (5). The classical perturbation result cited at the beginning of this note would give  $|\xi_i| \leq \|E\|$ . Since  $\|PE\| \leq \|E\|$ , our result is sharper; and in fact when  $n \gg p$  we may expect  $PE$  to be significantly smaller than  $\|E\|$ , so that (5) represents a true improvement over the classical result.

The third and perhaps most interesting observation is that when  $\sigma_j$  is of order  $\|E\|$ , the term  $\eta_j$  will tend to dominate. Now  $\eta_j$  always represents an increase in the singular value, and when  $n \gg p$  this increase can be significant, depending as it does on  $(I-P)E$ . To put the matter in other words, if one takes a matrix with a small singular value and perturbs it by quantities of the same size as that singular value, then one can expect the singular value to increase, not decrease. This tendency toward better conditioned matrices with larger  $\sigma_p$  has been observed in practice in connection with a regression problem in which simulated random perturbations in the data seriously biased the regression coefficients [1,3].



### References

1. A. E. Beaton, D. B. Rubin, and J. L. Barone, The acceptability of regression solutions: another look at computational stability, J. Amer. Stat. Assoc. 71 (1976) 158-168.
2. G. W. Stewart, Introduction to Matrix Computations, Academic Press, New York (1974).
3. \_\_\_\_\_, Another look at the Longley data set, University of Maryland, Computer Science Technical Report TR-(1978).