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

Lecture 9: Conditioning of Gaussian Elimination; Backward Stability of Gaussian Elimination

Xiangmin Jiao

Stony Brook University

Outline

1. Condition Number of Gaussian Elimination (NLA §22 & MC §3.3)
   - Perturbing Right-Hand Side
   - Perturbing Coefficient Matrix
   - Perturbing Both Sides

2. Backward Stability of LU Factorization (NLA §22 & MC §3.3)

3. Putting It All Together

Condition Number of Linear System

Theorem

Let A be nonsingular, and let x and $\hat{x} = x + \delta x$ be the solutions of $Ax = b$ and $A\hat{x} = b + \delta b$, respectively. Then

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A)\frac{\|\delta b\|}{\|b\|},$$

and there exist b and δb for which the equality holds.

- Proof sketch: $\|\delta x\| \leq \|A^{-1}\|\|\delta b\|$ and $\|b\| \leq \|A\|\|x\|$
- Question: For what b and δb is the equality achieved?

Condition Number of Linear System

Theorem

Let A be nonsingular, and let x and $\hat{x} = x + \delta x$ be the solutions of $Ax = b$ and $A\hat{x} = b + \delta b$, respectively. Then

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A)\frac{\|\delta b\|}{\|b\|},$$

and there exist b and δb for which the equality holds.

- Proof sketch: $\|\delta x\| \leq \|A^{-1}\|\|\delta b\|$ and $\|b\| \leq \|A\|\|x\|$
- Question: For what b and δb is the equality achieved?
  Answer: When b is in direction of *minimum* magnification of A^{-1}, and δb is in direction of *maximum* magnification of A^{-1}.
  In 2-norm, when b is in direction of *maximum* magnification of A^T, and δb is in direction of *minimum* magnification of A^T.
- We say a matrix is *nearly singular* if its condition number is very large.

Ill Conditioning Caused by Poor Scaling

- Some matrices are ill conditioned merely because they are out of scale.

Theorem

Let $A \in \mathbb{R}^{n\times n}$ be any nonsingular matrix, and let a_k, $1 \le k \le n$ denote the kth column of A. Then for any i and j with $1 \le i, j \le n$, $\kappa_p(A) \ge \|a_i\|_p / \|a_j\|_p$.

- Proof sketch: $\|A\|_p \ge \|a_i\|_p$ and $\|A^{-1}\|_p \ge 1/\|a_i\|_p$ for $1 \le i \le n$
- This theorem indicates that poor scaling inevitably leads to ill conditioning
- A *necessary* condition for a matrix to be well conditioned is that all of its rows and columns are of roughly the same magnitude.

Non-singularity of Perturbed Matrix

Theorem

If A is nonsingular and

$$\|\delta A\|/\|A\| < 1/\kappa(A),$$

then $A + \delta A$ is nonsingular.

Non-singularity of Perturbed Matrix

Theorem

If A is nonsingular and

$$\|\delta A\|/\|A\| < 1/\kappa(A),$$

then $A + \delta A$ is nonsingular.

Proof.

$\|\delta A\|/\|A\| < 1/\kappa(A)$ is equivalent to $\|\delta A\|\|A^{-1}\| < 1$. Suppose $A + \delta A$ is singular, then $\exists y \neq 0$ such that $(A + \delta A)y = 0$, and $y = -A^{-1}\delta A y$. Therefore, $\|y\| \leq \|A^{-1}\|\|\delta A\|\|y\|$, or $\|A^{-1}\|\|\delta A\| \geq 1$. $\square$

- If $A + \delta A$ is the singular matrix closest to A, in the sense that $\|\delta A\|_2$ is as small as possible, then $\|\delta A\|_2/\|A\|_2 = 1/\kappa_2(A)$

Linear System with Perturbed Matrix

- Suppose $Ax = b$ and $\hat{A}\hat{x} = b$ where $\hat{A} = A + \delta A$. Let $\delta x = \hat{x} - x$ and $\hat{x} = x + \delta x$.
- We would like to bound $\|\delta x\|/\|x\|$, but first we bound $\|\delta x\|/\|\hat{x}\|$

Theorem

If A is nonsingular, and let $b \neq 0$. Then

$$\frac{\|\delta x\|}{\|\hat{x}\|} \leq \kappa(A)\frac{\|\delta A\|}{\|A\|}.$$

Proof.

Rewrite $(A+\delta A)\hat{x} = b$ as $Ax + A\delta x + \delta A\hat{x} = b$, where $Ax = b$. Therefore,

$$\|\delta x\| \leq \|A^{-1}\|\|\delta A\|\|\hat{x}\|.$$

Therefore,

$$\frac{\|\delta x\|}{\|\hat{x}\|} \leq \|A^{-1}\|\|\delta A\| = \kappa(A)\frac{\|\delta A\|}{\|A\|}.$$

Linear System with Perturbed Matrix Cont'd

- $Ax = b$ and $\hat{A}\hat{x} = b$, where $\hat{A} = A + \delta A$. Let $\delta x = \hat{x} - x$ and $\hat{x} = x + \delta x$.

Theorem

If A is nonsingular and $\|\delta A\|/\|A\| < 1/\kappa(A)$, and let $b \neq 0$. Then

$$\frac{\|\delta x\|}{\|x\|} \leq \frac{\kappa(A)\|\delta A\|/\|A\|}{1 - \kappa(A)\|\delta A\|/\|A\|}.$$

Proof.

$\|\delta x\| \leq \|A^{-1}\|\|\delta A\|\|\hat{x}\| \leq \|A^{-1}\|\|\delta A\|(\|x\| + \|\delta x\|)$. Therefore,

$$\left(1 - \|A^{-1}\|\|\delta A\|\right)\|\delta x\| \leq \|A^{-1}\|\|\delta A\|\|x\|,$$

where $\|A^{-1}\|\|\delta A\| = \kappa(A)\|\delta A\|/\|A\|$. $\square$

We typically expect $\kappa(A)\|\delta A\| \ll \|A\|$, so the denominator is close to 1.

Perturbed RHS and Matrix

- $Ax = b$ and $\hat{A}\hat{x} = \hat{b}$, where $\hat{A} = A + \delta A$, $\hat{b} = b + \delta b$, and $\hat{x} = x + \delta x$.

Theorem (Thm 2.3.8, Fundamentals of Matrix Computations 3rd ed.)

Suppose A is nonsingular, $\hat{x} \neq 0$, and $\hat{b} \neq 0$. Then

$$\frac{\|\delta x\|}{\|\hat{x}\|} \leq \kappa(A)\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|\hat{b}\|} + \frac{\|\delta A\|}{\|A\|}\frac{\|\delta b\|}{\|\hat{b}\|}\right).$$

Proof.

$A\delta x = \delta b - \delta A\hat{x}$. Hence$\|\delta x\| \leq \|A^{-1}\|(\|\delta A\|\|\hat{x}\| + \|\delta b\|)$, and

$$\frac{\|\delta x\|}{\|\hat{x}\|} \leq \kappa(A)\left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|\hat{x}\|\|A\|}\right).$$

Furthermore, $\frac{1}{\|\hat{x}\|\|A\|} \leq \frac{\|\hat{A}\|}{\|A\|\|\hat{b}\|} \leq \frac{\|A\|+\|\delta A\|}{\|A\|\|\hat{b}\|} = \frac{1}{\|\hat{b}\|} + \frac{\|\delta A\|}{\|A\|\|\hat{b}\|}$. $\square$

Perturbed RHS and Matrix

We can simply the previous theorem to be

$$\frac{\|\delta x\|}{\|\hat{x}\|} \lesssim \kappa(A) \left(\frac{\|\delta A\|}{\|A\|} + \frac{\|\delta b\|}{\|\hat{b}\|} \right)$$

and then obtain the following result:

Theorem

If A is nonsingular and $\|\delta A\|/\|A\| < 1/\kappa(A)$, and let $b \neq 0$, then

$$\frac{\|\delta x\|}{\|x\|} \lesssim \frac{\kappa(A)(\|\delta A\|/\|A\| + \|\delta b\|/\|b\|)}{1 - \kappa(A)\|\delta A\|/\|A\|}.$$

Roughly speaking, $\kappa(A)$ determines extra loss of digits of accuracy in x in additional to input errors in A and b

Outline

1. Condition Number of Gaussian Elimination (NLA §22 & MC §3.3)
   - Perturbing Right-Hand Side
   - Perturbing Coefficient Matrix
   - Perturbing Both Sides

2. Backward Stability of LU Factorization (NLA §22 & MC §3.3)

3. Putting It All Together

Stability of LU without Pivoting

- For $A = LU$ computed without pivoting (Theorem 22.1 of NLA)

$$\tilde{L}\tilde{U} = A + \delta A, \qquad \frac{\|\delta A\|}{\|L\|\|U\|} = O(\epsilon_{\text{machine}})$$

(Theorem 3.3.1 of *Matrix Computations*, 4th ed., Golub & Van Loan)

- This is close to backward stability, except that we have $\|L\|\|U\|$ instead of $\|A\|$ in the denominator
- Instability of Gaussian elimination can happen only if one or both of the factors L and U is large relative to size of A
- Unfortunately, $\|L\|$ and $\|U\|$ can be arbitrarily large (even for well-conditioned A), e.g.,

$$A = \begin{bmatrix} 10^{-20} & 1 \\ 1 & a_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 10^{20} & 1 \end{bmatrix} \begin{bmatrix} 10^{-20} & 1 \\ 0 & a_{22} - 10^{20} \end{bmatrix}$$

- If $\mathrm{fl}(a_{22} - 10^{20}) = 10^{20}$, we obtain same L and U regardless of a_{22}
- L and U are not exact for a nearby A, so algorithm is *unstable*

Stability of LU with Partial Pivoting

- With pivoting, all entries of L are in $[-1, 1]$, so $\|L\| = O(1)$
- To measure growth in U, we introduce the growth factor $\rho = \frac{\max_{i,j} |u_{ij}|}{\max_{i,j} |a_{ij}|}$, and hence $\|U\| = O(\rho\|A\|)$
- We then have $PA = LU$

$$\tilde{L}\tilde{U} = \tilde{P}A + \delta A, \qquad \frac{\|\delta A\|}{\|A\|} = O(\rho\epsilon_{\text{machine}})$$

- If $|\ell_{ij}| < 1$ for each $i > j$ (i.e., there is no tie for the pivoting), then $\tilde{P} = P$ for sufficiently small $\epsilon_{\text{machine}}$
- If $\rho = O(1)$, then the algorithm is backward stable
- In fact, $\rho \le 2^{n-1}$, so by definition ρ is a constant but can be very large

The Growth Factor

- ρ can indeed be as large as 2^{n-1}. Consider matrix

$$\begin{bmatrix} 1 & & & & 1 \\ -1 & 1 & & & 1 \\ -1 & -1 & 1 & & 1 \\ -1 & -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & -1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & & & & 0 \\ -1 & 1 & & & 0 \\ -1 & -1 & 1 & & 0 \\ -1 & -1 & -1 & 1 & 0 \\ -1 & -1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & & & & 1 \\ & 1 & & & 2 \\ & & 1 & & 4 \\ & & & 1 & 8 \\ & & & & 16 \end{bmatrix}$$

  where growth factor $\rho = 16 = 2^{n-1}$
- $\rho = 2^{n-1}$ is as large as ρ can get. It can be catastrophic in practice
- Theoretically, Gaussian elimination with partial pivoting is backward stable according to formal definition
- However, in the worst case, Gaussian elimination with partial pivoting may be unstable for practical values of n

The Growth Factor in Practice

- **Good news**: Large ρ occurs only for very skewed matrices. Experimentally, one rarely see very large ρ
- Probability of large ρ decreases exponentially in ρ
- "If you pick a billion matrices at random, you will almost certainly not find one for which Gaussian elimination is unstable"
- In practice, ρ is no larger than $O(\sqrt{n})$. However, this behavior is not fully understood yet
- In conclusion,
  - Gaussian elimination with partial pivoting is backward stable
  - In theory, its error may grow exponentially in n
  - In practice, it is stable for matrices of practical interests

Outline

Accuracy of Linear Solver

- Solving $Ax = b$ using LU factorization with partial pivoting is also backward stable
  1. $PA = LU$
  2. $Ly = Pb$
  3. $Ux = y$
- Each step is backward stable (we omit detailed proof)
- Overall growth factor of error is bounded by product of growth factors of individual steps

A Posteriori Error Analysis Using Residual

- Suppose $\hat{x}$ is a computed solution of $Ax = b$, and residual $\hat{r} = b - A\hat{x}$.
- Let A be nonsingular and $b \neq 0$. Then $\frac{\|\delta x\|}{\|x\|} \leq O(\kappa(A))\frac{\|\hat{r}\|}{\|b\|}$.
- If the residual is tiny and A is well conditioned, then $\hat{x}$ is an accurate approximation to x.
- For *a posteriori* error bound, one needs to estimate $\|\hat{r}\|$ and $\kappa(A)$
- Typically one estimates $\kappa_1(A) = \|A\|_1\|A^{-1}\|_1$ without computing A^{-1}, but allow LU factorization of A
    - For any vector $w \in \mathbb{R}^n$ and $\|w\|_1 = 1$, we have lower bound $\kappa_1(A) \geq \|A\|_1\|A^{-1}w\|_1$
    - If w has a significant component in direction near maximum magnification by A^{-1}, then $\kappa_1(A) \approx \|A\|_1\|A^{-1}w\|_1$
    - Good estimators conduct systematic searches for w that approximately maximizes $\|A^{-1}w\|_1$