International Symposium on Nonlinear Theory and its Applications
Xi'an, PRC, October 7–11, 2002

Fast Inclusion Method for Singular Values of Complex Matrices

Kosuke MARUYAMA and Shin'ichi OISHI

Waseda University
3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555, Japan
E-mail: maruyama@oishi.info.waseda.ac.jp, oishi@oishi.info.waseda.ac.jp

1. Introduction

In this report, we are concerned with accuracy of calculated singular values obtained by

$$B = U\Sigma V \tag{1}$$

where B is an $m \times n$ complex matrix, Σ an $m \times n$ diagonal matrix whose diagonal elements σ_i are singular values of B, U an $m \times m$ Hermite matrix and V an $n \times n$ Hermite matrix. Numerical softwares usually produce an approximation $\tilde{\sigma}$ of a singular value of B. In this report, we shall discuss how to verify the existence of a true singular value σ^* of (1) and how to get rigorous error bound between $\tilde{\sigma}$ and σ^*.

For singular values of real matrices, an inclusion method by means of Weyl's perturbation theory has been presented [1]. In this report, based on this Oishi's method, we shall show an extension to apply this method to complex matrices.

2. Rounding Mode Controlled Computation [1]

In this section, we shall show that a very sharp upper bound of $||BCD - E||_\infty$ can be calculated rigorously via floating point arithmetic computation under suitably controlled rounding mode.

2.1. Floating Point System

For the purpose of fast numerical computation, let us discuss on floating point numbers floating point arithmetic. Let R be the set of real numbers. Let F be a set of floating point numbers. In the following, we assume that the floating point arithmetic considered in this paper satisfies the following assumptions. The floating point systems abiding by the IEEE standard 754 satisfy following assumptions:

First, we assume that F is symmetric, *i.e.*, $x \in \mathrm{F} \rightarrow -x \in \mathrm{F}$, such that $|x|$ is exact for $x \in \mathrm{F}$

Let $c \in \mathrm{R}$. Next, we assume that the following two types of rounding modes are defined:

1. **(1) rounding to $-\infty$** Round $c \in \mathrm{R}$ to the largest floating point number $f \in \mathrm{F}$ satisfying $f \leq c$. We represent it by $\bigtriangledown : \mathrm{R} \rightarrow \mathrm{F}$.

2. **(1) rounding to $+\infty$** Round $c \in \mathrm{R}$ to the smallest floating point number $f \in \mathrm{F}$ satisfying $f \geq c$. We represent it by $\bigtriangleup : \mathrm{R} \rightarrow \mathrm{F}$.

We denote the rounding operator as $\bigcirc : \mathrm{R} \rightarrow \mathrm{F}$. Then, $\bigcirc$ is one of $\bigtriangledown$ or $\bigtriangleup$.

Then, we assume that the floating point arithmetic is defined through the following formula:

$$\begin{array}{l} x \odot y = \bigcirc(x, y), \text{ for any } x, y \in \mathrm{F} \\ \text{for } \cdot \in \{+, -, \times, /\} \text{ and } \bigcirc \in \{\bigtriangledown, \bigtriangleup\}. \end{array} \tag{2}$$

Here, left hand side of formula (2), $\odot$ means the floating point arithmetic and in the right hand side of this formula, $x \cdot y$ represents the ordinary real arithmetic.

Moreover, we assume that if we take a floating point square root of any $x \in \mathrm{F}$ in the rounding mode to $+\infty$, then it is greater than or equal to the ordinary square root of x.

2.2. Verified Upper Bound of $||BCD - E||$

Let B, C, D and E be $n \times n$ complex matrices with elements in F. We shall present algorithms of computing rigorous upper bound of $||BC - D||_\infty$ and $||BCD - E||_\infty$. For this purpose, we need the following result. Let F, R and S be $n \times n$ matrices. If all elements of both real and imaginary part of R are nonnegative (in the following, $R \geq 0$ denotes this property), it is known that the following formula holds:

$$[F - R, F + R]S = [FS - R|S|, FS + R|S|]. \tag{3}$$

Here, —S— is the matrix with the $i - j$ elements $|S_{ij}|$, S_{ij} being the $i - j$ element of S.

The following script calculates an inclusion of BC.

function $[\underline{P},\overline{P}]$=**cproduct**$(B,C)$
B_r =**real**(B);
B_i =**imag**(B);
C_r =**real**(C);
C_i =**imag**(C);
down;
$\underline{P_r} = B_r * C_r + (-B_i) * C_i$;
$\underline{P_i} = B_r * C_i + B_i * C_r$;
up;
$\overline{P_r} = B_r * C_r + (-B_i) * C_i$;
$\overline{P_i} = B_r * C_i + B_i * C_r$;
$\underline{P} = \underline{P_r} + (i * \underline{P_i})$;
$\overline{P} = \overline{P_r} + (i * \overline{P_i})$;

Algorithm 1.
Rigorous inclusion of BC

Here, the instruction **up** means to adopt the mode of rounding to the $+\infty$ and the instruction **down** to adopt the mode of rounding to the $-\infty$. We assume the rounding mode is changed, it remains unchanged until the next instruction of changing rounding mode appears.

The following script calculates a rigorous upper bound of $||BC - D||_\infty$

function c=**cerrb**(B, C, D)
$[\underline{H}, \overline{H}]$=**cproduct**$(B, C)$;
$\underline{H_r}$=**real**$(\underline{H})$;
$\underline{H_i}$=**imag**$(\underline{H})$;
$\overline{H_r}$=**real**$(\overline{H})$;
$\overline{H_i}$=**imag**$(\overline{H})$;
D_r=**real**(D);
D_i=**imag**(D);
down;
$\underline{H_r} = \underline{H_r} - D_r$;
$\underline{H_i} = \underline{H_i} - D_i$;
up;
$\overline{H_r} = \overline{H_r} - D_r$;
$\overline{H_i} = \overline{H_i} - D_i$;
$H_a = \underline{H_r} + (i * \underline{H_i})$;
$H_b = \underline{H_r} + (i * \overline{H_i})$;
$H_c = \overline{H_r} + (i * \underline{H_i})$;
$H_d = \overline{H_r} + (i * \overline{H_i})$;
$H\mathrm{max}$=**max** (**abs**(H_a),**abs**(H_b),**abs**(H_c),**abs**(H_d));
c=**norm**$(H\mathrm{max}$,**inf**);

Algorithm 2.
Rigorous upper bound of $||BC - D||_\infty$

Here, **max**(**abs**(H_a), **abs**(H_b), **abs**(H_c), **abs**(H_d)) produces the matrix with the $i - j$ element $H\mathbf{max}_{ij} = \max\{|H_a|, |H_b|, |H_c|, |H_d|\}$ The instruction **norm**$(H$, **inf**) is to calculate the $||H\mathrm{max}||_\infty$:

$$||H\mathrm{max}||_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |H\mathrm{max}_{ij}| \tag{4}$$

The next script calculates $n \times n$ matrices F and R satisfying $[\underline{P}, \overline{P}] \subset [F - R, F + R]$ [1].

function $[F, R]$=**ccr**$(\underline{P}, \overline{P})$
$\underline{P_r}$=**real**$(\underline{P})$;
$\underline{P_i}$=**imag**$(\underline{P})$;
$\overline{P_r}$=**real**$(\overline{P})$;
$\overline{P_i}$=**imag**$(\overline{P})$;
up;
$F_r = \underline{P_r} + 0.5 * (\overline{P_r} - \underline{P_r})$;
$F_i = \underline{P_i} + 0.5 * (\overline{P_i} - \underline{P_i})$;
$R_r = F_r - \underline{P_r}$;
$R_i = F_i - \underline{P_i}$;
$F = F_r + (i * F_i)$;
$R = R_r + (i * R_i)$;

Algorithm 3.
Calculation of the center and the radius of $[\underline{P}, \overline{P}]$

Let A, F and R be $n \times n$ matrices with $R \ge 0$. The next script calculates $n \times n$ matrices $\underline{D}$ and $\overline{D}$ satisfying $[\underline{D}, \overline{D}] \supseteq A * [F - R, F + R]$. [1].

function $[\underline{D}, \overline{D}]$ =**s_i_product**(A, F, R)
G=**abs**(A);
down;
$\underline{D} = A * F + (-G) * R$;
up;
$\overline{D} = A * F + G * R$;

Algorithm 4.
Calculation of $A * [F - R, F + R]$

Here, **abs**(A) produces the matrix with th $i - j$ elements $|A_{ij}|$, A_{ij} being the $i - j$ element of the matrix A,

Let A, F and R be $n \times n$ complex matrices with $R \ge 0$. The following script calculates $n \times n$ complex matrices $\underline{D}$ and $\overline{D}$ satisfying $[\underline{D}, \overline{D}] \supseteq A * [F - R, F + R]$.

function $[\underline{D}, \overline{D}]$ =**cs_i_product_cr**(A, F, R)
A_r=**real**(A);
A_i=**imag**(A);
F_r=**real**(F);
F_i=**imag**(F);
R_r=**real**(R);
R_i=**imag**(R);
$[\underline{D_{r1}}, \overline{D_{r1}}]$ =**s_i_product_cr**(A_r, F_r, R_r);
$[\underline{D_{r2}}, \overline{D_{r2}}]$ =**s_i_product_cr**(A_i, F_i, R_i);
$[\underline{D_{i1}}, \overline{D_{i1}}]$ =**s_i_product_cr**(A_r, F_i, R_i);
$[\underline{D_{i2}}, \overline{D_{i2}}]$ =**s_i_product_cr**(A_i, F_r, R_r);
down;
$\underline{D_r} = \underline{D_{r1}} - \overline{D_{r2}}$;
$\underline{D_i} = \underline{D_{i1}} + \underline{D_{i2}}$;
up;
$\overline{D_r} = \overline{D_{r1}} - \underline{D_{r2}}$;
$\overline{D_i} = \overline{D_{i1}} + \overline{D_{i2}}$;
$\underline{D} = \underline{D_r} + (i * \underline{D_i})$;
$\overline{D} = \overline{D_r} + (i * \overline{D_i})$;

Algorithm 5.
Calculation of $A * [F - R, F + R]$
for complex matrices A, F and R

Combining Algorithms 1., 2., 3., 4. and 5. with the formula (3), we finally have the following script of calculating rigorous upper bound of $||BCD - E||_\infty$.

function t=**cerrt**(B, C, D, E)
$[\underline{P}, \overline{P}]$=**cproduct**$(C, D)$;
$[F, R]$=**ccr**$(\underline{P}, \overline{P})$;
$[\underline{H}, \overline{H}]$=**cs_i_product_cr**$(B, F, R)$;
E_r=**real**(E);
E_i=**imag**(E);
$\underline{H_r}$=**real**$(\underline{H})$;
$\underline{H_i}$=**imag**$(\underline{H})$;
$\overline{H_r}$=**real**$(\overline{H})$;
$\overline{H_i}$=**imag**$(\overline{H})$;
down;
$\underline{H_r} = \underline{H_r} - E_r$;
$\underline{H_i} = \underline{H_i} - E_i$;
up;
$\overline{H_r} = \overline{H_r} - E_r$;
$\overline{H_i} = \overline{H_i} - E_i$;
$H_a = \underline{H_r} + (i * \underline{H_i})$;
$H_b = \underline{H_r} + (i * \overline{H_i})$;
$H_c = \overline{H_r} + (i * \underline{H_i})$;
$H_d = \overline{H_r} + (i * \overline{H_i})$;
H=**max**([**abs**(H_a),**abs**(H_b), **abs**(H_c),**abs**(H_d)]);
t=**norm**$(H$,**inf**$)$;

Algorithm 6.
Rigorous upper bound of $||BCD - E||_\infty$

The algorithm 6. produces an inclusion $BCD - E \in [\underline{H}, \overline{H}]$. Moreover, the computed value t becomes a rigorous upper bound of $||BCD - E||_\infty$.

3. Perturbation theory for matrices

In order to derive mathematically rigorous results singular values via numerical computation, we have to modify the traditional perturbation theorems for matrix eigenvalues. In this section, we shall present one such theorem based on the Weyl theorem.

3.1. Weyl Type Perturbation Theorem

The following theorem is a modification of the Weyl theorem suited for numerical inclusion of matrix eigenvalues whose elements are complex values.

Theorem 1 *Let A be an $n \times n$ Hermite matrix. We assume that as a result of numerical computation we have an $n \times n$ real symmetric matrix D and $n \times n$ complex matrix P such that*

$$P^* DP = A + E,\ P^* P = I + F, \tag{5}$$

where E and F are matrices expressing computational errors. We assume that $||F||_2 \leq 1$ so that P is invertible. Let $\tilde{\lambda}$ and λ be eigenvalues of D and A, respectively, such that

$$\tilde{\lambda}_1 \leq \cdots \leq \tilde{\lambda}_n, \quad \lambda_1 \leq \cdots \leq \lambda_n. \tag{6}$$

Then, the following estimation holds.

$$|\lambda - \tilde{\lambda}| \leq |\tilde{\lambda}| ||F||_2 + ||E||_2. \tag{7}$$

In order to prove this theorem, we need two lemmas.

Lemma 1 (Weyl) *Let A and $A + E$ be both Hermite matrices and if*

$$\lambda_1 \leq \lambda_2 \cdots \lambda_n, \quad \hat{\lambda}_1 \leq \hat{\lambda}_2 \cdots \hat{\lambda}_n, \tag{8}$$

then the following estimation holds

$$|\lambda_i - \hat{\lambda}_i| \leq ||E||_2. \tag{9}$$

Lemma 2 (Eisenstat, Ipsen [2]) *Let real symmetric matrices D and $\tilde{D} = P^* DP$ have eigenvalues $\tilde{\lambda}$ and $\hat{\lambda}$, respectively, such that*

$$\tilde{\lambda}_1 \leq \tilde{\lambda}_2 \cdots \tilde{\lambda}_n \quad \hat{\lambda}_1 \leq \hat{\lambda}_2 \cdots \hat{\lambda}_n \tag{10}$$

If and $n \times n$ complex matrices P is nonsingular, then

$$|\hat{\lambda} - \tilde{\lambda}| \leq |\tilde{\lambda}_i| ||P^* P - I||_2. \tag{11}$$

Proof of Theorem 1

Let $\tilde{\lambda}_i$, $\hat{\lambda}_i$ and λ_i be eigenvalues of D, $P^* DP$ and A, respectively, such that

$$\begin{array}{l} \tilde{\lambda}_1 \leq \tilde{\lambda}_2 \cdots \tilde{\lambda}_n \\ \hat{\lambda}_1 \leq \hat{\lambda}_2 \cdots \hat{\lambda}_n \\ \lambda_1 \leq \lambda_2 \cdots \lambda_n. \end{array} \tag{12}$$

Using Lemmas 1 and 2, we have

$$\begin{array}{l} |\lambda_i - \tilde{\lambda}| \\ \leq |\lambda_i - \hat{\lambda}_i| + |\hat{\lambda}_i - \tilde{\lambda}| \\ \leq |\tilde{\lambda}_i| ||F||_2 + ||E||_2. \end{array} \tag{13}$$

∎

3.2. Perturbation Theory for Singular Values

From Theorem 1, we immediately have

Theorem 2 *Let B be $m \times n$ complex matrix. We assume that as a result of numerical computation we have an $m \times n$ real matrix D, an $m \times m$ matrix Y and $n \times n$ complex matrix X such that*

$$\begin{array}{l} Y^t D X = B + E, \\ X^t X - I = F, \quad Y^t Y - I = G \end{array} \tag{14}$$

where E, F and G are matrices expressing computational errors. We assume that $||F||_2 < 1$ and $||G||_2 < 1$ so that X and Y are invertible. Let $\tilde{\sigma}_i$ and σ_i be singular values of D and B, respectively, such that

$$\tilde{\sigma}_1 \geq \cdots \geq \tilde{\sigma}_n \geq 0, \quad \sigma_1 \geq \cdots \geq \sigma_n \geq 0 \tag{15}$$

Then , the following estimation holds:

$$|\sigma_i - \tilde{\sigma}_i| \leq |\tilde{\sigma}_i| \max(||F||_2, ||G||_2) + ||E||_2. \tag{16}$$

Proof of Theorem 2

Let

$$A = \begin{pmatrix} O & B^* \\ B & O \end{pmatrix}. \tag{17}$$

O is the zero matrix with a suitable size. The matrix A becomes a Hermite matrix. It is well known that for any matrix B, eigenvalues of A are the singular values of B. Let

$$P = \begin{pmatrix} X & O \\ O & Y \end{pmatrix}, \tag{18}$$

and

$$D = \begin{pmatrix} O & D^* \\ D & O \end{pmatrix}. \tag{19}$$

From equations (14),

$$P^* D P = A + \tilde{E}, \; P^* P = I + \begin{pmatrix} F & O \\ O & G \end{pmatrix}. \tag{20}$$

Here,

$$\tilde{E} = \begin{pmatrix} O & E^* \\ E & O \end{pmatrix}. \tag{21}$$

Thus, noticing $||\tilde{E}||_2 = ||E||_2$, from Theorem 1, we have a desired result. ∎

4. Verification of Singular Values

Let B an $m \times n$ complex matrix. If we apply numerical singular value decomposition method to B, then we may have a nearly unitary $m \times m$ matrix Y, a nearly unitary $n \times n$ matrix X and an $m \times n$ real matrix D such that.

$$\begin{array}{l} Y^t D X = B + E, \\ X^t X - I = F, \quad Y^t Y - I = G \end{array} \tag{22}$$

with error terms E, F and G. If $||F||_2 < 1$ and $||G||_2$ holds, the singular value perturbation theorem 2 provides a rigorous error bound between the singular values of D and B. Then, we have inclusions of all singular values of E as

$$\begin{array}{l} \sigma_i \in [\tilde{\sigma}_i - \epsilon, \tilde{\sigma}_i + \epsilon] \\ \epsilon = |\tilde{\sigma}_i| \max(||F||_2, ||G||_2) + ||E||_2). \end{array} \tag{23}$$

The following script calculates the estimation (23).

```
function e=cvsvd(B)
[Y,D,X]=svd(B);
S=diag(D);
Xt=X^t;
Yt=Y^t;
e1=cerrt(Y,D,Xt,B);
e2=cerrb(Y,Yt,eye(m));
e3=cerrb(X,Xt,eye(n));
ee=max(e2,e3);
up;
e=diag(diag(abs(S)) * ee + e1*eye(n));
```

Algorithm 7.
Rigorous upper bound of computed singular values

We assume that an $m \times n$ complex matrix B is given. The instruction **svd**(B) is to calculate singular value decomposition of B.

The following table shows example of calculating all singular values by this method. Results is obtained by GNU Free software Octave on the computer with Pentium III 800 Mhz CPU.

Table 1: Verification for complex matrices

Dimension m ($n = m$)	Time[s] (app.)	Time[s] (inclusion)	Error bound $\lVert\epsilon\rVert_\infty$
50	0.041	0.058	3.9×10^{-13}
100	0.31	0.30	2.5×10^{-12}
200	2.52	1.90	5.0×10^{-12}
500	38.2	22.5	2.3×10^{-11}
1000	301	156	8.2×10^{-11}

References

[1] Shin'ichi OISHI: "Fast Enclosure of Matrix Eigenvalues Via Rounding Mode Controlled Computation", Linear, Algebra and its Applications, 324 (2001) pp.133-146.

[2] Stanley C. Eisenstat and Ilse C. F. Ipsen: "Relative perturbation techniques for singular value problems", SIAM J. Numer. Anal., 32(6), pp.1972-1988 (1995).