EECS 275 Matrix Computation

Ming-Hsuan Yang

Electrical Engineering and Computer Science
University of California at Merced
Merced, CA 95344
http://faculty.ucmerced.edu/mhyang

Lecture 3
Overview

- Basic definition: matrix norm, range, rank, null space, matrix inverse
- Elementary analytical and topological properties
Reading

- Chapter 1-3 *Numerical Linear Algebra* by Trefethen and Bau
- Chapter 2 of *Matrix Computations* by Gene Golub and Charles Van Loan
- Chapter 5 of *Matrix Analysis and Applied Linear Algebra* by Carl Meyer
- Chapter 2 of *Optimization by Vector Space Methods* by David Luenberger
- Chapter 3 and Chapter 4 of *Matrix Algebra From a Statistician’s Perspective* by David Harville
Matrix norm

- Matrix form: a function, \( f : \mathbb{R}^{m \times n} \to \mathbb{R} \) is a matrix norm if the following properties hold:
  
  \[
  \begin{align*}
  f(A) & \geq 0 \\
  f(A + B) & \leq f(A) + f(B) \\
  f(\alpha A) & = |\alpha| f(A)
  \end{align*}
  \]
  \( A \in \mathbb{R}^{m \times n}, \ A, B \in \mathbb{R}^{m \times n}, \ \alpha \in \mathbb{R}, A \in R^{m \times n} \)

- Frobenius norm:
  \[
  \|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2} = \sum_{i=1}^{m} \|A(i, :)\|_2^2 = \sum_{j=1}^{n} \|A(:, j)\|_2^2 = \sqrt{\text{tr}(A^T A)}
  \]

- The Frobenius norm suggests
  \[
  \|Ax\|_2^2 = \sum_{i=1}^{n} |A(i, :)x|^2 \leq \sum_{i=1}^{n} \|A(i, :)\|_2^2 \|x\|_2^2 = \|A\|_F^2 \|x\|_2^2
  \]
  \[
  \|Ax\|_2 \leq \|A\|_F \|x\|_2
  \]

- For matrices \( A \) and \( B \)
  \[
  \|AB\|_F \leq \|A\|_F \|B\|_F
  \]
Matrix norm (cont’d)

- $p$-norms:

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

Note that matrix $p$-norms are defined in terms of vector $p$-norms.

- It is clear that $\|A\|_p$ is the $p$-norm of the largest vector obtained by applying $A$ to a unit $p$-norm vector

$$\|A\|_p = \sup_{x \neq 0} \left\| A \left( \frac{x}{\|x\|_p} \right) \right\|_p = \max_{\|x\|_p=1} \|Ax\|_p$$

- When $A$ is non-singular,

$$\min_{\|x\|_p=1} \|Ax\|_p = \frac{1}{\|A^{-1}\|_p}$$

- Frobenius norm and $p$-norms define families norms that

$$\|AB\|_p \leq \|A\|_p \|B\|_p \quad A \in \mathbb{R}^{m \times n}, B \in \mathbb{R}^{n \times q}$$

- For every $A \in \mathbb{R}^{m \times n}$ and $x \in \mathbb{R}^n$, we have

$$\|Ax\|_p \leq \|A\|_p \|x\|_p$$
Matrix norm (cont’d)

- Not all matrix norms satisfy the sub-multiplicative property

\[ \|AB\| \leq \|A\| \|B\| \]

- For example, if \( \|A\|_\Delta = \max |a_{ij}| \), and

\[ A = B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \]

then \( \|AB\|_\Delta > \|A\|_\Delta \|B\|_\Delta \)
More generally, for any vector norm $\| \cdot \|_\alpha$ on $\mathbb{R}^n$ and $\| \cdot \|_\beta$ on $\mathbb{R}^m$, we have $\|Ax\|_\beta \leq \|A\|_{\alpha,\beta} \|x\|_\alpha$ where $\|A\|_{\alpha,\beta}$ is a matrix norm defined by

$$\|A\|_{\alpha,\beta} = \sup_{x \neq 0} \frac{\|Ax\|_\beta}{\|x\|_\alpha}$$

We say that $\| \cdot \|_{\alpha,\beta}$ is subordinate to the vector norms $\| \cdot \|_\alpha$ and $\| \cdot \|_\beta$.

Since the set $\{x \in \mathbb{R}^n : \|x\|_\alpha = 1\}$ is compact and $\| \cdot \|_\beta$ is continuous, it follows that

$$\|A\|_{\alpha,\beta} = \max_{\|x\|_\alpha = 1} \|Ax\|_\beta = \|Ax^*\|_\beta$$

for some $x^* \in \mathbb{R}^n$ having unit $\alpha$-norm.
Matrix norm properties

For $A \in \mathbb{R}^{m \times n}$, the Frobenius and $p$-norms satisfy certain important properties

\[
\|A\|_2 \leq \|A\|_F \leq \sqrt{n} \|A\|_2 \\
\max_{i,j} |a_{ij}| \leq \|A\|_2 \leq \sqrt{mn} \max_{i,j} |a_{ij}| \\
\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |a_{ij}| \\
\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |a_{ij}| \\
\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{m} \|A\|_\infty \\
\frac{1}{\sqrt{m}} \|A\|_1 \leq \|A\|_2 \leq \sqrt{n} \|A\|_1 \\
\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty} \\
\|A(i_1 : i_2, j_1 : j_2)\|_p \leq \|A\|_p
\]
Example

Given a matrix,

\[
A = \begin{bmatrix}
3 & 5 & 7 \\
2 & 6 & 4 \\
0 & 2 & 8 \\
\end{bmatrix}
\]

- \(\|A\|_1 = \max(3 + 2 + 0, 5 + 6 + 2, 7 + 4 + 8) = 19\) (maximum absolute column sum)
- \(\|A\|_{\infty} = \max(3 + 5 + 7, 2 + 6 + 4, 0 + 2 + 8) = 15\) (maximum absolute row sum)
- \(\|A\|_F = 14.3875\)
Matrix 2-norm

**Theorem**

If $A \in \mathbb{R}^{m \times n}$, then there exists a unit 2-norm $z \in \mathbb{R}^n$ such that $A^\top A z = \mu^2 z$ where $\mu = \|A\|_2$.

**Proof.**

Suppose $z \in R^n$ is a unit vector such that $\|Az\|_2 = \|A\|_2$. Since $z$ maximizes the function

$$g(x) = \frac{1}{2} \frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{1}{2} \frac{x^\top A^\top A x}{x^\top x},$$

it follows that with by setting gradient $\nabla g(z) = 0$,

$$\frac{\partial g(z)}{\partial z_i} = \left[ (z^\top z) \sum_{j=1}^n (A^\top A)_{ij} z_j - (z^\top A^\top A z) z_i \right] / (z^\top z)^2, \forall i$$

In vector notation, $A^\top A z = (z^\top A^\top A z) z$. The theorem follows by setting $\mu = \|Az\|_2$. 

Matrix 2-norm (cont’d)

- It implies that $\|A\|_2^2$ is a zero of the polynomial $p(\lambda) = \det(A^\top A - \lambda I)$, i.e.,

\[
\|A\|_2 = \max_{\|x\|_2 = 1} \|Ax\|_2 = \sqrt{\lambda_{\text{max}}}
\]

where $\lambda_{\text{max}}$ is the largest eigenvalue.

- $\|A\|_2$ is the square root of the largest eigenvalue of $A^\top A$.

- When $A$ is non-singular

\[
\|A^{-1}\|_2 = \frac{1}{\min_{\|x\|_2 = 1} \|Ax\|_2} = \frac{1}{\sqrt{\lambda_{\text{min}}}}
\]

where $\lambda_{\text{min}}$ is the smallest eigenvalue of $A^\top A$. 

Matrix 2-norm (cont’d)

- $\|A\|$ represents the maximum extent to which a vector on the unit sphere can be stretched by $A$.
- $\frac{1}{\|A^{-1}\|}$ measures the extent to which a non-singular matrix $A$ can shrink vectors on the unit sphere.
Matrix 2-norm (cont’d)

- Computation of matrix 2-norm is iterative and more complicated than that of the matrix 1-norm or ∞-norm.
- The order of magnitude of $\|A\|_2$ can be computed easily.

**Corollary**

If $A \in \mathbb{IR}^{m \times n}$, then $\|A\|_2 \leq \sqrt{\|A\|_1 \|A\|_\infty}$

**Proof.**

If $z \neq 0$ is such that $A^T Az = \mu^2 z$ with $\mu = \|A\|_2$, then

$\mu^2 \|z\|_1 = \|A^T Az\|_1 \leq \|A^T\|_1 \|A\|_1 \|z\|_1 = \|A\|_\infty \|A\|_1 \|z\|_1$
Example

- Given the matrix
  \[
  \begin{bmatrix}
  1 & 2 \\
  0 & 2 \\
  \end{bmatrix}
  \]
  and two points \((1, 0)\) and \((0, 1)\)
- The amplification factors for 1-norm is 4, and \(\infty\)-norm is 3
- The amplification factor for 2-norm is at least \(\sqrt{8} \approx 2.8284\) as \((0, 1)\) is mapped to \((2, 2)\). In fact \(\|A\|_2 \approx 2.9208\)

![Diagram of unit balls and their images under matrix A]

Figure 3.1. *On the left, the unit balls of \(\mathbb{R}^2\) with respect to \(\| \cdot \|_1\), \(\| \cdot \|_2\), and \(\| \cdot \|_\infty\). On the right, their images under the matrix \(A\) of (3.7). Dashed lines mark the vectors that are amplified most by \(A\) in each norm.*
Independence, subspace, and span

- A set of vectors \( \{a_1, \ldots, a_n\} \in \mathbb{R}^m \) is linearly independent if
  \[
  \sum_{j=1}^{n} \alpha_j a_j = 0 \implies \alpha_j = 0
  \]
- Otherwise, nontrivial combination of the \( a_j \) is zero and \( \{a_1, \ldots, a_n\} \) is linearly dependent
- A subspace of \( \mathbb{R}^m \) is a subset that is also a vector space
- Given a collection of vectors, \( a_1, \ldots, a_n \in \mathbb{R}^m \), the set of all linear combinations of vectors in a subspace referred to as the span of \( \{a_1, \ldots, a_n\} \)
  \[
  \text{span}\{a_1, \ldots, a_n\} = \sum_{j=1}^{n} \beta_j a_j : \beta_j \in \mathbb{R}
  \]
- If \( \{a_1, \ldots, a_n\} \) is independent and \( b \in \text{span}\{a_1, \ldots, a_n\} \), then \( b \) is a unique linear combination of the \( a_j \)
Range, null space, and rank

- **Range:** The range of $A \in \mathbb{R}^{m \times n}$ is

  \[
  \text{ran}(A) = \{ y \in \mathbb{R}^m : y = Ax, x \in \mathbb{R}^n \}
  \]

- **Null space:**

  \[
  \text{null}(A) = \{ x \in \mathbb{R}^n : Ax = 0 \}
  \]

- If $A = [a_1, \ldots, a_n]$, then

  \[
  \text{ran}(A) = \text{span}\{a_1, \ldots, a_n\}
  \]

- **Rank:** the number of linear independent columns of $A$.

  \[
  \text{rank}(A) = \dim(\text{ran}(A))
  \]

- $A$ is rank deficient if $\text{rank}(A) < \min\{m, n\}$.

- If $A \in \mathbb{R}^{m \times n}$, then

  \[
  \dim(\text{null}(A)) + \text{rank}(A) = n
  \]
Matrix inverse

- The $n$-by-$n$ identity matrix $I_n$ is defined by the columns
  \[ I = [e_1, \ldots, e_n] \]
  where
  \[ e_k = (0, \ldots, 0, 1, 0, \ldots, 0)^\top \]

- If $A$ and $X$ are in $\mathbb{R}^{n \times n}$ and satisfy $AX = I$, then $X$ is the inverse of $A$ and is denoted by $A^{-1}$ (i.e., $AA^{-1} = I$).
- If $A^{-1}$ exists, then $A$ is said to be nonsingular. Otherwise, $A$ is singular.
- Several matrix inverse properties
  \[ (AB)^{-1} = B^{-1}A^{-1} \]
  \[ B^{-1} = A^{-1} - B^{-1}(B - A)A^{-1} \]
- Sherman-Morrison-Woodbury formula
  \[ (A + UV^\top)^{-1} = A^{-1} - A^{-1}U(I + V^\top A^{-1}U)^{-1}V^\top A^{-1} \]