Lecture 5 Singular value decomposition

Weinan E$^{1,2}$ and Tiejun Li$^2$

$^1$Department of Mathematics, Princeton University, weinan@princeton.edu

$^2$School of Mathematical Sciences, Peking University, tieli@pku.edu.cn

No.1 Science Building, 1575
Outline

Review and applications

QR for symmetric matrix

Numerical SVD
**Theorem (Singular value decomposition)**

Let \( A \in \mathbb{R}^{m \times n} \), then there exist \( U \in \mathbb{R}^{m \times m} \), \( V \in \mathbb{R}^{n \times n} \) and \( \Sigma \in \mathbb{R}^{m \times n} \) such that

\[
A = U \Sigma V
\]

where \( \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_r) \in \mathbb{R}^{m \times n} \). \( r \) is the rank of \( A \), \( \sigma_i > 0 \) are called singular values of \( A \), \( U^T U = I \), \( V^T V = I \) are orthogonal matrices.

It is straightforward that

\[
A^T A = V^T \Sigma^T \Sigma V = V^T \text{diag}(\sigma_1^2, \ldots, \sigma_r^2, 0, \ldots, 0) V
\]

i.e. the singular value \( \sigma_i = \sqrt{\lambda_i(A^T A)} \). Similarly we have \( \sigma_i = \sqrt{\lambda_i(A A^T)} \).
About singular values

- To find the orthogonal matrices $U$ and $V$ is equivalent to find the eigenvectors of matrices $A^T A$ and $AA^T$.

- If $A$ is symmetric, the singular value matrix $\Sigma = D$, where $D = \text{diag}(\lambda_1, \ldots, \lambda_r, 0, \ldots, 0)$. $\lambda_i$ is the eigenvalues of $A$, and $V = U^T$.

- The 2-norm of a matrix

\[ \|A\|_2 = \sqrt{\lambda_{\text{max}}(A^T A)} = \sigma_{\text{max}}. \]

- The 2-condition number

\[ \text{Cond}_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}}. \]
Generalized inverse of a matrix

- In general, if $A$ is singular, $A^{-1}$ doesn’t exist! If $A \in \mathbb{R}^{m \times n}$, there is no definition for $A^{-1}$.

- We define the Moore-Penrose generalized inverse of $A$ as

$$A^+ = V^T \text{diag}(\sigma_1^{-1}, \ldots, \sigma_r^{-1}, 0, \ldots, 0)U^T$$

for arbitrary matrix $A$!
Least square problems

- Least square problem 1: $Ax = b$ may have more than one solution. If it has more than one solution we wish to pick one with $\|x\|_2$ is the smallest, i.e., to find $x \in S = \{x | Ax = b\}$ such that
  $$\min_{x} \|x\|_2$$

- Least square problem 2: if it has no solution we wish to pick one which is the solution of the following minimization problem
  $$\min_{x} \|Ax - b\|_2$$

- In any case we have the following solution by generalized inverse
  $$x = A^+ b.$$
Multivariate linear regression

- Formulation
  Suppose we have a list of experimental data for a multi-variate function $Y = f(x_1, x_2, \ldots, x_m)$, after taking the zero-th and first order terms, we approximate $Y$ as
  
  $$Y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$$

  The problem is how to recover $\beta_i$ from the data?

- Naively consider the linear system
  
  $$Y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im}$$

  and $i = 1, \ldots, n$. It may have no solution or have infinite solutions. This is reduced to the least square problem for

  $$X\beta = Y$$
Multivariate linear regression

We have

\[
X = \begin{pmatrix}
1 & x_{11} & x_{12} & \cdots & x_{1m} \\
1 & x_{21} & x_{22} & \cdots & x_{2m} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & x_{n1} & x_{n2} & \cdots & x_{nm}
\end{pmatrix}, \quad \beta = \begin{pmatrix}
\beta_0 \\
\beta_1 \\
\vdots \\
\beta_m
\end{pmatrix}
\]

Least square solution

\[\beta = X^+ Y\]
Principal component analysis (PCA)

- **Object:** For a multi-component problem, is it possible to catch very few but very important characters to reduce the scale or dimension of the problem?

- **Answer:** Yes! PCA can do this job!
Principal component analysis (PCA)

- **PCA**

  Suppose we have experimental data to \( n \) characters of \( t \) units for a biological species, which can be proposed a matrix under experiments or investigations as

  \[
  Y = \begin{pmatrix}
  y_{11} & y_{12} & \cdots & y_{1n} \\
  y_{21} & y_{22} & \cdots & y_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  y_{t1} & y_{t2} & \cdots & y_{tn}
  \end{pmatrix}
  \]

  - **Object:** **Intuitively,** PCA is to find vectors
    \( a_i = (a_{1i}, a_{2i}, a_{ni}) \) \( (i = 1, \ldots, n) \) such that
    \[
    F_i = a_{1i}y_1 + a_{2i}y_2 + \cdots + a_{ni}y_n, \quad i = 1, \ldots, n
    \]

    are perpendicular each other, and pick up some large components among \( \|F_i\|_2 \). The analysis of \( a_i \) will give the main components of the problem.
Principal component analysis (PCA)

A geometrical interpretation of PCA for 2D coordinates analysis

▷ A mathematical rigorous interpretation (Projection maximization)

\[
\max_{\|a\|_2=1} \sum_{i=1}^{N} (x_i \cdot a)^2 = a^T X^T X a
\]

▷ Courant-Fisher’s theorem gives PCA.
Principal component analysis (PCA)

- Step 1: non-dimensionalization

Calculate the mean $\bar{y}_j = \frac{1}{t} \sum_{k=1}^{t} y_{kj}$, $j = 1, 2 \ldots, n$

Calculate variance $d_j = \sqrt{\sum_{k=1}^{t} (y_{kj} - \bar{y}_j)^2}$, $j = 1, 2 \ldots, n$

Transformation $x_{ij} = \frac{y_{ij} - \bar{y}_j}{d_j}$, $i = 1, 2 \ldots, t; j = 1, 2 \ldots, n$

Non-dimensionalization is used to eliminate the effect of choice of unit (单位).
Review and applications

Principal component analysis (PCA)

Step 2: Define principal component vector as

\[ F_i = a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{ni}x_n, \quad i = 1, \ldots, n \]

where \( x_i = (x_{1i}, x_{2i}, \ldots, x_{ti}) \). In order the vectors are independent each other, we need

\[ F_i^T F_j = 0, \quad i \neq j \]

i.e.

\[ F_i^T F_j = (a_{1i} \ a_{2i} \ \cdots \ a_{ni})X^T X \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix} = 0 \]
Principal component analysis (PCA)

Step 3: There exists orthogonal matrix $A$ such that

$$A^T X^T X A = \text{diag}(\lambda_1, \ldots, \lambda_n)$$

and $\lambda_k \geq 0$ ($k = 1, \ldots, n$). We have if $i \neq j$, the vectors $a_i, a_j$ in the $i$-th and $j$-th column will satisfy the independent condition, and

$$\|F_i\|_2 = \lambda_i$$

Step 4: Take the eigenvectors $a_i$ corresponding to the first $m$ biggest eigenvalues ($\lambda_1 > \lambda_2 > \cdots > \lambda_m > \cdots$), and make linear combination

$$F_i = a_{1i}x_1 + a_{2i}x_2 + \cdots + a_{ni}x_n, \quad i = 1, 2, \ldots, m$$

We will obtain the first $m$ principal component vectors.
PCA and SVD

If $X$ has SVD

\[ X = U \Sigma V \]

then we have $A = V^T$, and

\[ V X^T X V^T = \Sigma^T \Sigma \]

To find the first $m$ principal component vectors is equivalent to find the first $m$ principal (biggest) singular value and corresponding right singular vectors.
Outline

Review and applications

QR for symmetric matrix

Numerical SVD
Tri-diagonalization of symmetric matrix

- First transform symmetric $A$ into tri-diagonal matrix $T$

$$T = \begin{pmatrix}
\alpha_1 & \beta_1 & & & \\
\beta_1 & \alpha_2 & \ddots & \\
& \ddots & \ddots & \beta_{n-1} \\
& & \beta_{n-1} & \alpha_n
\end{pmatrix}$$

by a sequence of Householder transformations.

- The transformation procedure is the same as that for upper Hessenburg form with symmetry argument.
**Tri-diagonalization of symmetric matrix**

- The approach is to apply Householder transformation to $A$ column by column.

$$A = \begin{pmatrix}
  a_{11} & a_{12} & \cdots & a_{1n} \\
  a_{21} & a_{22} & \cdots & a_{2n} \\
  \vdots & \vdots & \ddots & \vdots \\
  a_{n1} & a_{n2} & \cdots & a_{nn}
\end{pmatrix}$$

- Suitably choose Householder matrix $H_1$ such that

$$H_1 \cdot \begin{pmatrix}
  a_{11} \\
  a_{21} \\
  a_{31} \\
  \vdots \\
  a_{n1}
\end{pmatrix} = \begin{pmatrix}
  a'_{11} \\
  a'_{21} \\
  0 \\
  \vdots \\
  0
\end{pmatrix}, \quad H_1 = \begin{pmatrix}
  1 & 0 \\
  0 & H_1'
\end{pmatrix}$$
Now we have

\[ A_1 = H_1 A H_1 = \begin{pmatrix} a'_{11} & a'_{12} & \cdots & 0 \\ a'_{21} & a'_{22} & \cdots & a'_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a'_{n2} & \cdots & a'_{nn} \end{pmatrix} \]

by symmetry of \( A \) and \( A_1 \).

The next step is the same for upper Hesseburg form. Finally we have tridiagonal form \( T \) and \( T \) has the same eigenvalues as \( A \).
Implicit shifted QR for symmetric tridiagonal matrix

Now we have symmetric tridiagonal $T$ with diagonal entries $\alpha_i (i = 1, \ldots, n)$ and off-diagonal entries $\beta_i (i = 1, \ldots, n - 1)$, one shifted QR step is

$$T - \mu I = QR$$

$$\hat{T} = RQ + \mu I$$

In fact

$$\hat{T} = Q^T T Q$$

If we can find $Q, \hat{T}$ directly, we doesn’t need the intermediate steps.

In fact

$$Q^T T Q = Q^T (QR + \mu I) Q = RQ + \mu I = \hat{T}.$$
Implicit shifted QR for symmetric tridiagonal matrix

- Find Givens matrix $G_1 = G(1, 2; \theta_1)$ such that

$$
\begin{pmatrix}
c & s \\
-s & c
\end{pmatrix}^T \begin{pmatrix}
\alpha_1 - \mu \\
\beta_1
\end{pmatrix} = \begin{pmatrix}
* \\
0
\end{pmatrix}
$$

- Define

$$T_1 = G_1^T T G_1.$$

We have

$$T_1 = \begin{pmatrix}
* & * & * \\
* & * & * \\
* & * & * \\
\vdots & \vdots & \ddots \\
* & * & * \\
* & * & *
\end{pmatrix}$$

- We should zero out the term *$. That only needs another Givens matrix $G_2$ multiplication.
**Implicit shifted QR for symmetric tridiagonal matrix**

- We can find Givens matrix $G_2 = G(2, 3; \theta_2)$ such that the term $*$ would be zero out.

- Define

$$T_2 = G_2^T G_1^T T G_1 G_2$$

We have

$$T_2 = \begin{pmatrix}
* & * & & & \\
* & * & * & & * \\
* & * & * & & \\
* & \cdots & \cdots & * \\
* & \cdots & \cdots & * \\
* & * & & & 
\end{pmatrix}$$

- We should zero out the term $*$ again. That needs a Givens matrix multiplication again.
Implicit shifted QR for symmetric tridiagonal matrix

- Sequentially we have

\[ T_{n-2} = \begin{pmatrix}
  * & * & & \\
  * & * & * & \\
  * & * & \ddots & * \\
  \vdots & \vdots & \ddots & * \\
  * & * & & *
\end{pmatrix} \]

- Finally we obtain

\[ \hat{T} = \begin{pmatrix}
  * & * & & \\
  * & * & * & \\
  * & * & \ddots & * \\
  \vdots & \vdots & \ddots & * \\
  * & * & & *
\end{pmatrix} \]
**Implicit shifted QR for symmetric tridiagonal matrix**

- Iterating for \( \hat{T} \) to obtain the next QR step!
- In general the shift is chosen as the famous Wilkinson’s shift: If the submatrix of \( T \)

\[
S = \begin{pmatrix}
\alpha_{n-1} & \beta_{n-1} \\
\beta_{n-1} & \alpha_n
\end{pmatrix}
\]

then choose \( \mu \) one of the eigenvalues of \( S \) which is more closer to \( \alpha_n \).

\[
\mu = \alpha_n + \delta - \text{sign}(\delta) \sqrt{\delta^2 + \beta_{n-1}^2}
\]

and \( \delta = \frac{\alpha_n + \alpha_{n-1}}{2} \).
- The convergence will be very fast with this shift.
Outline

Review and applications

QR for symmetric matrix

Numerical SVD
Implicit QR method for singular value computation

- First transform $A$ into upper bidiagonal matrix $B$

$$B = \begin{pmatrix}
    d_1 & f_2 \\
    & d_2 & \ddots \\
    & & \ddots & f_n \\
    & & & d_n
\end{pmatrix}$$

by a sequence of Householder transformations

$A \xrightarrow{U_1} \text{eliminate the first column} \xrightarrow{V_1} \text{eliminate the first row} \ldots$

$\xrightarrow{U_n} \text{eliminate the n-th column} = \begin{pmatrix}
    B \\
    0
\end{pmatrix}$

- $A$ has the same singular values as $B$. 
Implicit QR method for singular value computation

- First transform $A$ into upper bidiagonal matrix $B$

$$B = \begin{pmatrix}
  d_1 & f_2 \\
  & d_2 & \ddots \\
  & & \ddots & f_n \\
  & & & d_n
\end{pmatrix}$$

by a sequence of Householder transformations

$$A \xrightarrow{U_1} \text{eliminate the first column} \xrightarrow{V_1} \text{eliminate the first row} \rightarrow \cdots \rightarrow$$

$$U_n \xrightarrow{} \text{eliminate the n-th column} = \begin{pmatrix} B \\ 0 \end{pmatrix}$$

- Now we have

$$U_n \cdots U_1 A V_1 \cdots V_{n-1} = \begin{pmatrix} B \\ 0 \end{pmatrix}$$
Implicit shifted QR method for singular value computation

- Basic idea: Implicitly apply shifted QR method to symmetric tridiagonal matrix $B^T B$ but without forming it.

- Steps:
  - Determine the shift $\mu$. This is equivalent to the shift step for $B^T B$.
    Wilkinson shift: set $\mu$ is the eigenvalue of
    \[
    \begin{pmatrix}
    d_{n-1}^2 + f_{n-1}^2 & d_{n-1}f_n \\
    d_{n-1}f_n & d_n^2 + f_n^2
    \end{pmatrix}
    \]
    closer to $d_n^2 + f_n^2$ to make the convergence faster.
  - Find Givens matrix $G_1 = G(1, 2; \theta)$ such that
    \[
    \begin{pmatrix} c & s \\ -s & c \end{pmatrix}^T \begin{pmatrix} d_1^2 - \mu \\ d_1f_2 \end{pmatrix} = \begin{pmatrix} * \\ 0 \end{pmatrix}
    \]
    and compute $BG_1$.
    This is equivalent to apply $G_1$ step for $B^T B$. 
Implicit shifted QR method for singular value computation

- We have

\[
BG_1 = \begin{pmatrix}
  * & * \\
  * & * & * \\
  \vdots & \ddots & * \\
  \vdots & \ddots & * \\
  * & & & \\
\end{pmatrix}
\]

so we should zero out the term *. We want to find \( P_2 \) and \( G_2 \) such that \( P_2(BG_1)G_2 \) is bidiagonal and \( G_2e_1 = e_1 \).

This is equivalent to apply \( G_2 \) step for \( G_1^TB^TBG_1 \).
Implicit shifted QR method for singular value computation

- It is not difficult to find $P_2$ and $G_2$ by Givens transformation and we have

$$P_2BG_1G_2 = \begin{pmatrix}
    * & * \\
    * & * \\
    * & * \\
    \vdots & \vdots & \ddots \\
    \end{pmatrix}$$

so we should zero out the term *. We want to find $P_3$ and $G_3$ such that $P_3P_2BG_1G_2G_3$ is bidiagonal and $G_3e_i = e_i$, $i = 1, 2$.

These steps should be repeated until $BG_1$ becomes bidiagonal! It is equivalent to find $G_i$ steps for symmetric tridiagonal matrix.
Implicit shifted QR method for singular value computation

Finally we have

\[ P_{n-1} \cdots P_2BG_1 \cdots G_{n-1} = \begin{pmatrix} * & * \\ & * & * \\ & & * & \ddots \\ & & & & * \\ & & & & & * \end{pmatrix} \]

Iterate until the off-diagonal entries converge to 0, and the diagonal entries converge to singular values!