AMS526: Numerical Analysis I (Numerical Linear Algebra for Computational and Data Sciences)

Lecture 12: Givens Rotation; Least Squares Problems; Conditioning of Least Squares Problems

Xiangmin Jiao

Stony Brook University

Outline

Givens Rotations

- Instead of using reflection, we can rotate x to obtain $\|x\| e_1$
- A Given rotation $R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ rotates $x \in \mathbb{R}^2$ counterclockwise by θ
- Choose θ to be angle between $(x_i, x_j)^T$ and $(1, 0)^T$, and we have

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_i \\ x_j \end{bmatrix} = \begin{bmatrix} \sqrt{x_i^2 + x_j^2} \\ 0 \end{bmatrix}$$

where

$$\cos\theta = \frac{x_i}{\sqrt{x_i^2 + x_j^2}}, \ \sin\theta = \frac{-x_j}{\sqrt{x_i^2 + x_j^2}}$$

Givens QR

- Introduce zeros in column bottom-up, one zero at a time

$$\begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \end{bmatrix} \overset{(4,5)}{\longrightarrow} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \times & \times & \times \\ \mathsf{x} & \mathsf{x} & \mathsf{x} \\ 0 & \mathsf{x} & \mathsf{x} \end{bmatrix} \overset{(3,4)}{\longrightarrow} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ \mathsf{x} & \mathsf{x} & \mathsf{x} \\ 0 & \mathsf{x} & \mathsf{x} \\ & \times & \times \end{bmatrix} \dots$$

- To zero a_{ij}, left-multiply matrix F with $F_{i:i+1,i:i+1}$ being rotation matrix and $F_{kk} = 1$ for $k \neq i, i+1$
- Flop count of Givens QR is $3mn^2 - n^3$, which is about 50% more expensive than Householder triangularization

Adding a Row

- Suppose $A \in \mathbb{R}^{m \times n}$ with $m \geq n$, and A has full rank
- Let $\tilde{A} = \begin{bmatrix} A_1 \\ z^T \\ A_2 \end{bmatrix}$, where $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ and z^T is a new row inserted
- Obtain $\tilde{A} = \tilde{Q}\tilde{R}$ from $A = QR$ efficiently using Givens rotation:
  - Suppose $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix} = \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} R$.
  - Then $\tilde{A} = \begin{bmatrix} A_1 \\ z^T \\ A_2 \end{bmatrix} = \begin{bmatrix} 0 & Q_1 \\ 1 & 0^T \\ 0 & Q_2 \end{bmatrix} \begin{bmatrix} z^T \\ R \end{bmatrix}$
  - Perform series of Givens rotation $\tilde{R} = U_n^T \dots U_2^T U_1^T \begin{bmatrix} z^T \\ R \end{bmatrix}$, and

    then $\tilde{Q} = \begin{bmatrix} 0 & Q_1 \\ 1 & 0^T \\ 0 & Q_2 \end{bmatrix} U_1 U_2 \dots U_n$
  - Updating $\tilde{R}$ costs $3n^2$ flops, and updating $\tilde{Q}$ costs $6mn$ flops

Adding a Column

- Suppose $A \in \mathbb{R}^{m\times n}$ with $m \geq n$, and A has full rank
- Let $\tilde{A} = \begin{bmatrix} A_1 & z & A_2 \end{bmatrix}$, where $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix}$ and z is new column
- Obtain $\tilde{A} = \tilde{Q}\tilde{R}$ from $A = QR$ efficiently using Givens rotation:
  - Suppose $A = \begin{bmatrix} A_1 & A_2 \end{bmatrix} = Q \begin{bmatrix} R_1 & R_2 \end{bmatrix}$
  - Then $\tilde{A} = \begin{bmatrix} A_1 & z & A_2 \end{bmatrix} = Q \begin{bmatrix} R_1 & w & R_2 \end{bmatrix}$, where $w = Q^T z$
  - Perform series of Givens rotation $\tilde{R} = U_{k+1} \cdots U_n \begin{bmatrix} R_1 & w & R_2 \end{bmatrix}$, where U_n performs on rows n and $n-1$, U_{n-1} performs on rows $n-1$ and $n-2$, etc.
  - $\tilde{Q} = QU_n^T \cdots U_{k+1}^T$
  - It takes $O(mn)$ time overall

Outline

Linear Least Squares Problems

- Overdetermined system of equations $Ax \approx b$, where A has more rows than columns and has full rank, in general has no solutions
- Example application: Polynomial least squares fitting
- In general, minimize the residual $r = b - Ax$
- In terms of 2-norm, we obtain linear least squares problem: Given $A \in \mathbb{R}^{m \times n}$, $m \geq n$, and $b \in \mathbb{R}^m$, find $x \in \mathbb{R}^n$ such that $\|b - Ax\|_2$ is minimized
- If A has full rank, the minimizer x is the solution to the normal equation

  $$A^T A x = A^T b$$

  or in terms of the *pseudoinverse* A^+,

  $$x = A^+ b, \quad \text{where } A^+ = (A^T A)^{-1} A^T \in \mathbb{R}^{n \times m}$$

Geometric Interpretation

- Ax is in range(A), and the point in range(A) closest to b is its orthogonal projection onto range(A)
- Residual r is then orthogonal to range(A), and hence $A^T r = A^T(b - Ax) = 0$
- Ax is orthogonal projection of b, where $x = A^+ b$, so $P = AA^+ = A(A^T A)^{-1} A^T$ is orthogonal projection

Solution of Lease Squares Problems

- One approach is to solve normal equation $A^T Ax = A^T b$ directly using Cholesky factorization
  - Is unstable, but is very efficient if $m \gg n$ $(mn^2 + \frac{1}{3}n^3)$
- 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 $\hat{Q}\hat{R}x = \hat{Q}\hat{Q}^T b$
  - Left-multiply by $\hat{Q}^T$ and we get $\hat{R}x = \hat{Q}^T b$ (note $A^+ = \hat{R}^{-1}\hat{Q}^T$)

Least squares via QR Factorization

Compute reduced QR factorization $A = \hat{Q}\hat{R}$
Compute vector $c = \hat{Q}^T b$
Solve upper-triangular system $\hat{R}x = c$ for x

- Computation is dominated by QR factorization $(2mn^2 - \frac{2}{3}n^3)$
- Question: If Householder QR is used, how to compute $\hat{Q}^T b$?

Solution of Lease Squares Problems

- One approach is to solve normal equation $A^T Ax = A^T b$ directly using Cholesky factorization
  - Is unstable, but is very efficient if $m \gg n$ $(mn^2 + \frac{1}{3}n^3)$
- 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 $\hat{Q}\hat{R}x = \hat{Q}\hat{Q}^T b$
  - Left-multiply by $\hat{Q}^T$ and we get $\hat{R}x = \hat{Q}^T b$ (note $A^+ = \hat{R}^{-1}\hat{Q}^T$)

Least squares via QR Factorization

  Compute reduced QR factorization $A = \hat{Q}\hat{R}$
  Compute vector $c = \hat{Q}^T b$
  Solve upper-triangular system $\hat{R}x = c$ for x

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

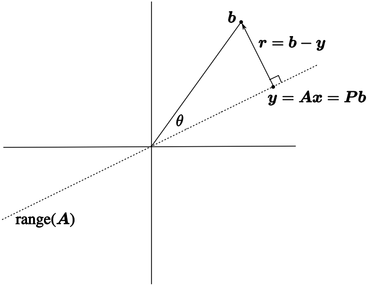
Outline

Four Conditioning Problems

- Least squares problem: Given $A \in \mathbb{R}^{m \times n}$ with full rank and $b \in \mathbb{R}^m$,

$$\min_{x \in \mathbb{R}^n} \|b - Ax\|$$

- Its solution is $x = A^+ b$. Another quantity is $y = Ax = Pb$, where $P = AA^+$



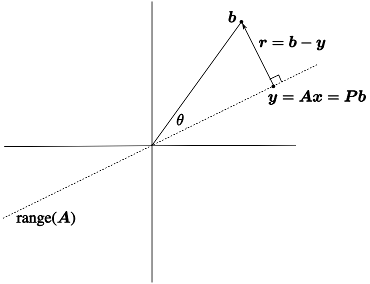
- Consider A and b as input data, and x and y as output. We then have four conditioning problems:

Input \ Output	y	x
b	$\kappa_{b \to y}$	$\kappa_{b \to x}$
A	$\kappa_{A \to y}$	$\kappa_{A \to x}$

- These conditioning problems are important and subtle.

Some Prerequisites

- We focus on the second column, namely $\kappa_{b\to x}$ and $\kappa_{A\to x}$
- However, understanding $\kappa_{b\to y}$ and $\kappa_{A\to y}$ is prerequisite
- Three quantities: (All in 2-norms)
  - Condition number of A: $\kappa(A) = \|A\|\|A^+\| = \sigma_1/\sigma_n$
  - Angle between b and y: $\theta = \arccos\frac{\|y\|}{\|b\|}$. $(0 \le \theta \le \pi/2)$
  - Orientation of y with range(A): $\eta = \frac{\|A\|\|x\|}{\|y\|}$. $(1 \le \eta \le \kappa(A))$



Sensitivity of y to Perturbations in b

- Intuition: The larger θ is, the more sensitive y is in terms of relative error
- Analysis: $y = Pb$, so

$$\kappa_{b \to y} = \frac{\|P\|}{\|y\|/\|b\|} = \frac{\|b\|}{\|y\|} = \frac{1}{\cos\theta},$$

where $\|P\| = 1$

Input \ Output	y	x
b	$\frac{1}{\cos\theta}$	
A		

- Question: When is the maximum attained for perturbation δb?

Sensitivity of y to Perturbations in b

- Intuition: The larger θ is, the more sensitive y is in terms of relative error
- Analysis: $y = Pb$, so

$$\kappa_{b \to y} = \frac{\|P\|}{\|y\|/\|b\|} = \frac{\|b\|}{\|y\|} = \frac{1}{\cos\theta},$$

where $\|P\| = 1$

Input \ Output	y	x
b	$\frac{1}{\cos\theta}$	
A		

- Question: When is the maximum attained for perturbation δb?
- Answer: When is δb in range(A)

Sensitivity of x to Perturbations in b

- Intuition: It depends on how sensitive y is to b, and how y lies within range(A)
- Analysis: $x = A^+ b$, so

$$\kappa_{b \to x} = \frac{\|A^+\|}{\|x\|/\|b\|} = \|A^+\| \frac{\|b\|}{\|y\|} \frac{\|y\|}{\|x\|} = \|A^+\| \frac{1}{\cos\theta} \frac{\|A\|}{\eta} = \frac{\kappa(A)}{\eta \cos\theta},$$

where $\eta = \|A\| \|x\| / \|y\|$

Input \ Output	y	x
b	$\frac{1}{\cos\theta}$	$\frac{\kappa(A)}{\eta\cos\theta}$
A		

Sensitivity of x to Perturbations in b

- Assume $\cos\theta = O(1)$, $\kappa_{b\to x} = \frac{\kappa(A)}{\eta\cos\theta}$ can lie anywhere between 1 and $O(\kappa(A))$!
- Question: When is the maximum attained for perturbation δb?

Sensitivity of x to Perturbations in b

- Assume $\cos\theta = O(1)$, $\kappa_{b\to x} = \frac{\kappa(A)}{\eta\cos\theta}$ can lie anywhere between 1 and $O(\kappa(A))$!
- Question: When is the maximum attained for perturbation δb?
- Answer: When δb is in subspace spanned by left singular vectors corresponding to smallest singular values
- Question: What if A is a nonsingular matrix?

Sensitivity of x to Perturbations in b

- Assume $\cos\theta = O(1)$, $\kappa_{b\to x} = \frac{\kappa(A)}{\eta\cos\theta}$ can lie anywhere between 1 and $O(\kappa(A))$!
- Question: When is the maximum attained for perturbation δb?
- Answer: When δb is in subspace spanned by left singular vectors corresponding to smallest singular values
- Question: What if A is a nonsingular matrix?
- Answer: $\kappa_{b\to x}$ can lie anywhere between 1 and $\kappa(A)$!

Sensitivity of y and x to Perturbations in A

- The relationship are nonlinear, because range(A) changes due to δA
- Intuitions:
  - The larger θ is, the more sensitive y is in terms of relative error.
  - Tilting of range(A) depends on $\kappa(A)$.
  - For x, it depends where y lies within range(A)

Input \ Output	y	x
b	$\frac{1}{\cos\theta}$	$\frac{\kappa(A)}{\eta\cos\theta}$
A	$\leq \frac{\kappa(A)}{\cos\theta}$	$\leq \kappa(A) + \frac{\kappa(A)^2\tan\theta}{\eta}$

- For second row, bounds are not necessarily tight
- Assume $\cos\theta = O(1)$, $\kappa_{A\to x}$ can lie anywhere between $\kappa(A)$ and $O(\kappa(A)^2)$

Condition Numbers of Linear Systems

- Linear system $Ax = b$ for nonsingular $A \in \mathbb{R}^{m \times m}$ is a special case of least squares problems, where $y = b$
- If $m = n$, then $\theta = 0$, so $\cos \theta = 1$ and $\tan \theta = 0$.

Input \ Output	y	x
b	1	$\kappa(A)/\eta$
A	-	$\leq \kappa(A)$