AMS526: Numerical Analysis I  
(Numerical Linear Algebra)  
Lecture 14: Solution of Least Squares Problems; 
Singular-Value Decomposition

Xiangmin Jiao

Stony Brook University
Outline

1. Solution of Least Squares Problems

2. Singular-Value Decomposition
Solution of Least Squares Problems

- More robust approach is to use QR factorization \( A = \hat{Q}\hat{R} \)
  - \( b \) can be projected onto \( \text{range}(A) \) by \( P = \hat{Q}\hat{Q}^T \), and therefore \( \hat{Q}\hat{R}x = \hat{Q}\hat{Q}^Tb \)
  - Left-multiply by \( \hat{Q}^T \) and we get \( \hat{R}x = \hat{Q}^Tb \) (note \( A^+ = \hat{R}^{-1}\hat{Q}^T \))

<table>
<thead>
<tr>
<th>Least squares via QR Factorization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute reduced QR factorization ( A = \hat{Q}\hat{R} )</td>
</tr>
<tr>
<td>Compute vector ( c = \hat{Q}^Tb )</td>
</tr>
<tr>
<td>Solve upper-triangular system ( \hat{R}x = c ) for ( x )</td>
</tr>
</tbody>
</table>

- Computation is dominated by QR factorization \( (2mn^2 - \frac{2}{3}n^3) \)
- Question: If Householder QR is used, how to compute \( \hat{Q}^Tb \)?
Solution of Least Squares Problems

- More robust approach is to use QR factorization $A = \hat{Q}\hat{R}$
  - $b$ can be projected onto range($A$) by $P = \hat{Q}\hat{Q}^T$, and therefore $\hat{Q}\hat{R}x = \hat{Q}\hat{Q}^Tb$
  - Left-multiply by $\hat{Q}^T$ and we get $\hat{R}x = \hat{Q}^Tb$ (note $A^+ = \hat{R}^{-1}\hat{Q}^T$)

<table>
<thead>
<tr>
<th>Least squares via QR Factorization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compute reduced QR factorization $A = \hat{Q}\hat{R}$</td>
</tr>
<tr>
<td>Compute vector $c = \hat{Q}^Tb$</td>
</tr>
<tr>
<td>Solve upper-triangular system $\hat{R}x = c$ for $x$</td>
</tr>
</tbody>
</table>

- Computation is dominated by QR factorization $(2mn^2 - \frac{2}{3}n^3)$
- Question: If Householder QR is used, how to compute $\hat{Q}^Tb$?
- Answer: Compute $Q^Tb$ (where $Q$ is from full QR factorization) and then take first $n$ entries of resulting $Q^Tb$
Backward Stability of Householder QR

- For a QR factorization $A = QR$ computed by Householder triangularization, the factors $\tilde{Q}$ and $\tilde{R}$ satisfy

\[ \tilde{Q}\tilde{R} = A + \delta A, \quad \|\delta A\|/\|A\| = O(\epsilon_{\text{machine}}), \]

i.e., exact QR factorization of a slightly perturbed $A$

- $\tilde{R}$ is $R$ computed by algorithm using floating points

- However, $\tilde{Q}$ is product of exactly orthogonal reflectors

\[ \tilde{Q} = \tilde{Q}_1 \tilde{Q}_2 \ldots \tilde{Q}_n \]

where $\tilde{Q}_k$ is given by computed $\tilde{v}_k$, since $Q$ is not formed explicitly
Backward Stability of Solving $Ax = b$ with QR

Algorithm: Solving $Ax = b$ by QR Factorization

1. Compute $A = QR$ using Householder, represent $Q$ by reflectors
2. Compute vector $y = Q^T b$ implicitly using reflectors
3. Solve upper-triangular system $Rx = y$ for $x$

- All three steps are backward stable
- Overall, we can show that

$$(A + \Delta A)\tilde{x} = b, \quad \|\Delta A\|/\|A\| = O(\epsilon_{\text{machine}})$$

as we prove next
Backward Stability of Solving $Ax = b$ with Householder QR

Proof: Step 2 gives

\[(\tilde{Q} + \delta Q)\tilde{y} = b, \quad \|\delta Q\| = O(\epsilon_{\text{machine}})\]

Step 3 gives

\[(\tilde{R} + \delta R)\tilde{x} = \tilde{y}, \quad \|\delta R\|/\|\tilde{R}\| = O(\epsilon_{\text{machine}})\]

Therefore,

\[b = (\tilde{Q} + \delta Q)(\tilde{R} + \delta R)\tilde{x} = \left[\tilde{Q}\tilde{R} + (\delta Q)\tilde{R} + \tilde{Q}(\delta R) + (\delta Q)(\delta R)\right]\tilde{x}\]

Step 1 gives

\[b = \left[ A + \delta A + (\delta Q)\tilde{R} + \tilde{Q}(\delta R) + (\delta Q)(\delta R) \right] \tilde{x} \]

where $\tilde{Q}\tilde{R} = A + \delta A$
Proof of Backward Stability Cont’d

\( \tilde{Q}\tilde{R} = A + \delta A \) where \( \|\delta A\|/\|A\| = O(\epsilon_{\text{machine}}) \), and therefore

\[
\frac{\|\tilde{R}\|}{\|A\|} \leq \|\tilde{Q}^T\| \frac{\|A + \delta A\|}{\|A\|} = O(1)
\]

Now show that each term in \( \Delta A \) is small

\[
\frac{\|(\delta Q)\tilde{R}\|}{\|A\|} \leq \|(\delta Q)\| \frac{\|\tilde{R}\|}{\|A\|} = O(\epsilon_{\text{machine}})
\]

\[
\frac{\|\tilde{Q}(\delta R)\|}{\|A\|} \leq \|\tilde{Q}\| \frac{\|\delta R\| \|\tilde{R}\|}{\|A\|} = O(\epsilon_{\text{machine}})
\]

\[
\frac{\|(\delta Q)(\delta R)\|}{\|A\|} \leq \|\delta Q\| \frac{\|\delta R\|}{\|A\|} = O(\epsilon^2_{\text{machine}})
\]

Overall,

\[
\frac{\|\Delta A\|}{\|A\|} \leq \frac{\|\delta A\|}{\|A\|} + \frac{\|(\delta Q)\tilde{R}\|}{\|A\|} + \frac{\|\tilde{Q}(\delta R)\|}{\|A\|} + \frac{\|(\delta Q)(\delta R)\|}{\|A\|} = O(\epsilon_{\text{machine}})
\]

Since the algorithm is backward stable, it is also accurate.
Stability of Gram-Schmidt Orthogonalization

- Gram-Schmidt QR is unstable, due to loss of orthogonality
- Gram-Schmidt can be stabilized using augmented system of equations
  1. Compute QR factorization of augmented matrix: \([Q,R1]=\text{mgs}([A,b])\)
  2. Extract \(R\) and \(Q^Tb\) from \(R1\): \(R=R1(1:n,1:n); Qb=R1(1:n,n+1)\)
  3. Back solve: \(x=R\backslash Qb\)

Theorem

The solution of the full-rank least squares problem by Gram-Schmidt orthogonality is backward stable in the sense that the computed solution \(\tilde{x}\) has the property

\[
\|(A + \delta A)\tilde{x} - b\| = \min, \quad \frac{\|\delta A\|}{\|A\|} = O(\epsilon_{\text{machine}})
\]

for some \(\delta A \in \mathbb{R}^{m \times n}\), provided that \(Q^Tb\) is formed implicitly.
Other Methods

- The method of *normal equation* solves \( x = (A^T A)^{-1} A^T b \), due to squaring of condition number of \( A \)

### Theorem

*The solution of the full-rank least squares problem via normal equation is unstable. Stability can be achieved, however, by restriction to a class of problems in which \( \kappa(A) \) is uniformly bounded above.*

- Another method is to SVD (coming up next)
Summary of Algorithms for Least Squares

- Householder QR (with/without pivoting, explicit or implicit $Q$): **Backward stable**
- Classical Gram-Schmidt: **Unstable**
- Modified Gram-Schmidt with explicit $Q$: **Unstable**
- Modified Gram-Schmidt with augmented system of equations with implicit $Q$: **Backward stable**
- Normal equations (solve $A^T A x = A^T b$): **Very unstable**
- Singular value decomposition: **Backward stable**
Outline

1. Solution of Least Squares Problems

2. Singular-Value Decomposition
Geometric Observation

- The image of unit sphere under any $m \times n$ matrix is a hyperellipsoid.
- Give a unit sphere $S$ in $\mathbb{R}^n$, let $AS$ denote the shape after transformation.
- SVD is

$$A = U\Sigma V^T$$

where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ is orthogonal and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal.

- *Singular values* are diagonal entries of $\Sigma$, correspond to the principal semiaxes, with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$.

- *Left singular vectors* of $A$ are column vectors of $U$ and are oriented in the directions of the principal semiaxes of $AS$.

- *Right singular vectors* of $A$ are column vectors of $V$ and are the preimages of the principal semiaxes of $AS$.

- $Av_j = \sigma_j u_j$ for $1 \leq j \leq n$. 


Two Different Types of SVD

- **Full SVD**: $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ is
  \[ A = U \Sigma V^T \]
Two Different Types of SVD

- **Full SVD:** $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ is

  $$A = U \Sigma V^T$$

- **Reduced SVD:** $\hat{U} \in \mathbb{R}^{m \times n}$, $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ (assume $m \geq n$)

  $$A = \hat{U} \hat{\Sigma} V^T$$
Two Different Types of SVD

- **Full SVD**: $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ is
  \[ A = U\Sigma V^T \]

- **Reduced SVD**: $\hat{U} \in \mathbb{R}^{m \times n}$, $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ (assume $m \geq n$)
  \[ A = \hat{U}\hat{\Sigma}V^T \]

- Furthermore, notice that
  \[ A = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^T \]
  so we can keep only entries of $U$ and $V$ corresponding to nonzero $\sigma_i$. 

Xiangmin Jiao
Numerical Analysis I
SVD vs. Eigenvalue Decomposition

- Eigenvalue decomposition of nondefective matrix $A$ is $A = X\Lambda X^{-1}$

- Differences between SVD and Eigenvalue Decomposition
  - Not every matrix has eigenvalue decomposition, but every matrix has singular value decomposition
  - Eigenvalues may not always be real numbers, but singular values are always non-negative real numbers
  - Eigenvectors are not always orthogonal to each other (orthogonal for symmetric matrices), but left (or right) singular vectors are orthogonal to each other

- Similarities
  - Singular values of $A$ are square roots of eigenvalues of $AA^T$ and $A^TA$,
    and their eigenvectors are left and right singular vectors, respectively
  - Singular values of symmetric matrices are absolute values of eigenvalues, and eigenvectors are singular vectors
  - This relationship can be used to compute singular values by hand
SVD vs. Eigenvalue Decomposition

- Eigenvalue decomposition of nondefective matrix $A$ is $A = X\Lambda X^{-1}$

- Differences between SVD and Eigenvalue Decomposition
  - Not every matrix has eigenvalue decomposition, but every matrix has singular value decomposition
  - Eigenvalues may not always be real numbers, but singular values are always non-negative real numbers
  - Eigenvectors are not always orthogonal to each other (orthogonal for symmetric matrices), but left (or right) singular vectors are orthogonal to each other
SVD vs. Eigenvalue Decomposition

- Eigenvalue decomposition of nondefective matrix $A$ is $A = X\Lambda X^{-1}$

- Differences between SVD and Eigenvalue Decomposition
  - Not every matrix has eigenvalue decomposition, but every matrix has singular value decomposition
  - Eigenvalues may not always be real numbers, but singular values are always non-negative real numbers
  - Eigenvectors are not always orthogonal to each other (orthogonal for symmetric matrices), but left (or right) singular vectors are orthogonal to each other

- Similarities
  - Singular values of $A$ are square roots of eigenvalues of $AA^T$ and $A^TA$, and their eigenvectors are left and right singular vectors, respectively
  - Singular values of symmetric matrices are absolute values of eigenvalues, and eigenvectors are singular vectors
  - This relationship can be used to compute singular values by hand
Existence of SVD

**Theorem**

*(Existence) Every matrix* $A \in \mathbb{R}^{m \times n}$ *has an SVD.*

**Proof:** Let $\sigma_1 = \|A\|_2$. There exists $v_1 \in \mathbb{R}^n$ with $\|v_1\|_2 = 1$ and $\|Av_1\|_2 = \sigma_1$. Let $U_1$ and $V_1$ be orthogonal matrices whose first columns are $u_1 = Av_1/\sigma_1$ (or any unit-length vector if $\sigma_1 = 0$) and $v_1$, respectively. Note that

$$U_1^T AV_1 = S = \begin{bmatrix} \sigma_1 & \omega^T \\ 0 & B \end{bmatrix}. \quad (1)$$

Furthermore, $\omega = 0$ because $\|S\|_2 = \sigma_1$, and

$$\|\begin{bmatrix} \sigma_1 & \omega^T \\ 0 & B \end{bmatrix}\begin{bmatrix} \sigma_1 \\ \omega \end{bmatrix}\|_2 \geq \sigma_1^2 + \omega^T \omega = \sqrt{\sigma_1^2 + \omega^T \omega} \|\begin{bmatrix} \sigma_1 \\ \omega \end{bmatrix}\|_2,$$

implying that $\sigma_1 \geq \sqrt{\sigma_1^2 + \omega^T \omega}$ and $\omega = 0$. 

Xiangmin Jiao

Numerical Analysis I
Existence of SVD Cont’d

We then prove by induction using (1). If \( m = 1 \) or \( n = 1 \), then \( B \) is empty and we have \( A = U_1 S V_1^T \). Otherwise, suppose \( B = U_2 \Sigma_2 V_2^T \), and then

\[
A = U_1 \begin{bmatrix} 1 & 0^T \\ 0 & U_2 \end{bmatrix} \begin{bmatrix} \sigma_1 & 0^T \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} 1 & 0^T \\ 0 & V_2^T \end{bmatrix} V_1^T,
\]

where \( U \) and \( V \) are orthogonal.