A shift strategy for superquadratic convergence of the dqds algorithm for computing singular values

Graduate School of Information Science and Technology
The University of Tokyo
Kensuke Aishima

2012.4.11

Joint work with T. Matsuo, K. Murota, M. Sugihara, S. Tamura

Computation of singular values

Singular value decomposition

$$A = U \Sigma V^{\mathrm{T}}$$

$$\Sigma = \left(\begin{array}{ccc|c} \sigma_1 & & 0 & \\ & \ddots & & 0 \\ 0 & & \sigma_r & \\ \hline & 0 & & \end{array}\right)$$

Singular values

U, V: Orthogonal matrices

(Application: least square method, Image compression, etc.)

Step 1 $A \xrightarrow{\text{orthogonal transformation}}$ bidiagonal matrix B

Step 2 Computing singular values of B : $(\sigma_1, \ldots, \sigma_r)$

(e.g.) dqds algorithm is used

History of the method of computing singular values (relevant to the dqds)

- QR method (Golub – Kahan, 1965)
- QR method was improved (Demmel–Kahan, 1990)
- dqds algorithm (Fernando – Parlett, 1994)

  differential **q**uotient **d**ifference with **s**hifts
  - High speed, high accuracy
  - DLASQ routine in LAPACK (Parlett-Marques, 2000)

List of contents

- The dqds algorithm
- Our shift strategy for superquadratic convergence of the dqds algorithm
  - Algorithm, Convergence theorem, numerical experiment, Practical implementation
- Conclusions

List of contents

- The dqds algorithm
- Our shift strategy for superquadratic convergence of the dqds algorithm
  - Algorithm, Convergence theorem, numerical experiment, Practical implementation
- Conclusions

Notation and normalization

$$B = \begin{pmatrix} a_1 & b_1 & & 0 \\ & a_2 & \ddots & \\ & & \ddots & b_{m-1} \\ 0 & & & a_m \end{pmatrix} \quad m$$

m

Suppose that

$$a_k, b_k > 0$$

Singular values

$$\sigma_1 > \cdots > \sigma_m > 0$$

(There is no loss of generality in this assumption)

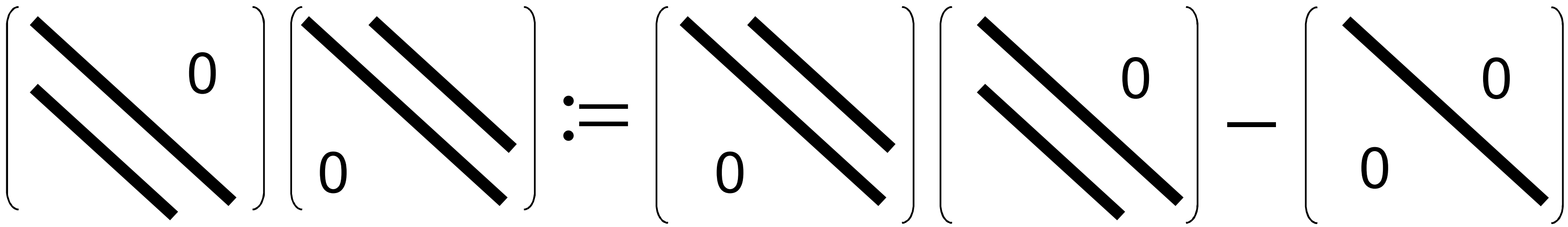
The dqds compute the singular values of B

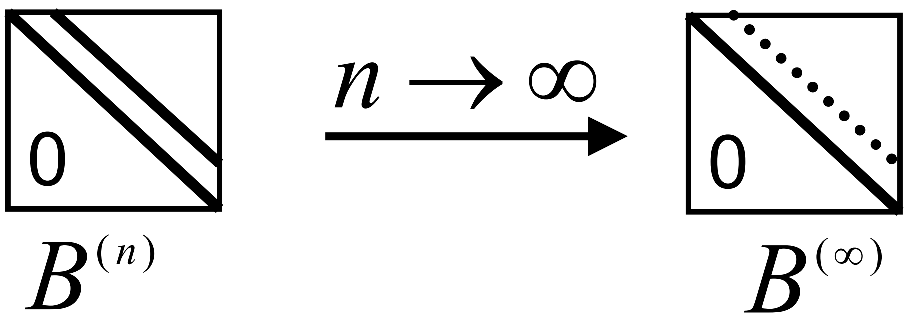
The dqds algorithm (Fernando – Parlett, 1994)

Initialization: $B^{(0)} := B$

Iterations: $(B^{(n+1)})^{\mathrm{T}} B^{(n+1)} := B^{(n)} (B^{(n)})^{\mathrm{T}} - s^{(n)} I$

shift (to accelerate convergence)





The diagonal elements are the singular values

The dqds algorithm (Fernando – Parlett, 1994)

Initialization:

$$q_k^{(0)} = (a_k)^2, e_k^{(0)} = (b_k)^2$$

Iterations:

for $n := 0, 1, \cdots$ do

choose $s^{(n)} \geq 0$

$d_1^{(n+1)} := q_1^{(n)} - s^{(n)}$

for $k := 1, \cdots, m-1$ do

$q_k^{(n+1)} := d_k^{(n+1)} + e_k^{(n)}$

$e_k^{(n+1)} := e_k^{(n)} q_{k+1}^{(n)} / q_k^{(n+1)}$

$d_{k+1}^{(n+1)} := d_k^{(n+1)} q_{k+1}^{(n)} / q_k^{(n+1)} - s^{(n)}$

end for

$q_m^{(n+1)} := d_m^{(n+1)}$

end for

dqds algorithm:

$$(B^{(n+1)})^{\mathrm{T}} B^{(n+1)} = B^{(n)} (B^{(n)})^{\mathrm{T}} - s^{(n)} I$$

shift (to accelerate convergence)

$$B^{(n)} = \begin{pmatrix} \sqrt{q_1^{(n)}} & \sqrt{e_1^{(n)}} & & \\ & \sqrt{q_2^{(n)}} & \ddots & \\ & & \ddots & \sqrt{e_{m-1}^{(n)}} \\ & & & \sqrt{q_m^{(n)}} \end{pmatrix}$$

Theorem (Convergence of the dqds)

(Aishima et. al. , 2008, SIMAX)

$\sigma_{\min}^{(n)}$:Smallest singular value of $B^{(n)}$

Shift $0 \le s^{(n)} < (\sigma_{\min}^{(n)})^2$

⇒

$$\sum_{n=0}^{\infty} s^{(n)} \le \sigma_m^2,$$

$$\lim_{n\to\infty} e_k^{(n)} = 0 \ ,$$

$$\lim_{n\to\infty} q_k^{(n)} = \sigma_k^2 - \sum_{n=0}^{\infty} s^{(n)}$$

$$B^{(n)} = \begin{pmatrix} \sqrt{q_1^{(n)}} & \sqrt{e_1^{(n)}} & & \\ & \sqrt{q_2^{(n)}} & \ddots & \\ & & \ddots & \sqrt{e_{m-1}^{(n)}} \\ & & & \sqrt{q_m^{(n)}} \end{pmatrix}$$

(Singular values $\sigma_1 > \cdots > \sigma_m$)

Theorem (Convergence of the dqds)

(Aishima et. al. , 2008, SIMAX)

$\sigma_{\min}^{(n)}$:Smallest singular value of $B^{(n)}$

Shift $0 \le s^{(n)} < (\sigma_{\min}^{(n)})^2$

$$\sum_{n=0}^{\infty} s^{(n)} \le \sigma_m^2,$$

$$\lim_{n\to\infty} e_k^{(n)} = 0 ,$$

$$\lim_{n\to\infty} q_k^{(n)} = \sigma_k^2 - \sum_{n=0}^{\infty} s^{(n)}$$

$$B^{(n)} = \begin{pmatrix} \sqrt{q_1^{(n)}} & \sqrt{e_1^{(n)}} & & \\ & \sqrt{q_2^{(n)}} & \ddots & \\ & & \ddots & \sqrt{e_{m-1}^{(n)}} \\ & & & \sqrt{q_m^{(n)}} \end{pmatrix}$$

(Singular values $\sigma_1 > \cdots > \sigma_m$)

Convergence of the lower right elements are fast

Deflation strategy

$$B^{(n)} = \begin{pmatrix} \sqrt{q_1^{(n)}} & \sqrt{e_1^{(n)}} & & & \\ & \sqrt{q_2^{(n)}} & \ddots & & \\ & & \ddots & \sqrt{e_{m-1}^{(n)}} & \\ & & & & \sqrt{q_m^{(n)}} \end{pmatrix}$$

Convergence of the lower right elements are fast

- $e_{m-1}^{(N)} \approx 0$ then terminate
- approximate $\sigma_m^2 \approx q_m^{(N)} + \sum_{n=0}^{N-1} s^{(n)}$
- repeat deflation
- $\sigma_m, \sigma_{m-1}, \ldots, \sigma_1$ can be computed

Convergence rate of the dqds

Convergence rate of the dqds is the convergence rate of $e_{m-1}^{(n)}$

$$B^{(n)} = \begin{pmatrix} \sqrt{q_1^{(n)}} & \sqrt{e_1^{(n)}} & & \\ & \sqrt{q_2^{(n)}} & \ddots & \\ & & \ddots & \sqrt{e_{m-1}^{(n)}} \\ & & & \sqrt{q_m^{(n)}} \end{pmatrix}$$

- $s^{(n)} \approx (\sigma_{\min}^{(n)})^2 \Rightarrow$ Accelerate the convergence !
- But the value of $\sigma_{\min}^{(n)}$ is unknown

A strict lower bound of $(\sigma_{\min}^{(n)})^2$ is used for the shift in the implementation of the dqds in LAPACK (DLASQ routine)

Theorem (superquadratic convergence)

(Aishima et. al. , 2010, JCAM)

The dqds in the LAPACK routine is executed

⇒

$$\lim_{n\to\infty} \frac{e_{m-1}^{(n+1)}}{(e_{m-1}^{(n)})^2} = 0 \qquad B^{(n)} = \begin{pmatrix} \sqrt{q_1^{(n)}} & \sqrt{e_1^{(n)}} & & \\ & \sqrt{q_2^{(n)}} & \ddots & \\ & & \ddots & \sqrt{e_{m-1}^{(n)}} \\ & & & \sqrt{q_m^{(n)}} \end{pmatrix}$$

Superquadratic convergence !

We propose another new superquadratic shift strategy

Direct approach for superquadratic convergence

- The shift strategy in the dqds in LAPACK (DLASQ routine)
  - Shifts are chosen such that $s^{(n)} \approx (\sigma_{\min}^{(n)})^2,\ s^{(n)} < (\sigma_{\min}^{(n)})^2$ based on the perturbation theory. As a result, superquadratic convergence is indirectly achieved

- Our proposed shift strategy
  - Superquadratic convergence: $\dfrac{e_{m-1}^{(n+1)}}{(e_{m-1}^{(n)})^2} \approx 0$ is directly and intuitively achieved !

List of contents

Motivation

Initialization:

$$q_k^{(0)} = (a_k)^2, e_k^{(0)} = (b_k)^2$$

Iterations:

for $n := 0, 1, \cdots$ do
  choose $s^{(n)} \geq 0$
  $d_1^{(n+1)} := q_1^{(n)} - s^{(n)}$
  for $k := 1, \cdots, m-1$ do
    $q_k^{(n+1)} := d_k^{(n+1)} + e_k^{(n)}$
    $e_k^{(n+1)} := e_k^{(n)} q_{k+1}^{(n)} / q_k^{(n+1)}$
    $d_{k+1}^{(n+1)} := d_k^{(n+1)} q_{k+1}^{(n)} / q_k^{(n+1)} - s^{(n)}$
  end for
  $q_m^{(n+1)} := d_m^{(n+1)}$
end for

From the dqds algorithm we see

$$\frac{e_{m-1}^{(n+2)}}{(e_{m-1}^{(n+1)})^2}$$

$$= \frac{q_m^{(n+1)}}{e_{m-1}^{(n+1)} q_{m-1}^{(n+2)}}$$

$$= \frac{1}{q_{m-1}^{(n+2)}} \left(\frac{q_m^{(n)} - s^{(n)}}{e_{m-1}^{(n+1)}} - 1 \right)$$

$$= \frac{q_{m-1}^{(n+1)}}{q_{m-1}^{(n+2)}} \left(\frac{q_m^{(n)} - s^{(n)}}{e_{m-1}^{(n+1)} q_m^{(n)}} - \frac{1}{q_{m-1}^{(n+1)}} \right)$$

$$= \frac{q_{m-1}^{(n+1)}}{q_{m-1}^{(n+2)}} \left(\frac{q_m^{(n)} - s^{(n)}}{e_{m-1}^{(n+1)} q_m^{(n)}} - \frac{1}{q_{m-1}^{(n)} - e_{m-2}^{(n+1)} + e_{m-1}^{(n)} - s^{(n)}} \right)$$

We want to make it 0 for superquadratic convergence

Motivation

$$\frac{e_{m-1}^{(n+2)}}{(e_{m-1}^{(n+1)})^2} = \frac{q_{m-1}^{(n+1)}}{q_{m-1}^{(n+2)}} \cdot \left(\frac{q_m^{(n)} - s^{(n)}}{e_{m-1}^{(n)} q_m^{(n)}} - \frac{1}{q_{m-1}^{(n)} - e_{m-2}^{(n+1)} + e_{m-1}^{(n)} - s^{(n)}} \right)$$

- We see $\frac{q_{m-1}^{(n+1)}}{q_{m-1}^{(n+2)}} \to 1$ from the global convergence theorem
- We hope that the shift $s^{(n)}$ is chosen such that

$$\frac{q_m^{(n)} - s^{(n)}}{e_{m-1}^{(n)} q_m^{(n)}} - \frac{1}{q_{m-1}^{(n)} - \boxed{e_{m-2}^{(n+1)}} + e_{m-1}^{(n)} - s^{(n)}} \approx 0$$

Unknown

- Hence, we set the shift $s^{(n)}$ that satisfies the following equation:

$$\boxed{\frac{q_m^{(n)} - s^{(n)}}{e_{m-1}^{(n)} q_m^{(n)}} - \frac{1}{q_{m-1}^{(n)} - \boxed{e_{m-2}^{(n)}} + e_{m-1}^{(n)} - s^{(n)}} = 0}$$

Replace

Motivation

$$\frac{e_{m-1}^{(n+2)}}{(e_{m-1}^{(n+1)})^2} = \frac{q_{m-1}^{(n+1)}}{q_{m-1}^{(n+2)}} \cdot \left(\frac{q_m^{(n)} - s^{(n)}}{e_{m-1}^{(n)} q_m^{(n)}} - \frac{1}{q_{m-1}^{(n)} - e_{m-2}^{(n+1)} + e_{m-1}^{(n)} - s^{(n)}} \right)$$

$$\frac{q_m^{(n)} - s^{(n)}}{e_{m-1}^{(n)} q_m^{(n)}} - \frac{1}{q_{m-1}^{(n)} - e_{m-2}^{(n)} + e_{m-1}^{(n)} - s^{(n)}} = 0$$

⇕

Quadratic equation: $(s^{(n)})^2 - X^{(n)} s^{(n)} + Y^{(n)} = 0$

where $X^{(n)} = q_{m-1}^{(n)} + q_m^{(n)} - e_{m-2}^{(n)} + e_{m-1}^{(n)}$, $Y^{(n)} = 4 q_m^{(n)} (q_{m-1}^{(n)} - e_{m-2}^{(n)})$

Solution: $s^{(n)} = \frac{1}{2}(X^{(n)} \pm \sqrt{(X^{(n)})^2 - Y^{(n)}})$

Superquadratic convergence is expected !

Proposed shift strategy

- Quadratic equation: $(s^{(n)})^2 - X^{(n)} s^{(n)} + Y^{(n)} = 0$

  $(X^{(n)} = q_{m-1}^{(n)} + q_m^{(n)} - e_{m-2}^{(n)} + e_{m-1}^{(n)},\ Y^{(n)} = 4 q_m^{(n)} (q_{m-1}^{(n)} - e_{m-2}^{(n)}))$

- One solution: $\lambda^{(n)} := \frac{1}{2}(X^{(n)} - \sqrt{(X^{(n)})^2 - Y^{(n)}})$
- Condition of convergence $0 \le s^{(n)} < (\sigma_{\min}^{(n)})^2$

Shift strategy (Q)

$$s^{(n)} = \begin{cases} \lambda^{(n)} & \text{If} \quad 0 < \lambda^{(n)} < (\sigma_{\min}^{(n)})^2 \\ \text{Other shifts} & \text{Otherwise} \end{cases}$$

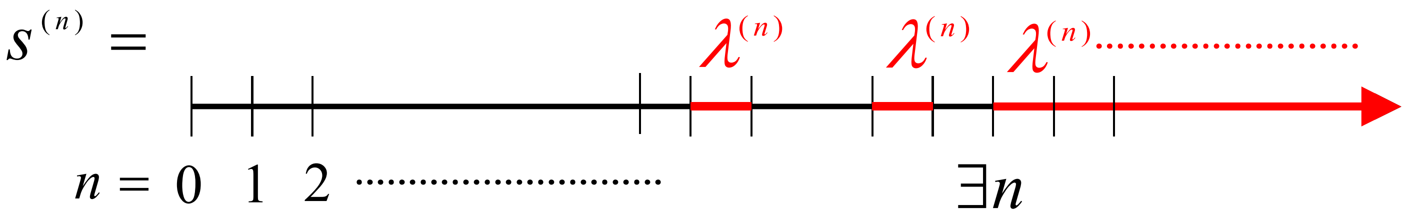
Theoretical analysis

Lemma (shifts in the final phase)

Shift strategy (Q) is used

$s^{(n)} = \lambda^{(n)}$ (superquadratic convergence shift)

for all sufficiently large n



Theoretical analysis

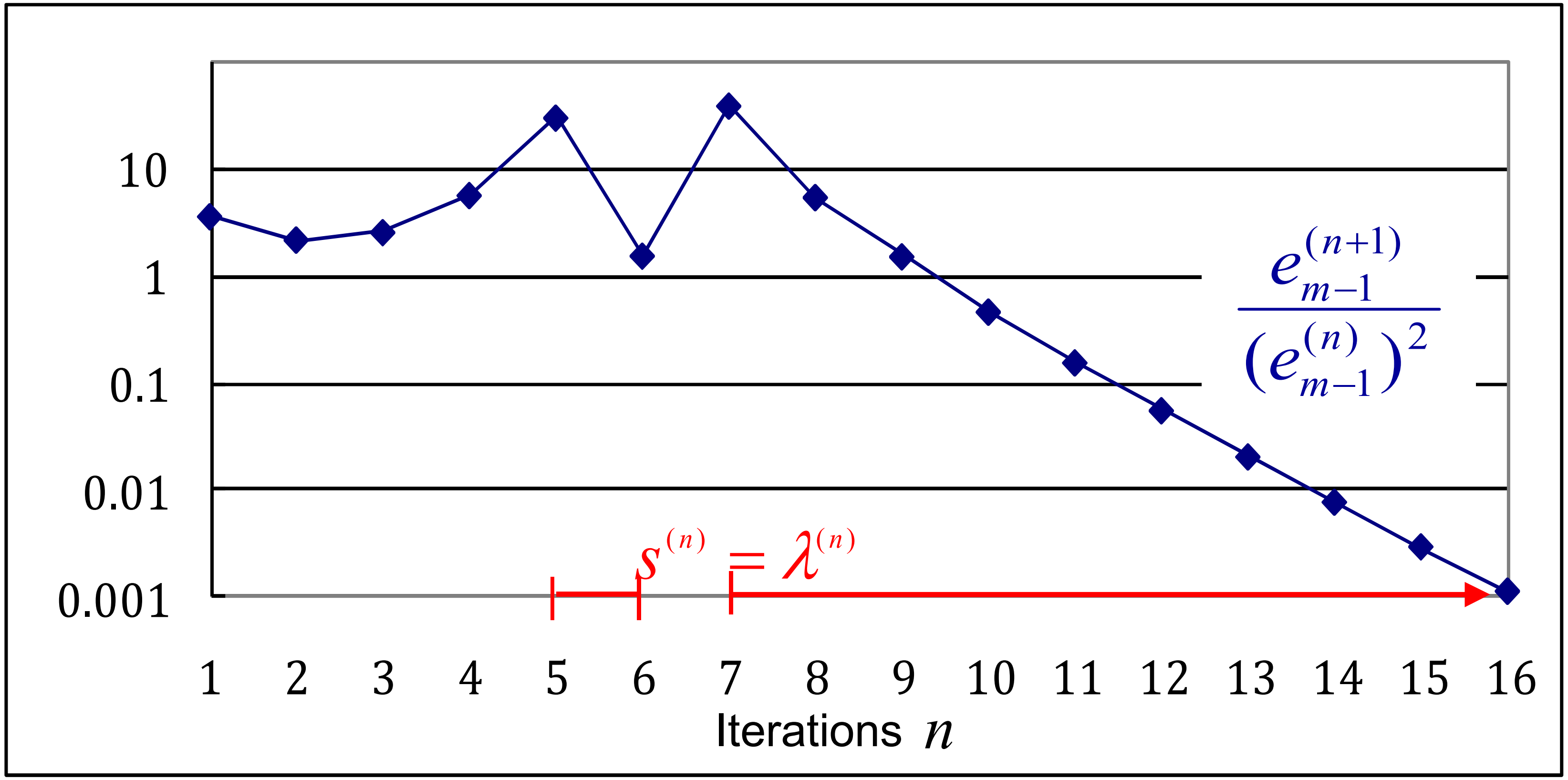
Theorem (superquadratic convergence)

Shift strategy (Q) is used

$$\Rightarrow \quad \lim_{n \to \infty} \frac{e_{m-1}^{(n+1)}}{\left(e_{m-1}^{(n)}\right)^2} = 0 \qquad B^{(n)} = \begin{pmatrix} \sqrt{q_1^{(n)}} & \sqrt{e_1^{(n)}} & & \\ & \sqrt{q_2^{(n)}} & \ddots & \\ & & \ddots & \sqrt{e_{m-1}^{(n)}} \\ & & & \sqrt{q_m^{(n)}} \end{pmatrix}$$

Superquadratic convergence !

A numerical experiment to illustrate the superquadratic convergence



$$\frac{e_{m-1}^{(n+1)}}{(e_{m-1}^{(n)})^2} \longrightarrow 0\text{: Superquadratic convergence}$$

The arbitrary-precision arithmetic library “exflib” , Fortran77

Practical implementation

Shift strategy (Q)

$$s^{(n)} = \begin{cases} \lambda^{(n)} & \text{If} \quad 0 < \lambda^{(n)} < (\sigma_{\min}^{(n)})^2 \\ \underline{\text{Other shifts}} & \text{Otherwise} \end{cases}$$

We employ the LAPACK shift strategy

- The ratio of cpu time to compute all singular values: (Our algorithm)/(LAPACK) is from 0.8 to 1.3 (almost same)

Test matrices are the same matrices in [Parlett-Marques, 2000, LAA]

List of contents

Conclusions

- A new shift strategy for superquadratic convergence was proposed
- Superquadratic convergence theorem was shown
- A numerical experiment to illustrate the superquadratic convergence was shown
- Practical implementation similar to LAPACK routine was given