SOME SIMPLE ESTIMATES FOR SINGULAR VALUES OF A MATRIX

by

Liqun Qi

Computer Sciences Technical Report #472

May 1982

Some Simple Estimates for Singular Values of a Matrix*

Liqun Qi**

Computer Sciences Department
University of Wisconsin-Madison
Madison, Wisconsin 53706

Abstract

Some simple estimation theorems for singular values of a rectangular matrix A are given. They only use the elements of A itself, and in some cases they yield better results than does the Gerschgorin theorem applied to A*A. A bound for the condition number of A may be obtained from them. When A is square a bound is derived which explains why scaling improves the performance of Gauss elimination when row or column norms differ widely in magnitude. Their application to perturbation theory of singular values is also discussed.

Subject Classification: AMS(MOS): 15A18, 65F99; CR: 5.14.

---

*Sponsored by the National Science Foundation under Grant No. MCS7901066, Mod. 2.

**Permanent Address: Department of Applied Mathematics, Tsinghua University, Beijing, The People's Republic of China.

1. Introduction

For eigenvalues of a square matrix $A = (a_{ij})$ there is a widely used theorem, the Gerschgorin theorem [5].

Theorem 1. (Gerschgorin) Let $A = (a_{ij}) \in C^{nxn}$, then each eigenvalue of A lies in one of the disks in the complex plane

$$(1.1) \qquad D_i = \{\lambda : |\lambda - a_{ii}| \le r_i := \sum_{\substack{j=1 \\ j \ne i}}^{n} |a_{ij}|\}, \quad i=1,\dots,n.$$

Furthermore, if k disks constitute a connected region but are disconnected from the other $n - k$ disks, then exactly k eigenvalues lie in this region.

For singular values [4] of a rectangular matrix A, we can apply the Gerschgorin theorem to A^*A to get estimates. However, there are two disadvantages: (1) it is a little complicated to use the elements of A^*A; (2) the smallest singular value will be very badly conditioned in this process [1]. In many cases, we cannot use this process to give a non-zero lower bound for the smallest singular value.

In Section 2, we give an estimation theorem for the singular values of a rectangular matrix A. This estimation theorem only uses the elements of A itself. For a square matrix $A = (a_{ij})$, this theorem simply uses the n real intervals

$$(1.2) \qquad B_i = [(|a_{ii}| - s_i)_+, \ |a_{ii}| + s_i], \quad i=1,\dots,n$$

to replace the n disks in Theorem 1, where

$$(1.3) \qquad s_i = \max \Big(\sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ij}|, \sum_{\substack{j=1 \\ j\neq i}}^{n} |a_{ji}| \Big), \; i=1,\ldots,n$$

and for a real member a, we denote

$$a_+ := \max(0,a).$$

A simple example shows that this theorem gives a sharper bound for the smallest singular values of A than the Gerschgorin theorem applied to A^*A. In fact, it gives an upper bound for the condition number of A though the Gerschgorin theorem does not work for this example.

In Section 3, a sharper estimation theorem is given. A few square root operations improve the results up to a factor of 2.

In Section 4, the scaling technique is discussed. In Section 5, another simple estimate for the largest and the smallest singular value is given. In Section 6, the diagonalization technique is discussed. They can be combined with the theorem in Sections 2 and 3 to improve the results.

In Section 5, we also use the estimate to explain the fact, observed in practice, that scaling improves the performance of the Gaussian elimination method for linear equations.

In Section 7, an application to perturbation theory of singular values is given.

2. A Gerschgorin-Type Theorem

Suppose $A = (a_{ij}) \in C^{m\times n}$. Write

$$r_i := \sum_{\substack{j=1\\ j\neq i}}^{n} |a_{ij}|, \quad c_i := \sum_{\substack{j=1\\ j\neq i}}^{m} |a_{ji}|, \tag{2.1}$$

$$s_i := \max(r_i, c_i), \quad a_i := |a_{ii}|,$$

for $i=1,2,\ldots,\min(m,n)$. For $m \neq n$, define

$$s := \begin{cases} \max\limits_{n+1\leq i\leq m} \{\sum\limits_{j=1}^{n} |a_{ij}|\}, & \text{for } m > n \\ \\ \max\limits_{m+1\leq i\leq n} \{\sum\limits_{j=1}^{m} |a_{ji}|\}, & \text{for } m < n \end{cases}$$

We give the theorem for $m \geq n$. For $m < n$, the result is similar.

Theorem 2. With the above notation, each singular value of A lies in one of the real intervals

$$B_i = [(a_i - s_i)_+, \ a_i + s_i], \ i=1,\ldots,n$$

(2.2)

$$B_{n+1} = [0,s].$$

If $m = n$ or if $m > n$ and $a_i \geq s_i + s$, $i=1,\ldots n$ then B_{n+1} is not needed in the above statement. Furthermore, every component interval of the union of B_i, $i=1,2,\ldots,n+1$ (n for $m = n$) contains exactly k singular values if it contains k intervals of $B_1,\ldots,B_n$.

Proof. Suppose λ is a singular value of A. Then there are two nonzero vectors $x \in C^n$, $y \in C^m$, $x = (x_1,\ldots,x_n)$, $y = (y_1,\ldots,y_m)$ such that

$$\lambda x = A^* y,\ \lambda y = Ax.$$

Suppose $|y_i| = \max\ \{|x_1|,\ldots,|x_n|,|y_1|,\ldots,|y_m|\}$ (it is similar if the maximum is attained by $|x_i|$). If $m > n$ and $i > n$, then

$$\lambda y_i = \sum_{j=1}^{n} a_{ij} x_j$$

$$|\lambda| \le \sum_{j=1}^{n} |a_{ij}| \le s,$$

i.e., $\lambda \in [0,s]$.

Suppose $i \le n$. Then

$$\lambda x_i - \bar{a}_{ii} y_i = \sum_{\substack{j=1 \\ j \ne i}}^{m} \bar{a}_{ji} y_i \tag{2.3}$$

$$\lambda y_i - a_{ii} x_i = \sum_{\substack{j=1 \\ j \ne i}}^{n} a_{ij} x_j \tag{2.4}$$

Write $\delta = \frac{x_i}{y_i}$, then

$$|\lambda\delta - \bar{a}_{ii}| \le \sum_{\substack{j=1 \\ j \ne i}}^{m} |a_{ji}| = c_i \tag{2.5}$$

$$|\lambda - \delta a_{ii}| \le \sum_{\substack{j=1 \\ j \ne i}}^{n} |a_{ij}| = r_i \tag{2.6}$$

If $\lambda \ge a_i$, then

$$|\lambda - a_i \quad | \le |\lambda - |\delta|\ a_i \quad | \le |\lambda - \delta a_{ii}| \le r_i.$$

If $\lambda \leq a_i$, then

$$|\lambda - a_i| \leq ||\delta|\lambda - a_i| \leq |\lambda\delta - \bar{a}_{ii}| \leq c_i.$$

In any case, we have

$$|\lambda - a_i| \leq s_i.$$

Since $\lambda \geq 0$, we know $\lambda \in B_i$. This proves the first part of the theorem.

Since singular values are the square roots of eigenvalues of A^*A, therefore, they are also continuous functions of the elements of A [4] [5]. Consider $D + \varepsilon B$, where $D = \hat{D}$ for $m = n$, $D = \binom{\hat{D}}{0} \in R^{m\times n}$ for $m > n$, $\hat{D} = \mathrm{diag}(a_{11},\ldots,a_{nn})$, $B = A - D$. Let ε change continuously from 0 to 1; we get the whole conclusion of this theorem.

We use the simple example $A = \begin{pmatrix} 10 & 1 \\ 0 & 3 \end{pmatrix}$ to compare this theorem and the Gerschgorin theorem applied to A^*A.

(1) Apply the Gerschgorin theorem to $A^*A = \begin{pmatrix} 100 & 10 \\ 10 & 10 \end{pmatrix}$. Suppose the singular values of A are λ_1 and λ_2, $\lambda_1 \geq \lambda_2$. We know that

$$\lambda_1^2 \in [90,\ 110],\ \lambda_2^2 \in [0\ ,20],$$

i.e.,

$$\lambda_1 \in \{9.486,\ 10.489],\ \lambda_2 \in [0,\ 4.473].$$

Since the condition number $k(A)$ of A in the Euclidean norm is $\frac{\lambda_1}{\lambda_2}$, we get a lower bound for $k(A)$ but no upper bound for $k(A)$:

$$k(A) \in [\frac{9.487}{4.473}, +\infty) = [2.120, +\infty).$$

(2) Apply Theorem 2 directly to A. We get

$$\lambda_1 \in [9, 11], \lambda_2 \in [2, 4],$$

$$k(A) \in [\frac{9}{4}, \frac{11}{2}] = [2.25, 5.5].$$

We see that Theorem 2 is not only simpler, especially for a large matrix, but also better in the above situations.

3. A Sharper Theorem

Theorem 2 is simple enough. However, we can get a sharper estimate by a few square root operations.

Theorem 3. Theorem 2 remains true if we replace B_i, $i=1,\ldots,n$ by

(3.1) $$G_i = [\ell_{i_+}, u_i], \ i=1,\ldots,n$$

where

$$\ell_i = \min \left\{ \sqrt{a_i^2 - a_i r_i + \frac{c_i^2}{4}} - \frac{c_i}{2}, \right.$$

$$\left. \sqrt{a_i^2 - a_i c_i + \frac{r_i^2}{4}} - \frac{r_i}{2} \right\},$$

$$u_i = \max \left\{ \sqrt{a_i^2 + a_i r_i + \frac{c_i^2}{4}} + \frac{c_i}{2}, \right.$$

$$\left. \sqrt{a_i^2 + a_i c_i + \frac{r_i^2}{4}} + \frac{r_i}{2} \right\},$$

for $i=1,\ldots,n$, where if one of the numbers in the minimum is not real, we omit it.

Proof. Following the same argument as in the proof of Theorem 2, we get (2.3) and (2.4). Substitute (2.3) in (2.4). We get

$$\lambda^2 y_i - a_i^2 y_i = \lambda \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ij} x_j + a_{ii} \sum_{\substack{j=1 \\ j\neq i}}^{m} \bar{a}_{ji} y_j$$

$$|\lambda^2 - a_i^2| \leq \lambda r_i + a_i c_i$$

Suppose $\lambda \geq a_i$; we have

$$\lambda^2 - a_i^2 \leq \lambda r_i + a_i c_i$$

or

$$\lambda^2 - r_i \lambda - a_i^2 - a_i c_i \leq 0$$

The necessary condition is

$$\lambda \leq \sqrt{a_i^2 + a_i c_i + \frac{r_i^2}{4}} + \frac{r_i}{2}$$

Similarly, if $\lambda \leq a_i$, we have

$$\lambda \geq \sqrt{a_i^2 - a_i c_i + \frac{r_i^2}{4}} - \frac{r_i}{2}$$

If $|x_i| = \max\{|x_1|,\dots,|x_n|;\ |y_1|,\dots,|y_m|\}$, we get

$$\lambda \leq \sqrt{a_i^2 + a_i r_i + \frac{c_i^2}{4}} + \frac{c_i}{2},$$

$$\lambda \geq \sqrt{a_i^2 - a_i r_i + \frac{c_i^2}{4}} - \frac{c_i}{2}.$$

(Notice that if $a_i^2 - a_i c_i + \frac{r_i^2}{4} < 0$, we have $\lambda \geq a_i$, the same argument applies to above.) Combining them together, we get the conclusion. The other part of the proof of Theorem 2 is not changed, including the argument about B_{n+1}.

Remark. If $r_i = c_i$, then $\ell_i = a_i - s_i$, $u_i = a_i + s_i$. If $r_i \neq c_i$, then $\ell_i > a_i - s_i$, $u_i < a_i + s_i$. Therefore, we get a sharper result. Applying this theorem to the example in Section 2, $A = \begin{pmatrix} 10 & 1 \\ 0 & 3 \end{pmatrix}$ we get

$$\lambda_1 \in [9.486,\ 10.513],\ \lambda_2 \in [2.449,\ 3.542]$$

$$k(A) \in \left[\frac{9.486}{3.542},\ \frac{10.513}{2.449}\right] = [2.678,\ 4.293].$$

This is better than the results in Section 2. For this matrix, the true singular values are

$$\lambda_1 = 10.05474,\ \lambda_2 = 2.98367$$

$$k(A) = 3.370.$$

4. Scaling

In practice, Theorem 2 and 3 can be combined with scaling techniques to improve the results.

Theorem 4. Theorem 2 and 3 remain true if we replace the definition of r_i and c_i in (2.1), and s by

$$r_i := \sum_{\substack{j=1\\j\neq i}}^{n} \frac{k_j}{k_i}|a_{ij}|,\quad c_i := \sum_{\substack{j=1\\j\neq i}}^{m} \frac{k_j}{k_i}|a_{ji}|, \tag{4.1}$$

for $i=1,2,\ldots,\min(m,n)$, and

$$s := \begin{cases} \max\limits_{n+1\leq i\leq m} \{\sum_{j=1}^{n} \frac{k_j}{k_i}|a_{ij}|\}, & \text{for } m > n, \\ \max\limits_{m+1\leq i\leq n} \{\sum_{j=1}^{m} \frac{k_j}{k_i}|a_{ji}|\}, & \text{for } m < n, \end{cases}$$

where k_i, $i=1,2,\ldots,\max(m,n)$, are any positive numbers.

Proof. Let $x_i = k_i\hat{x}_i$, $i=1,\ldots,n$, $y_i = k_i\hat{y}_i$, $i=1,\ldots,m$. Then our fundamental equations become

$$\lambda\hat{x}_i = \sum_{j=1}^{m} \bar{a}_{ji} \frac{k_j}{k_i} \hat{y}_j,\quad i=1,\ldots,n,$$

$$\lambda\hat{y}_i = \sum_{j=1}^{n} a_{ij} \frac{k_j}{k_i} \hat{x}_j,\quad i=1,\ldots,m.$$

Considering the maximum of $\{|\hat{x}_1|,\ldots,|\hat{x}_n|,|\hat{y}_1|,\ldots,|\hat{y}_m|\}$, and using the same technique of the proofs of Theorem 2 and

Theorem 3, we get the conclusion.

Applying this idea to our simple example $A = \begin{pmatrix} 10 & 1 \\ 0 & 3 \end{pmatrix}$, we get a fairly good result. We get from Theorem 2 and Theorem 4 that two singular values lie in

$$B_1 = [10-\frac{k_2}{k_1},\ 10+\frac{k_2}{k_1}],\ B_2 = [3-\frac{k_1}{k_2},\ 3+\frac{k_1}{k_2}].$$

Let $d = \frac{k_1}{k_2}$. Then

$$B_1 = [10-d^{-1},\ 10+d^{-1}],\ B_2 = [3-d,\ 3+d].$$

The best lower bound for λ_1 is

$$\ell_1 = 10 - d^{-1} = u_2 = 3 + d = 9.854, \quad \text{for} \quad d^{-1} = 0.146.$$

The best upper bound for λ_1 is

$$u_1 = 10 + d^{-1} = u_2 = 3 + d = 10.140, \quad \text{for} \quad d^{-1} = 0.140.$$

The best lower bound for λ_2 is

$$\ell_2 = 3 - d = \ell_1 = 10 - d^{-1} = 2.860, \quad \text{for} \quad d = 0.140.$$

The best upper bound for λ_2 is

$$u_2 = 3 + d = \ell_1 = 10 - d^{-1} = 3.146, \quad \text{for} \quad d = 0.146.$$

Therefore, we have

$$\lambda_1 \in [9.854,\ 10.140],\ \lambda_2 \in [2.860,\ 3.46],$$

$$k(A) \in [3.132,\ 3.545].$$

If we use Theorem 3 and Theorem 4, we get even better results:

$$\ell_1 = 9.944, \quad u_2 = 9.940, \quad \text{for } d = 9.035$$

$$u_1 = 10.05475, \quad u_2 = 10.05415, \quad \text{for } d = 9.159$$

$$\ell_1 = 3.0577, \quad u_2 = 3.0557 \quad \text{for } d^{-1} = 9.065$$

$$\ell_1 = 2.9496, \quad \ell_2 = 2.9447 \quad \text{for } d^{-1} = 9.130.$$

Therefore, we have

$$\lambda_1 \in [9.944, 10.05475], \; \lambda_2 \in [2.9447, 3.0557]$$

$$k(A) \in [3.254, 3.415].$$

Varga and Levinger have discussed the minimal Gerschgorin set of a square matrix in [6], [3], [2], [7]. One can hope that there may be a similar theory for the singular values.

5. An Estimate for the Largest and the Smallest Singular Values

In practice, the most important estimates are for the largest and the smallest singular values. The following simple theorem is useful as a supplement to the above theorems.

Let $a_{i.}$ denote the ith row of A, $a_{.j}$ denote the jth column of A. All the norm used here is the Euclidean norm.

Theorem 5. With the same notation, a lower bound for the largest singular value is

$$(5.1) \qquad \ell_{max} = \max \{ \max_{1 \le i \le m} \{||a_{i.}||\}, \max_{1 \le j \le n} \{||a_{.j}||\}\}$$

and an upper bound for the smallest singular value when $m = n$ is

$$(5.2) \qquad u_{min} = \min \{ \min_{1 \le i \le m} \{||a_{i.}||\}, \min_{1 \le j \le n} \{||a_{.j}||\}\}$$

Proof. The largest singular value is $||A||$, and

$$||A|| \ge \max_{1 \le j \le n} \{||Ae_j||\} = \max_{1 \le j \le n} \{||a_{.j}||\},$$

$$||A|| = ||A^*|| \ge \max_{1 \le i \le m} \{||A^* e_i'||\} = \max_{1 \le i \le m} \{||a_{i.}||\},$$

where $e_j \in C^n$, $j=1,\dots,n$, $e_i' \in C^m$, $i=1,\dots,m$ are unit vectors. This proves (5.1). For (5.2), if A is singular, the smallest singular value is zero and (5.2) holds. If A is nonsingular, the smallest singular value is $(||A^{-1}||)^{-1}$. We have

$$||A^{-1}|| \geq \max_{1 \leq j \leq n} \{ \frac{||e_j||}{||Ae_j||} \} = \max_{1 \leq j \leq n} \{ ||a_{\cdot j}||^{-1} \}$$

and

$$||A^{-1}|| = ||(A^*)^{-1}|| \geq \max_{1 \leq i \leq m} \{ \frac{||e_i'||}{||A^* e_i'||} \} = \max_{1 \leq i \leq m} \{ ||a_{i \cdot}||^{-1} \},$$

which leads to (5.2). This proves the theorem.

Applying this theorem to $A = \begin{pmatrix} 10 & 1 \\ 0 & 3 \end{pmatrix}$, we get ℓ_{min} = 10.04987. Combining this with the results we obtained from Theorem 3 and Theorem 4, we have

$$\lambda_1 \in [10.04987, 10.05475], \lambda_2 \in [2.9447, 3.0557]$$

$$k(A) \in [3.289, 3.415]$$

This is a sharp result.

Notice that (5.2) is not true in general when $m \neq n$.

On page 46 of [9], there is a footnote which says that in Gaussian elimination for $Ax = f$, "experience indicates that we usually achieve greater accuracy in the single precision solution, if we first scale the matrix A. That is, if with $B = D_1AD_2$, we solve

$$By = D_1f$$

for y, and then determine x from $D_2y = x$. Here D_1 and D_2 are some diagonal matrices chosen so that the n columns and the n rows of the matrix B have approximately equal norms. A complete mathematical explanation of this phenomenon is not available." Theorem 5 tells us that if the norms of the n columns and n rows differ greatly in magnitude, the condition number cannot be small. Therefore, it explains the phenomenon to some extent.

6. Diagonalization

Usually, A is not a diagonally dominant matrix. We can multiply A by several simple orthogonal matrices to reduce the sum of the squares of its off-diagonal elements, since such multiplications preserve the singular values.

To simplify the discussion, in this section, we restrict A to be a real matrix.

Suppose $m \geq n$ (the argument is similar if $m < n$). Consider

$$\max_{i \neq j} \{a_{ij}^2 + a_{ji}^2\}, \tag{6.1}$$

where we let $a_{ji} = a_{ij}$ for $i > n$, $j \leq n$.

(a) If the maximum is attained by (r,s), $r > n$, then the treatment is relatively simple. Let

$$B = PA, \tag{6.2}$$

where $P = (p_{ij}) \in R^{m \times m}$, $p_{ij} = \delta_{ij}$ except

$$p_{rr} = p_{ss} = \frac{a_{ss}}{\sqrt{a_{rs}^2 + a_{ss}^2}}, \quad p_{sr} = -p_{rs} = \frac{a_{rs}}{\sqrt{a_{rs}^2 + a_{ss}^2}} \tag{6.3}$$

It is easy to see that P is an orthogonal matrix, $B = (b_{ij}) \in R^{m \times n}$ has the same set of singular values as A, $\sum_{i,j} b_{ij}^2 = \sum_{i,j} a_{ij}^2$, $b_{ij} = a_{ij}$ for $i, j \neq r$ or s, $b_{rs} = 0$, $b_{ss}^2 = a_{rs}^2 + a_{ss}^2$, i.e.,

$$\sum_{i\neq j} b_{ij}^2 = \sum_{i\neq j} a_{ij}^2 - a_{rs}^2 \leq \left(1-\frac{1}{(m-1)n}\right) \sum_{i\neq j} a_{ij}^2 \tag{6.4}$$

(b) If the maximum is attained by (r,s), $r, s \leq n$, then let

$$B = PAQ \tag{6.5}$$

where $P = (p_{ij}) \in R^{m\times m}$, $p_{ij} = \delta_{ij}$ except

$$p_{rr} = p_{ss} = \cos\theta,\ p_{sr} = -p_{rs} = \sin\theta, \tag{6.6}$$

$$\tan 2\theta = \frac{2(a_{rr}a_{sr} + a_{rs}a_{ss})}{a_{sr}^2 + a_{ss}^2 - a_{rr}^2 - a_{rs}^2},\ -\frac{\pi}{4} \leq \theta \leq \frac{\pi}{4}$$

and $Q = (q_{ij}) \in R^{n\times n}$, $q_{ij} = \delta_{ij}$ except

$$q_{rr} = q_{ss} = \cos\phi,\ q_{sr} = -q_{rs} = \sin\phi, \tag{6.7}$$

$$\tan 2\phi = \frac{2(a_{rr}a_{rs} + a_{sr}a_{ss})}{a_{rr}^2 + a_{sr}^2 - a_{rs}^2 - a_{ss}^2},\ -\frac{\pi}{4} \leq \phi \leq \frac{\pi}{4}$$

It is also not difficult to verify that P and Q are orthogonal matrices, $B = (b_{ij}) \in R^{m\times n}$, $b_{ij} = a_{ij}$ for $i, j \neq r$ or s, $\sum_{i,j} b_{ij}^2 = \sum_{i,j} a_{ij}^2$, $b_{rs} = b_{sr} = 0$, and $b_{rr}^2 + b_{ss}^2 = a_{rr}^2 + a_{rs}^2 + a_{sr}^2 + a_{ss}^2$.

Therefore,

$$\sum_{i\neq j} b_{ij}^2 = \sum_{i\neq j} a_{ij}^2 - a_{rs}^2 - a_{sr}^2 \leq \left(1-\frac{2}{(m-1)n}\right) \sum_{i\neq j} a_{ij}^2. \tag{6.8}$$

We can repeat this process to diagonalize A, until we can apply Theorem 2, 3, 4 and 5 to get a satisfactory estimate for singular values of A. Notice that this process reduces the sum of the squares of off-diagonal elements geometrically by a factor less than $(1-\frac{1}{(m-1)n})$.

Another way to treat the singular value problem is first to transform A into an upper triangular matrix by Householder transformations:

$$B = P_h \ldots P_1 A, \; B = \binom{\hat{B}}{0}, \; \hat{B} \in R^{nxn},$$

where $P_1, \ldots, P_h$ are Householder matrices [3], $\hat{B}$ is upper triangular. Then we use the above technique to diagonalize $\hat{B}$.

7. Application to Perturbation Theory

Like the Gerschgorin theorem, our theorems can be applied to the perturbation theory of singular values. (Compare the discussion of [8, p. 72-81]).

Suppose $A = (a_{ij}) \in C^{m\times n}$, $B = (b_{ij}) \in C^{m\times n}$, and A has the singular decomposition $A = PDQ^*$, where

$$P = (p_1, p_2, \ldots, p_m) \in C^{m\times m},\ P^*P = I_m,$$

$$Q = (q_1, q_2, \ldots, q_n) \in C^{n\times n},\ Q^*Q = I_n,$$

$$D = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ 0 & & \ddots & \\ & & & \lambda_n \\ & 0 & & \end{bmatrix} \in C^{m\times n}$$

(Here, we suppose $m \geq n$. The argument is similar if $m < n$).

<u>Theorem 6</u>. If λ is a simple singular value of A, then $A + \varepsilon B$ has a singular value

$$\lambda = \lambda_i + \varepsilon\frac{p_i^* B q_i}{p_i^* q_i} + o(\varepsilon^2).$$

<u>Proof</u>.

$$P^*(A+\varepsilon B)Q = D + \varepsilon P^*BQ = D + \varepsilon\left(\frac{p_i^* B q_j}{p_i^* q_j}\right)$$

Applying Theorem 2 and Theorem 4 with $k_j = \varepsilon$, for $j \neq i$,

$$k_i = \frac{1}{2b} \min_{j \neq i} (|\lambda_j - \lambda_i|) \quad \text{where} \quad b = \max_{\substack{1 \le i \le m \\ 1 \le j \le n}} |\beta_{ij}|, \ (\beta_{ij}) := P^*BQ,$$

we can construct an interval around $\lambda_i + \varepsilon \frac{p_i^* B q_i}{p_i^* q_i}$ with length $o(\varepsilon^2)$. This interval is separated from other intervals. Thus, we prove the theorem.

Acknowledgment

I am thankful to Professor B. Noble who essentially suggested Theorems 4 and 6, and to Professor S. Robinson who carefully read the manuscript and made some good suggestions. I thank the referee who pointed out that most of the above results do not depend on assuming that A is real since my first draft has this assumption. I am also grateful to Professor R. Varga for his nice help.

References

[1] G. Golub and W. Kahan, Calculating the singular values and pseudo-inverse of a matrix, SIAM J. Numer. Anal., 2 (1965), pp. 205-224.

[2] B. W. Levinger, Minimal Gerschgorin sets, III, Linear Alg. Appli., 2 (1969), pp. 13-19.

[3] B. W. Levinger and R. S. Varga, Minimal Gerschgorin sets, II, Pacific J. Math., 17 (1966), pp. 199-210.

[4] B. Noble and J. W. Daniel, Applied Linear Algebra, 2nd Ed., Prentice-Hall, Englewood Cliffs, N.J., 1977.

[5] G. W. Stewart, Introduction to Matrix Computations, Academic Press, N.Y., 1973.

[6] R. S. Varga, Minimal Gerschgorin sets, Pacific J. Math., 15 (1965), pp. 719-729.

[7] R. S. Varga, Minimal Gerschgorin sets for partitioned matrices, SIAM J. Numer. Anal., 7 (1970), pp. 493-507.

[8] J. H. Wilkinson, The Algebraic Eigenvalue Problems, Clarendon Press, Oxford, 1965.

[9] E. Isaacson and H. B. Keller, Analysis of Numerical Methods, John Wiley & Sons, Inc., N. Y., 1966.