

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

Lecture 5: Projector; Conditioning and Condition Number

Xiangmin Jiao

SUNY Stony Brook

Outline

1 Projectors (NLA§6)

2 Conditioning and Condition Numbers (NLA§12)

Projectors (Projection Matrices)

- A *projector* (aka *projection matrix*) $P \in \mathbb{R}^{m \times m}$ (or $P \in \mathbb{C}^{m \times m}$) satisfies $P^2 = P$. They are also said to be *idempotent*.
 - ▶ *Orthogonal* projector
 - ▶ *Oblique* projector
- An *orthogonal projector* is one that projects onto a subspace S_1 along a space S_2 , where S_1 and S_2 are orthogonal.
 - ▶ $S_1 = \text{range}(P)$
 - ▶ $S_2 = \text{null}(P)$
- Example: $\begin{bmatrix} 0 & 0 \\ \alpha & 1 \end{bmatrix}$
 - ▶ is an oblique projector if $\alpha \neq 0$,
 - ▶ is orthogonal projector if $\alpha = 0$.

Orthogonal Projector

Theorem

A projector $P \in \mathbb{R}^{m \times m}$ is orthogonal if and only if $P = P^T$.

Note: An alternative definition of *orthogonal projection* is $P^2 = P$ and $P = P^T$, and it projects onto $S = \text{range}(P)$. For complex matrices, we need to replace $P = P^T$ with $P = P^H$.

Proof.

“If” direction: If $P = P^T$, then $(Px)^T(I - P)y = x^T(P - P^2)y = 0$.

“Only if” direction: Use SVD. Suppose P projects onto S_1 along S_2 where $S_1 \perp S_2$, and S_1 has dimension n . Let $\{q_1, \dots, q_n\}$ be orthonormal basis of S_1 and $\{q_{n+1}, \dots, q_m\}$ be a basis for S_2 . Let Q be orthogonal matrix whose j th column is q_j , and we have $PQ = (q_1, q_2, \dots, q_n, 0, \dots, 0)$, so $Q^T P Q = \text{diag}(\underbrace{1, 1, \dots, 1}_n, 0, \dots) = \Sigma$, and $P = Q \Sigma Q^T$. □

Orthogonal Projector

Theorem

A projector $P \in \mathbb{R}^{m \times m}$ is orthogonal if and only if $P = P^T$.

Note: An alternative definition of *orthogonal projection* is $P^2 = P$ and $P = P^T$, and it projects onto $S = \text{range}(P)$. For complex matrices, we need to replace $P = P^T$ with $P = P^H$.

Proof.

“If” direction: If $P = P^T$, then $(Px)^T(I - P)y = x^T(P - P^2)y = 0$.

“Only if” direction: Use SVD. Suppose P projects onto S_1 along S_2 where $S_1 \perp S_2$, and S_1 has dimension n . Let $\{q_1, \dots, q_n\}$ be orthonormal basis of S_1 and $\{q_{n+1}, \dots, q_m\}$ be a basis for S_2 . Let Q be orthogonal matrix whose j th column is q_j , and we have $PQ = (q_1, q_2, \dots, q_n, 0, \dots, 0)$, so $Q^T P Q = \text{diag}(\underbrace{1, 1, \dots, 1}_n, 0, \dots) = \Sigma$, and $P = Q \Sigma Q^T$. □

Discussion: Are orthogonal projectors orthogonal matrices?

Complementary Projectors

- Complementary projectors: P vs. $I - P$.
- What space does $I - P$ project onto?

Complementary Projectors

- Complementary projectors: P vs. $I - P$.
- What space does $I - P$ project onto?
 - ▶ Answer: $\text{null}(P)$.
 - ▶ $\text{range}(I - P) \supseteq \text{null}(P)$ because $Pv = 0 \Rightarrow (I - P)v = v$.
 - ▶ $\text{range}(I - P) \subseteq \text{null}(P)$ because for any v $(I - P)v = v - Pv \in \text{null}(P)$.
- A projector separates \mathbb{R}^m into two complementary subspaces: range space and null space (i.e., $\text{range}(P) + \text{null}(P) = \mathbb{R}^m$ and $\text{range}(P) \cap \text{null}(P) = \{0\}$ for projector $P \in \mathbb{R}^{m \times m}$)
- It projects onto range space along null space
 - ▶ In other words, $x = Px + r$, where $r = (I - P)x \in \text{null}(P)$
- Discussion: Are range space and null space of projector orthogonal to each other?

Uniqueness of Orthogonal Projectors

- Orthogonal projector for a subspace is unique
- In other words, for $S \subseteq \mathbb{R}^m$ be a subspace, if P_1 and P_2 are each orthogonal projector onto S , then $P_1 = P_2$
- Proof: For any $z \in \mathbb{R}^m$,

$$\begin{aligned}\|(P_1 - P_2)z\|_2^2 &= z^T(P_1 - P_2)(P_1 - P_2)z \\ &= z^T P_1(P_1 - P_2)z - z^T P_2(P_1 - P_2)z \\ &= z^T P_1(I - P_2)z + z^T P_2(I - P_1)z \\ &= (P_1 z)^T(I - P_2)z + (P_2 z)^T(I - P_1)z \\ &= 0\end{aligned}$$

Therefore, $\|P_1 - P_2\|_2 = 0$, and $P_1 = P_2$.

Projections with Orthonormal Basis

- Given unit vector q , $P_q = qq^T$ and $P_{\perp q} = I - P_q$
- Given any matrix $Q \in \mathbb{R}^{m \times n}$ whose columns q_j are orthonormal, $P = QQ^T = \sum_j q_j q_j^T$ is orthogonal projector onto $\text{range}(Q)$
- SVD-related projections
 - ▶ Suppose $A = U\Sigma V^T \in \mathbb{R}^{m \times n}$ is SVD of A , and $r = \text{rank}(A)$
 - ▶ Partition U and V to

$$U = \begin{bmatrix} U_r & \tilde{U}_r \end{bmatrix}, \quad V = \begin{bmatrix} V_r & \tilde{V}_r \end{bmatrix}$$

$\begin{matrix} r & m-r \end{matrix} \qquad \qquad \begin{matrix} r & n-r \end{matrix}$

- ▶ What do $U_r U_r^T$, $\tilde{U}_r \tilde{U}_r^T$, $V_r V_r^T$, and $\tilde{V}_r \tilde{V}_r^T$ project onto, respectively?

Projections with Orthonormal Basis

- Given unit vector q , $P_q = qq^T$ and $P_{\perp q} = I - P_q$
- Given any matrix $Q \in \mathbb{R}^{m \times n}$ whose columns q_j are orthonormal, $P = QQ^T = \sum_j q_j q_j^T$ is orthogonal projector onto $\text{range}(Q)$
- SVD-related projections
 - Suppose $A = U\Sigma V^T \in \mathbb{R}^{m \times n}$ is SVD of A , and $r = \text{rank}(A)$
 - Partition U and V to

$$U = \begin{bmatrix} U_r & \tilde{U}_r \\ r & m-r \end{bmatrix}, \quad V = \begin{bmatrix} V_r & \tilde{V}_r \\ r & n-r \end{bmatrix}$$

- What do $U_r U_r^T$, $\tilde{U}_r \tilde{U}_r^T$, $V_r V_r^T$, and $\tilde{V}_r \tilde{V}_r^T$ project onto, respectively?
 - ★ Answer: $\text{range}(A)$, $\text{null}(A^T)$, $\text{range}(A^T)$, $\text{null}(A)$

Projection with Arbitrary Basis

- For arbitrary vector $a \in \mathbb{R}^m$, we write $P_a = \frac{aa^T}{a^T a}$ and $P_{\perp a} = I - P_a$
- Given any matrix $A \in \mathbb{R}^{m \times n}$ that **has full rank** and $m \geq n$. Let $A = U\Sigma V^T$ be its SVD

$$P = UU^T = A(A^T A)^{-1}A^T$$

is orthogonal projection onto $\text{range}(A)$

- $(A^T A)^{-1}A^T$ is called the *pseudo-inverse* of A , denoted as A^+
- Therefore,

$$P = UU^T = AA^+$$

- In addition, $A^+A = I$
- Note: If $m < n$, $A^+ = A^T(AA^T)^{-1}$, and $AA^+ = I$ and A^+A is orthogonal projection onto $\text{range}(A^T)$

Distance Between Subspaces

- Suppose S_1 and S_2 are subspaces of \mathbb{R}^m , $\dim(S_1) = \dim(S_2)$, and P_i is orthogonal projection onto S_i
- The *distance* between S_1 and S_2 is

$$\text{dist}(S_1, S_2) = \|P_1 - P_2\|_2$$

- Suppose $W = \begin{bmatrix} W_1 & W_2 \end{bmatrix}$, $Z = \begin{bmatrix} Z_1 & Z_2 \end{bmatrix}$ are $n \times n$ orthogonal matrices. If $S_1 = \text{range}(W_1)$ and $S_2 = \text{range}(Z_1)$, then

$$\text{dist}(S_1, S_2) = \|W_1^T Z_2\|_2 = \|Z_1^T W_2\|_2$$

(proof omitted here)

- In general, $0 \leq \text{dist}(S_1, S_2) \leq 1$
- Discussion: If S_1 and S_2 are lines in 2D, what is $\text{dist}(S_1, S_2)$?

Outline

1 Projectors (NLA§6)

2 Conditioning and Condition Numbers (NLA§12)

Overview of Error Analysis

- Error analysis is important subject of numerical analysis
- Given a problem f and an algorithm \tilde{f} with an input x , the *absolute error* is $\|\tilde{f}(x) - f(x)\|$ and relative error is $\|\tilde{f}(x) - f(x)\|/\|f(x)\|$
- What are possible sources of errors?

Overview of Error Analysis

- Error analysis is important subject of numerical analysis
- Given a problem f and an algorithm \tilde{f} with an input x , the *absolute error* is $\|\tilde{f}(x) - f(x)\|$ and relative error is $\|\tilde{f}(x) - f(x)\|/\|f(x)\|$
- What are possible sources of errors?
 - ▶ Round-off error (input, computation) – main concern of NLA
 - ▶ truncation (approximation) error – main concern for AMS 527
- We would like the solution to be *accurate*, i.e., with small errors
- Error depends on property (*conditioning*) of the problem, property (*stability*) of the algorithm

Example Using SVD Analysis

- Suppose $A \in \mathbb{R}^{n \times n}$ is nonsingular, and let $A = U\Sigma V^T$ be its SVD
- Then

$$Ax = U\Sigma V^T x = \sum_{i=1}^n \sigma_i v_i^T x u_i$$

and

$$x = A^{-1}b = \left(U\Sigma V^T\right)^{-1} b = \sum_{i=1}^n \frac{u_i^T b}{\sigma_i} v_i$$

- Whether matrix multiplication and linear system are sensitive to small changes in A or b depends on distribution of singular values; nearly zero σ_n can amplify errors in b along direction of u_n
- How do we formalize this analysis?

Absolute Condition Number

- Condition number is a measure of *sensitivity* of a problem
- *Absolute condition number* of a problem f at x is

$$\hat{\kappa} = \lim_{\varepsilon \rightarrow 0} \sup_{\|\delta x\| \leq \varepsilon} \frac{\|\delta f\|}{\|\delta x\|}$$

where $\delta f = f(x + \delta x) - f(x)$

- Less formally, $\hat{\kappa} = \sup_{\delta x} \frac{\|\delta f\|}{\|\delta x\|}$ for infinitesimally small δx
- If f is differentiable, then

$$\hat{\kappa} = \|J(x)\|$$

where J is the Jacobian of f at x , with $J_{ij} = \partial f_i / \partial x_j$, and matrix norm is induced by vector norms on ∂f and ∂x

- Discussion: What is absolute condition number of $f(x) = \alpha x$?
- Discussion: Is absolute condition number scale dependent?

Relative Condition Number

- Relative condition number of f at x is

$$\kappa = \lim_{\varepsilon \rightarrow 0} \sup_{\|\delta x\| \leq \varepsilon} \frac{\|\delta f\| / \|f(x)\|}{\|\delta x\| / \|x\|}$$

- Less formally, $\kappa = \sup_{\delta x} \frac{\|\delta f\| / \|\delta x\|}{\|f(x)\| / \|x\|}$ for infinitesimally small δx
- Note: we can use different types of norms to get different condition numbers
- If f is differentiable, then

$$\kappa = \frac{\|J(x)\|}{\|f(x)\| / \|x\|}$$

- Discussion: What is relative condition number of $f(x) = \alpha x$?
- Discussion: Is relative condition number scale dependent?
- In numerical analysis, we in general use relative condition number
- A problem is *well-conditioned* if κ is small and is *ill-conditioned* if κ is large

Examples

- Example: Function $f(x) = \sqrt{x}$

Examples

- Example: Function $f(x) = \sqrt{x}$
 - ▶ Absolute condition number of f at x is $\hat{\kappa} = \|J\| = 1/(2\sqrt{x})$
 - ★ Note: We are talking about the condition number of the problem for a given x
 - ▶ Relative condition number $\kappa = \frac{\|J\|}{\|f(x)\|/\|x\|} = \frac{1/(2\sqrt{x})}{\sqrt{x}/x} = 1/2$
- Example: Function $f(x) = x_1 - x_2$, where $x = (x_1, x_2)^T$

Examples

- Example: Function $f(x) = \sqrt{x}$
 - ▶ Absolute condition number of f at x is $\hat{\kappa} = \|J\| = 1/(2\sqrt{x})$
 - ★ Note: We are talking about the condition number of the problem for a given x
 - ▶ Relative condition number $\kappa = \frac{\|J\|}{\|f(x)\|/\|x\|} = \frac{1/(2\sqrt{x})}{\sqrt{x}/x} = 1/2$
- Example: Function $f(x) = x_1 - x_2$, where $x = (x_1, x_2)^T$
 - ▶ Absolute condition number of f at x in ∞ -norm is $\hat{\kappa} = \|J\|_\infty = \|(1, -1)\|_\infty = 2$
 - ▶ Relative condition number $\kappa = \frac{\|J\|_\infty}{\|f(x)\|_\infty/\|x\|_\infty} = \frac{2}{|x_1 - x_2|/\max\{|x_1|, |x_2|\}}$
 - ▶ κ is arbitrarily large (f is ill-conditioned) if $x_1 \approx x_2$ (hazard of cancellation error)
- Note: From now on, we will talk about only relative condition number

Condition Number of Matrix

- Consider $f(x) = Ax$, with $A \in \mathbb{R}^{m \times n}$

$$\kappa = \frac{\|J\|}{\|f(x)\|/\|x\|} = \frac{\|A\|\|x\|}{\|Ax\|}$$

- If A is square and nonsingular, since $\|x\|/\|Ax\| \leq \|A^{-1}\|$

$$\kappa \leq \|A\|\|A^{-1}\|$$

- Note that for $f(b) = A^{-1}b$, its condition number $\kappa \leq \|A\|\|A^{-1}\|$
- We define *condition number of matrix* A as

$$\kappa(A) = \|A\|\|A^{-1}\|$$

- It is the upper bound of the condition number of $f(x) = Ax$ for any x
- For any induced matrix norm, $\kappa(I) = 1$ and $\kappa(A) \geq 1$
- Note about the distinction between the condition number of a *problem* (the map $f(x)$) and the condition number of a *problem instance* (the evaluation of $f(x)$ for specific x)

Geometric Interpretation of Condition Number

- Another way to interpret $\kappa(A)$ is

$$\kappa(A) = \sup_{\delta x, x} \frac{\|\delta f\|/\|\delta x\|}{\|f(x)\|/\|x\|} = \frac{\sup_{\delta x} \|A\delta x\|/\|\delta x\|}{\inf_x \|Ax\|/\|x\|}$$

- Question: For what x and δx is the equality achieved?

Geometric Interpretation of Condition Number

- Another way to interpret $\kappa(A)$ is

$$\kappa(A) = \sup_{\delta x, x} \frac{\|\delta f\|/\|\delta x\|}{\|f(x)\|/\|x\|} = \frac{\sup_{\delta x} \|A\delta x\|/\|\delta x\|}{\inf_x \|Ax\|/\|x\|}$$

- Question: For what x and δx is the equality achieved?
 - ▶ Answer: When x is in direction of minimum magnification, and δx is in direction of maximum magnification
- Define *maximum magnification* of A as

$$\text{maxmag}(A) = \max_{\|x\|=1} \|Ax\|$$

and *minimum magnification* of A as

$$\text{minmag}(A) = \min_{\|x\|=1} \|Ax\|$$

- Then condition number of matrix is $\kappa(A) = \text{maxmag}(A)/\text{minmag}(A)$
- For 2-norm, $\kappa(A) = \sigma_1/\sigma_n$, ratio of largest and smallest singular values

Example of Ill-Conditioned Matrix

Example

Let $A = \begin{bmatrix} 1000 & 999 \\ 999 & 998 \end{bmatrix}$. It is easy to verify that

$$A^{-1} = \begin{bmatrix} -998 & 999 \\ 999 & -1000 \end{bmatrix}. \text{ So}$$

$$\kappa_{\infty}(A) = \kappa_1(A) = 1999^2 = 3.996 \times 10^6.$$

Example of Ill-Conditioned Matrix

Example

A famous example of ill-conditioning is Hilbert matrix, defined by $h_{ij} = 1/(i + j - 1)$, $1 \leq i, j \leq n$. For example, for $n = 4$, we have

$$H_4 = \begin{bmatrix} 1 & 1/2 & 1/3 & 1/4 \\ 1/2 & 1/3 & 1/4 & 1/5 \\ 1/3 & 1/4 & 1/5 & 1/6 \\ 1/4 & 1/5 & 1/6 & 1/7 \end{bmatrix}.$$

This sequence of matrices are ill-conditioned even for quite small n . In particular, $\kappa_2(H_4) \approx 1.6 \times 10^4$, $\kappa_2(H_8) \approx 1.5 \times 10^{10}$, and $\kappa_2(H_{10}) \approx 1.6 \times 10^{13}$. The condition number grows exponentially in n .

Note that this example is an extreme case. Most matrices arising from practical applications are not nearly as bad.