L5b Singular Value Decomposition - Perturbation Theory

April 19, 2018

1 Singular Value Decomposition - Perturbation Theory

1.1 Prerequisites

The reader should be familiar with eigenvalue decomposition, singular value decompostion, and perturbation theory for eigenvalue decomposition.

1.2 Competences

The reader should be able to understand and check the facts about perturbations of singular values and vectors.

1.3 Peturbation bounds

For more details and the proofs of the Facts below, see Section **??**, and the references therein.

1.3.1 Definitions

Let $A \in \mathbb{C}^{m\times n}$ and let $A = U\Sigma V^*$ be its SVD.

The set of A's singular values is $sv(B) = \{\sigma_1, \sigma_2, \ldots)$, with $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$, and let $sv_{ext}(B) = sv(B)$ unless $m > n$ for which $sv_{ext}(B) = sv(B) \cup \{0, \ldots, 0\}$ (additional $|m-n|$ zeros).

Triplet $(u, \sigma, v) \in \times\mathbb{C}^m \times \mathbb{R} \times \mathbb{C}^n$ is a **singular triplet** of A if $\|u\|_2 = 1$, $\|v\|_2 = 1$, $\sigma \geq 0$, and $Av = \sigma u$ and $A^* u = \sigma v$.

$\tilde{A} = A + \Delta A$ is a **perturbed matrix**, where ΔA is **perturbation**. *The same notation is adopted to $\tilde{A}$, except all symbols are with tildes.*

Spectral condition number of A is $\kappa_2(A) = \sigma_{\max}(A)/\sigma_{\min}(A)$.

Let $X, Y \in \mathbb{C}^{n\times k}$ with $\operatorname{rank}(X) = \operatorname{rank}(Y) = k$. The **canonical angles** between their column spaces are $\theta_i = \cos^{-1}\sigma_i$, where σ_i are the singular values of $(Y^*Y)^{-1/2}Y^*X(X^*X)^{-1/2}$. The **canonical angle matrix** between X and Y is

$$\Theta(X, Y) = \operatorname{diag}(\theta_1, \theta_2, \ldots, \theta_k).$$

1.3.2 Facts

1. **Mirsky Theorem.** $\|\Sigma - \tilde{\Sigma}\|_2 \leq \|\Delta A\|_2$ and $\|\Sigma - \tilde{\Sigma}\|_F \leq \|\Delta A\|_F$.

2. **Residual bounds.** Let $\|\tilde{u}\|_2 = \|\tilde{v}\|_2 = 1$ and $\tilde{\mu} = \tilde{u}^* A \tilde{v}$. Let residuals $r = A\tilde{v} - \tilde{\mu}\tilde{u}$ and $s = A^*\tilde{u} - \tilde{\mu}\tilde{v}$, and let $\varepsilon = \max\{\|r\|_2, \|s\|_2\}$. Then $|\tilde{\mu} - \mu| \leq \varepsilon$ for some singular value μ of A.

3. The smallest error matrix ΔA for which $(\tilde{u}, \tilde{\mu}, \tilde{v})$ is a singular triplet of $\tilde{A}$ satisfies $\|\Delta A\|_2 = \varepsilon$.

4. Let μ be the closest singular value in $sv_{ext}(A)$ to $\tilde{\mu}$ and (u, μ, v) be the associated singular triplet, and let

$$\eta = \operatorname{gap}(\tilde{\mu}) = \min_{\mu \neq \sigma \in sv_{ext}(A)} |\tilde{\mu} - \sigma|.$$

If $\eta > 0$, then

$$|\tilde{\mu} - \mu| \leq \frac{\varepsilon^2}{\eta},$$

$$\sqrt{\sin^2 \theta(u, \tilde{u}) + \sin^2 \theta(v, \tilde{v})} \leq \frac{\sqrt{\|r\|_2^2 + \|s\|_2^2}}{\eta}.$$

5. Let

$$A = \begin{bmatrix} M & E \\ F & H \end{bmatrix}, \quad \tilde{A} = \begin{bmatrix} M & 0 \\ 0 & H \end{bmatrix},$$

where $M \in \mathbb{C}^{k \times k}$, and set $\eta = \min |\mu - \nu|$ over all $\mu \in sv(M)$ and $\nu \in sv_{ext}(H)$, and $\varepsilon = \max\{\|E\|_2, \|F\|_2\}$. Then

$$\max |\sigma_j - \tilde{\sigma}_j| \leq \frac{2\varepsilon^2}{\eta + \sqrt{\eta^2 + 4\varepsilon^2}}.$$

6. Let $m \geq n$ and let

$$\begin{bmatrix} U_1^* \\ U_2^* \end{bmatrix} A \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} A_1 & \\ & A_2 \end{bmatrix}, \quad \begin{bmatrix} \tilde{U}_1^* \\ \tilde{U}_2^* \end{bmatrix} \tilde{A} \begin{bmatrix} \tilde{V}_1 & \tilde{V}_2 \end{bmatrix} = \begin{bmatrix} \tilde{A}_1 & \\ & \tilde{A}_2 \end{bmatrix},$$

where $\begin{bmatrix} U_1 & U_2 \end{bmatrix}$, $\begin{bmatrix} V_1 & V_2 \end{bmatrix}$, $\begin{bmatrix} \tilde{U}_1 & \tilde{U}_2 \end{bmatrix}$, and $\begin{bmatrix} \tilde{V}_1 & \tilde{V}_2 \end{bmatrix}$ are unitary, and $U_1, \tilde{U}_1 \in \mathbb{C}^{m \times k}$, $V_1, \tilde{V}_1 \in \mathbb{C}^{n \times k}$. Set

$$R = A\tilde{V}_1 - \tilde{U}_1 \tilde{A}_1, \quad S = A^* \tilde{U}_1 - \tilde{V}_1 \tilde{A}_1.$$

Let $\eta = \min |\tilde{\mu} - \nu|$ over all $\tilde{\mu} \in sv(\tilde{A}_1)$ and $\nu \in sv_{ext}(A_2)$. If $\eta > 0$, then

$$\sqrt{\| \sin \Theta(U_1, \tilde{U}_1)\|_F^2 + \| \sin \Theta(V_1, \tilde{V}_1)\|_F^2} \leq \frac{\sqrt{\|R\|_F^2 + \|S\|_F^2}}{\eta}.$$

1.3.3 Example

In [1]:
```
m=8
n=5
s=srand(421)
k=min(m,n)
A=rand(-9:9,m,n)
```

Out[1]:
```
8×5 Array{Int64,2}:
 -8  -1  -2   2  -9
 -6   7   1   1   5
```

```
  3   2   9  -3   1
  7  -4  -8   0  -5
 -7   1  -4   3  -1
  0   0  -2   3  -8
  2   2  -2  -7  -6
 -7  -3  -8  -8  -5
```

In [2]:
```
A=rand(m,n)/100
B=A+A
```

Out[2]:
```
8Œ5 Array{Float64,2}:
 -7.99884     -0.993217    -1.99292   2.00887     -8.9908
 -5.99699      7.00114      1.00599   1.00344      5.00415
  3.00058      2.0068       9.00169  -2.99631      1.00099
  7.00973     -3.99895     -7.99876   0.00214899  -4.9979
 -6.99657      1.0013      -3.99663   3.00789     -0.994583
  0.00280957   0.00909776  -1.99083   3.00391     -7.99085
  2.00282      2.00033     -1.99167  -6.9938      -5.99554
 -6.99697     -2.99003     -7.99155  -7.99737     -4.99452
```

In [3]:
```
U,,V=svd(A)
U,,V=svd(B)
```

Out[3]:
```
([-0.447204 -0.338019  0.537286 -0.206075; 0.185802 -0.545543  -0.147543 0.633483;  ; -0
```

In [4]:
```
# Mirsky's Theorems
maximum(abs,-), norm(A), vecnorm(-), vecnorm(A)
```

Out[4]:
```
(0.02113859742292945, 0.03196314118428111, 0.022004821983059253, 0.035347859375243)
```

In [5]:
```
# Residual bounds - how close is (x,,y) to (U[:,j],[j],V[:,j])
j=rand(2:k-1)
x=round.(U[:,j],3)
y=round.(V[:,j],3)
x=x/norm(x)
y=y/norm(y)
=(x'*A*y)[]
, j,
```

Out[5]:
```
([21.167, 16.1921, 11.6027, 10.1944, 6.01866], 3, 11.602684964155396)
```

In [6]:
```
# Fact 2
r=A*y-*x
s=A'*x-*y
=max(norm(r),norm(s))
```

Out[6]:
```
0.012738790322717082
```

In [7]:
```
minimum(abs,-),
```

Out[7]: (5.33004160274686e-6, 0.012738790322717082)

In [8]: # Fact 4
```
=min(abs(-[j-1]),abs(-[j+1]))
```

Out[8]: 1.4082960857652687

In [9]:
```
-[j], ^2/
```

Out[9]: (-5.33004160274686e-6, 0.00011522916276371607)

In [10]:
```
# Eigenvector bound
# cos()
cosU=dot(x,U[:,j])
cosV=dot(y,V[:,j])
# Bound
sqrt(1-cosU^2+1-cosV^2), sqrt(norm(r)^2+norm(s)^2)/
```

Out[10]: (0.0008779073741076272, 0.011499971182415613)

In [11]:
```
# Fact 5 - we create small off-diagonal block perturbation
j=3
M=A[1:j,1:j]
H=A[j+1:m,j+1:n]
B=cat([1,2],M,H)
```

Out[11]:
```
8Œ5 Array{Int64,2}:
 -8  -1  -2   0   0
 -6   7   1   0   0
  3   2   9   0   0
  0   0   0   0  -5
  0   0   0   3  -1
  0   0   0   3  -8
  0   0   0  -7  -6
  0   0   0  -8  -5
```

In [12]:
```
E=rand(size(A[1:j,j+1:n]))/100
F=rand(size(A[j+1:m,1:j]))/100
C=map(Float64,B)
C[1:j,j+1:n]=E
C[j+1:m,1:j]=F
C
```

Out[12]:
```
8Œ5 Array{Float64,2}:
 -8.0          -1.0          -2.0          0.00782879   0.00361801
 -6.0           7.0           1.0          0.00546641   0.00490995
  3.0           2.0           9.0          0.00801844   0.00382256
  0.0073099     0.00245328    0.00521644   0.0         -5.0
  0.000391762   0.00951957    0.00368332   3.0         -1.0
  0.0096813     0.0070503     0.00130663   3.0         -8.0
  0.00175984    0.00405742    0.00229319  -7.0         -6.0
  0.000518073   0.00437298    0.00264824  -8.0         -5.0
```

```
In [13]: svdvals(M)

Out[13]: 3-element Array{Float64,1}:
          11.701
           9.71185
           4.21514

In [14]: svdvals(H)'

Out[14]: 1Œ2 RowVector{Float64,Array{Float64,1}}:
          14.0322  9.22487

In [15]: svdvals(M).-svdvals(H)'

Out[15]: 3Œ2 Array{Float64,2}:
          -2.3312    2.47609
          -4.32032   0.486977
          -9.81703  -5.00974

In [16]: =max(norm(E), norm(F))
         =svdvals(B)
         =svdvals(C)
         =minimum(abs,svdvals(M).-svdvals(H)')
         [ ], maximum(abs,-), 2*^2/(+sqrt(^2+4*^2))

Out[16]: ([14.0322 14.0322; 11.701 11.701;  ; 9.22487 9.22485; 4.21514 4.21514], 2.9571432486719
```

1.4 Relative perturbation theory

1.4.1 Definitions

Matrix $A \in \mathbb{C}^{m\times n}$ is **multiplicatively pertubed** to $\tilde{A}$ if $\tilde{A} = D_L^* A D_R$ for some $D_L \in \mathbb{C}^{m\times m}$ and $D_R \in \mathbb{C}^{n\times n}$.

Matrix A is (highly) **graded** if it can be scaled as $A = GS$ such that G is *well-behaved* (that is, $\kappa_2(G)$ is of modest magnitude), where the **scaling matrix** S is often diagonal. Interesting cases are when $\kappa_2(G) \ll \kappa_2(A)$.

Relative distances between two complex numbers α and $\tilde{\alpha}$ are:

$$\zeta(\alpha,\tilde{\alpha}) = \frac{|\alpha-\tilde{\alpha}|}{\sqrt{|\alpha\tilde{\alpha}|}}, \quad \text{for } \alpha\tilde{\alpha} \neq 0,$$

$$\varrho(\alpha,\tilde{\alpha}) = \frac{|\alpha-\tilde{\alpha}|}{\sqrt{|\alpha|^2+|\tilde{\alpha}|^2}}, \quad \text{for } |\alpha|+|\tilde{\alpha}| > 0.$$

1.4.2 Facts

1. If D_L and D_R are non-singular and $m \geq n$, then

$$\frac{\sigma_j}{\|D_L^{-1}\|_2\|D_R^{-1}\|_2} \leq \tilde{\sigma}_j \leq \sigma_j\|D_L\|_2\|D_R\|_2, \quad \text{for } i = 1,\ldots,n,$$

$$\|\operatorname{diag}(\zeta(\sigma_1,\tilde{\sigma}_1),\ldots,\zeta(\sigma_n,\tilde{\sigma}_n)\|_{2,F} \leq \frac{1}{2}\|D_L^* - D_L^{-1}\|_{2,F} + \frac{1}{2}\|D_R^* - D_R^{-1}\|_{2,F}.$$

2. Let $m \geq n$ and let

$$\begin{bmatrix} U_1^* \\ U_2^* \end{bmatrix} A \begin{bmatrix} V_1 & V_2 \end{bmatrix} = \begin{bmatrix} A_1 & \\ & A_2 \end{bmatrix}, \quad \begin{bmatrix} \tilde{U}_1^* \\ \tilde{U}_2^* \end{bmatrix} \tilde{A} \begin{bmatrix} \tilde{V}_1 & \tilde{V}_2 \end{bmatrix} = \begin{bmatrix} \tilde{A}_1 & \\ & \tilde{A}_2 \end{bmatrix},$$

where $\begin{bmatrix} U_1 & U_2 \end{bmatrix}$, $\begin{bmatrix} V_1 & V_2 \end{bmatrix}$, $\begin{bmatrix} \tilde{U}_1 & \tilde{U}_2 \end{bmatrix}$, and $\begin{bmatrix} \tilde{V}_1 & \tilde{V}_2 \end{bmatrix}$ are unitary, and $U_1, \tilde{U}_1 \in \mathbb{C}^{m\times k}$, $V_1, \tilde{V}_1 \in \mathbb{C}^{n\times k}$. Set

$$R = A\tilde{V}_1 - \tilde{U}_1\tilde{A}_1, \quad S = A^*\tilde{U}_1 - \tilde{V}_1\tilde{A}_1.$$

Let $\eta = \min \varrho(\mu, \tilde{\mu})$ over all $\mu \in sv(A_1)$ and $\tilde{\mu} \in sv_{ext}(A_2)$. If $\eta > 0$, then

$$\begin{aligned} &\sqrt{\|\sin\Theta(U_1, \tilde{U}_1)\|_F^2 + \|\sin\Theta(V_1, \tilde{V}_1)\|_F^2} \\ &\leq \frac{1}{\eta}(\|(I - D_L^*)U_1\|_F^2 + \|(I - D_L^{-1})U_1\|_F^2 \\ &\quad + \|(I - D_R^*)V_1\|_F^2 + \|(I - D_R^{-1})V_1\|_F^2)^{1/2}. \end{aligned}$$

3. Let $A = GS$ and $\tilde{A} = \tilde{G}S$, and let $\Delta G = \tilde{G} - G$. Then $\tilde{A} = DA$, where $D = I + (\Delta G)G^\dagger$, and Fact 1 applies with $D_L = D$, $D_R = I$, and

$$\|D^* - D^{-1}\|_{2,F} \leq \left(1 + \frac{1}{1 - \|(\Delta G)G^\dagger\|_2}\right) \frac{\|(\Delta G)G^\dagger\|_{2,F}}{2}.$$

According to the notebook on Jacobi Method and High Relative Accuracy, nearly optimal diagonal scaling is such that all columns of G have unit norms, $S = \operatorname{diag}\left(\|A_{:,1}\|_2, \ldots, \|A_{:,n}\|_2\right)$.

4. Let A be an real upper-bidiagonal matrix with diagonal entries $a_1, a_2, \ldots, a_n$ and the super-diagonal entries $b_1, b_2, \ldots, b_{n-1}$. Let the diagonal entries of $\tilde{A}$ be $\alpha_1 a_1, \alpha_2 a_2, \ldots, \alpha_n a_n$, and its super-diagonal entries be $\beta_1 b_1, \beta_2 b_2, \ldots, \beta_{n-1} b_{n-1}$. Then $\tilde{A} = D_L^* A D_R$ with

$$\begin{aligned} D_L &= \operatorname{diag}\left(\alpha_1, \frac{\alpha_1\alpha_2}{\beta_1}, \frac{\alpha_1\alpha_2\alpha_3}{\beta_1\beta_2}, \cdots\right), \\ D_R &= \operatorname{diag}\left(1, \frac{\beta_1}{\alpha_1}, \frac{\beta_1\beta_2}{\alpha_1\alpha_2}, \cdots\right). \end{aligned}$$

Let $\alpha = \prod_{j=1}^{n} \max\{\alpha_j, 1/\alpha_j\}$ and $\beta = \prod_{j=1}^{n-1} \max\{\beta_j, 1/\beta_j\}$. Then

$$(\alpha\beta)^{-1} \leq \|D_L^{-1}\|_2 \|D_R^{-1}\|_2 \leq \|D_L\|_2 \|D_R\|_2 \leq \alpha\beta,$$

and Fact 1 applies.

5. Consider the block partitioned matrices

$$\begin{aligned} A &= \begin{bmatrix} B & C \\ 0 & D \end{bmatrix}, \\ \tilde{A} &= \begin{bmatrix} B & 0 \\ 0 & D \end{bmatrix} = A \begin{bmatrix} I & -B^{-1}C \\ 0 & I \end{bmatrix} \equiv AD_R. \end{aligned}$$

By Fact 1, $\zeta(\sigma_j, \tilde{\sigma}_j) \leq \frac{1}{2}\|B^{-1}C\|_2$. This is used as a deflation criterion in the SVD algorithm for bidiagonal matrices.

1.4.3 Example - Bidiagonal matrix

In order to illustrate Facts 1 to 3, we need an algorithm which computes the singular values with high relative acuracy. Such algorithm, the one-sided Jacobi method, is discussed in the following notebook.

The algorithm actually used in the function `svdvals()` for `Bidiagonal` is the zero-shift bidiagonal QR algorithm, which attains the accuracy given by Fact 4: if all $1-\varepsilon \leq \alpha_i, \beta_j \leq 1+\varepsilon$, then

$$(1-\varepsilon)^{2n-1} \leq (\alpha\beta)^{-1} \leq \alpha\beta \leq (1-\varepsilon)^{2n-1}.$$

In other words, ε relative changes in diagonal and super-diagonal elements, cause at most $(2n-1)\varepsilon$ relative changes in the singular values.

However, if singular values and vectors are desired, the function `svd()` calls the standard algorithm, described in the next notebook, which **does not attain this accuracy** .

In [17]:
```
n=50
=100000
# The starting matrix
a=exp.(50*(rand(n)-0.5))
b=exp.(50*(rand(n-1)-0.5))
A=Bidiagonal(a,b, true)
# Multiplicative perturbation
DL=ones(n)+(rand(n)-0.5)/
DR=ones(n)+(rand(n)-0.5)/
# The perturbed matrix
=DL.*a.*DR
=DL[1:end-1].*b.*DR[2:end]
B=Bidiagonal(,,true)
(A.dv-B.dv)./A.dv
```

Out[17]:
```
50-element Array{Float64,1}:
  3.39812e-6
 -3.61322e-6
 -3.25847e-6
  1.31523e-6
  5.93668e-6
  7.9479e-6
  4.54888e-6
 -9.73189e-7
  1.0755e-6
  2.96194e-6
 -7.34928e-6
  2.99746e-7
 -1.35595e-6

 -4.0277e-6
  5.48266e-6
  4.66065e-6
  2.08409e-6
```

```
 4.37049e-6
 3.43698e-6
 7.09017e-6
 8.261e-6
 1.13622e-6
-2.21721e-6
-2.26233e-6
-5.49369e-6
```

In [18]:
```
(a-)./a, (b-)./b
```

Out[18]:
```
([3.39812e-6, -3.61322e-6, -3.25847e-6, 1.31523e-6, 5.93668e-6, 7.9479e-6, 4.54888e-6,
```

In [19]:
```
@which svdvals(A)
```

Out[19]:
```
svdvals(A::AbstractArray{#s268,2} where #s268<:Union{Complex{Float32}, Complex{Float64}
```

In [20]:
```
=svdvals(A)
=svdvals(B)
[ (-)./]
```

Out[20]:
```
50Œ2 Array{Float64,2}:
 3.71809e9     2.99746e-7
 1.1309e9      1.31523e-6
 7.36508e8    -5.84417e-6
 4.7042e8      6.50936e-6
 3.44534e8     9.3746e-6
 1.27054e8    -4.04487e-6
 9.8205e7     -5.2759e-7
 5.52383e7    -1.68669e-6
 5.3739e7      2.65906e-6
 1.85736e7     2.77895e-6
 1.34581e7    -2.26233e-6
 1.16407e7    -1.02195e-9
 3.3998e6     -5.11021e-6

 0.000135607  -1.17994e-6
 5.92811e-5   -1.69424e-6
 1.43836e-6    9.1791e-6
 1.02905e-6   -1.77238e-6
 2.56819e-8   -9.38028e-6
 2.22915e-8   -1.83292e-6
 3.12743e-9   -5.49369e-6
 4.11692e-10   5.96403e-7
 2.41415e-11  -1.4086e-6
 2.03616e-25   3.67804e-6
 1.77974e-27   8.8293e-7
 4.00726e-50  -1.18825e-7
```

In [21]:
```
cond(A)
```

```
Out[21]: 9.278376847680432e58

In [22]: # The standard algorithm
         U,,V=svd(A);

In [23]: (-)./

Out[23]: 50-element Array{Float64,1}:
             0.0
             2.10823e-16
             0.0
             0.0
            -1.73001e-16
             0.0
             4.55206e-16
             0.0
             0.0
             0.0
             2.76807e-16
             0.0
             1.36967e-16

             1.65358e-7
             2.28615e-16
            -5.08431e-10
            -2.0578e-16
           -13.4659
           -15.6661
          -117.791
          -901.316
           -17.9171
           -31.1363
          -136.616
            -5.09731e8

In [ ]:
```