Lecture 03
Matrix Norms; Singular Value Decomposition

Songting Luo

Department of Mathematics
Iowa State University

MATH 562 Numerical Analysis II

Outline

Outline

1. Matrix Norms

2. Singular Value Decomposition

Matrix Norms Induced by Vector Norms

- Viewing $m \times n$ matrix as mn-vectors is not always useful, as operations involving $m \times n$ matrices do not behave this way
- Induced matrix norms capture such behavior.

Matrix Norms Induced by Vector Norms

- Viewing $m \times n$ matrix as mn-vectors is not always useful, as operations involving $m \times n$ matrices do not behave this way
- Induced matrix norms capture such behavior.

Definition

Given vector norms $\|\cdot\|_{(n)}$ and $\|\cdot\|_{(m)}$ on domain and range of $\mathbf{A} \in \mathbb{C}^{m\times n}$, respectively, the induced matrix norm $\|\mathbf{A}\|_{(m,n)}$ is the smallest number $C \in \mathbb{R}$ for which the following inequality holds for all $\mathbf{x} \in \mathbb{C}^n$:

$$\|\mathbf{A}\mathbf{x}\|_{(m)} \leqslant C\|\mathbf{x}\|_{(n)}.$$

Matrix Norms Induced by Vector Norms

- Viewing $m \times n$ matrix as mn-vectors is not always useful, as operations involving $m \times n$ matrices do not behave this way
- Induced matrix norms capture such behavior.

Definition

Given vector norms $\|\cdot\|_{(n)}$ and $\|\cdot\|_{(m)}$ on domain and range of $\mathbf{A} \in \mathbb{C}^{m\times n}$, respectively, the induced matrix norm $\|\mathbf{A}\|_{(m,n)}$ is the smallest number $C \in \mathbb{R}$ for which the following inequality holds for all $\mathbf{x} \in \mathbb{C}^n$:

$$\|\mathbf{A}\mathbf{x}\|_{(m)} \leqslant C\|\mathbf{x}\|_{(n)}.$$

- In other words, it is supremum of ratio $\|\mathbf{A}\mathbf{x}\|_{(m)}/\|\mathbf{x}\|_{(n)}$ for all nonzero vectors $\mathbf{x} \in \mathbb{C}^n$. Maximum factor by which $\mathbf{A}$ can "strech" $\mathbf{x} \in \mathbb{C}^n$

$$\|\mathbf{A}\|_{(m,n)} = \sup_{\mathbf{x}\in\mathbb{C}^n, \mathbf{x}\neq\mathbf{0}} \|\mathbf{A}\mathbf{x}\|_{(m)}/\|\mathbf{x}\|_{(n)} = \sup_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_{(n)}=1} \|\mathbf{A}\mathbf{x}\|_{(m)}$$

Matrix Norms Induced by Vector Norms

- Viewing $m \times n$ matrix as mn-vectors is not always useful, as operations involving $m \times n$ matrices do not behave this way
- Induced matrix norms capture such behavior.

Definition

Given vector norms $\|\cdot\|_{(n)}$ and $\|\cdot\|_{(m)}$ on domain and range of $\mathbf{A} \in \mathbb{C}^{m\times n}$, respectively, the induced matrix norm $\|\mathbf{A}\|_{(m,n)}$ is the smallest number $C \in \mathbb{R}$ for which the following inequality holds for all $\mathbf{x} \in \mathbb{C}^n$:

$$\|\mathbf{A}\mathbf{x}\|_{(m)} \leqslant C\|\mathbf{x}\|_{(n)}.$$

- In other words, it is supremum of ratio $\|\mathbf{A}\mathbf{x}\|_{(m)}/\|\mathbf{x}\|_{(n)}$ for all nonzero vectors $\mathbf{x} \in \mathbb{C}^n$. Maximum factor by which $\mathbf{A}$ can "strech" $\mathbf{x} \in \mathbb{C}^n$

$$\|\mathbf{A}\|_{(m,n)} = \sup_{\mathbf{x}\in\mathbb{C}^n, \mathbf{x}\neq\mathbf{0}} \|\mathbf{A}\mathbf{x}\|_{(m)}/\|\mathbf{x}\|_{(n)} = \sup_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_{(n)}=1} \|\mathbf{A}\mathbf{x}\|_{(m)}$$

- Is vector norm consistent with matrix norm of $m \times 1$ matrix?

1-norm

- By definition

$$\|\mathbf{A}\|_1 = \sup_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1$$

1-norm

- By definition

$$\|\mathbf{A}\|_1 = \sup_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1$$

- What is it equal to?
    - Maximum of 1-norm of column vectors of **A**
    - "maximum column sum" of **A**

1-norm

- By definition

$$\|\mathbf{A}\|_1 = \sup_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_1=1} \|\mathbf{A}\mathbf{x}\|_1$$

- What is it equal to?
  - Maximum of 1-norm of column vectors of **A**
  - "maximum column sum" of **A**
- To prove it, note that for $\mathbf{x} \in \mathbb{C}^n$, $\|\mathbf{x}\|_1 = 1$

$$\|\mathbf{A}\mathbf{x}\|_1 = \|\sum_{j=1}^{n} x_j \mathbf{a}_j\|_1 \leqslant \sum_{j=1}^{n} |x_j| \|\mathbf{a}_j\|_1 \leqslant \max_{1\leqslant j\leqslant n} \|\mathbf{a}_j\|_1 \|\mathbf{x}\|_1$$

- Let $k = arg\max_{1\leqslant j\leqslant n} \|\mathbf{a}_j\|_1$, then $\|\mathbf{A}\mathbf{e}_k\|_1 = \|\mathbf{a}_k\|_1$, so $\max_{1\leqslant j\leqslant n} \|\mathbf{a}_j\|_1$ is tight upper bound

∞-norm

- By definition

$$\|\mathbf{A}\|_\infty = \sup_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_\infty = 1} \|\mathbf{A}\mathbf{x}\|_\infty$$

∞-norm

- By definition

$$\|\mathbf{A}\|_\infty = \sup_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_\infty = 1} \|\mathbf{A}\mathbf{x}\|_\infty$$

- What is it equal to?
    - Maximum of 1-norm of row vectors of $\mathbf{A}$
    - Maximum of 1-norm of column vectors of $\mathbf{A}^*$
    - "maximum row sum" of $\mathbf{A}$

∞-norm

- By definition

$$\|\mathbf{A}\|_\infty = \sup_{\mathbf{x}\in\mathbb{C}^n, \|\mathbf{x}\|_\infty = 1} \|\mathbf{A}\mathbf{x}\|_\infty$$

- What is it equal to?
  - Maximum of 1-norm of row vectors of $\mathbf{A}$
  - Maximum of 1-norm of column vectors of $\mathbf{A}^*$
  - "maximum row sum" of $\mathbf{A}$
- To prove it, note that for $\mathbf{x} \in \mathbb{C}^n$, $\|\mathbf{x}\|_\infty = 1$

$$\|\mathbf{A}\mathbf{x}\|_\infty = \max_{1\leqslant i\leqslant m} |\mathbf{a}_i^*\mathbf{x}| \leqslant \max_{1\leqslant i\leqslant m} \|\mathbf{a}_i^*\|_1 \|\mathbf{x}\|_\infty$$

  where $\mathbf{a}_i^*$ denotes i-th row vector of $\mathbf{A}$.
- Furthermore, $\max_{1\leqslant i\leqslant m} \|\mathbf{a}_i^*\|_1$ is a tight bound.
  - Which vector can we choose to reach the bound?
  -

2-norm

- What is 2-norm of a matrix?

2-norm

- What is 2-norm of a matrix?
- Answer: Its largest singular value.
- We will talk more about singular-value decomposition

2-norm

- What is 2-norm of a matrix?
- Answer: Its largest singular value.
- We will talk more about singular-value decomposition

- What is 2-norm of a diagonal matrix?

2-norm

- What is 2-norm of a matrix?
- Answer: Its largest singular value.
- We will talk more about singular-value decomposition

- What is 2-norm of a diagonal matrix?

Cauchy-Schwarz and Hölder Inequalities

- Hölder Inequality: let p and q satisfy $1/p + 1/q = 1$ with $1 \leqslant p, q \leqslant \infty$, then
$$|\mathbf{x}^*\mathbf{y}| \leqslant \|\mathbf{x}\|_p \|\mathbf{y}\|_q$$
- Cauchy-Schwarz inequality ($p = q = 2$),
$$|\mathbf{x}^*\mathbf{y}| \leqslant \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

Cauchy-Schwarz and Hölder Inequalities

- Hölder Inequality: let p and q satisfy $1/p + 1/q = 1$ with $1 \leqslant p, q \leqslant \infty$, then

$$|\mathbf{x}^* \mathbf{y}| \leqslant \|\mathbf{x}\|_p \|\mathbf{y}\|_q$$

- Cauchy-Schwarz inequality ($p = q = 2$),

$$|\mathbf{x}^* \mathbf{y}| \leqslant \|\mathbf{x}\|_2 \|\mathbf{y}\|_2$$

- Example: What is 2-norm of rank-one matrix? Hint: Use Cauchy-Schwarz inequality.

Bounding Matrix-Matrix Multiplication

- Let **A** be an $I \times m$ matrix and **B** be an $m \times n$ matrix, then

$$\|\mathbf{AB}\|_{(l,n)} \leqslant \|\mathbf{A}\|_{(l,m)}\|\mathbf{B}\|_{(m,n)}$$

Bounding Matrix-Matrix Multiplication

- Let **A** be an $I \times m$ matrix and **B** be an $m \times n$ matrix, then

$$\|\mathbf{AB}\|_{(l,n)} \leqslant \|\mathbf{A}\|_{(l,m)}\|\mathbf{B}\|_{(m,n)}$$

- To prove it, note

$$\|\mathbf{ABx}\| \leqslant \|\mathbf{A}\|\|\mathbf{Bx}\| \leqslant \|\mathbf{A}\|\|\mathbf{B}\|\|\mathbf{x}\|$$

Bounding Matrix-Matrix Multiplication

- Let **A** be an $I \times m$ matrix and **B** be an $m \times n$ matrix, then

$$\|\mathbf{AB}\|_{(l,n)} \leqslant \|\mathbf{A}\|_{(l,m)}\|\mathbf{B}\|_{(m,n)}$$

- To prove it, note

$$\|\mathbf{ABx}\| \leqslant \|\mathbf{A}\|\|\mathbf{Bx}\| \leqslant \|\mathbf{A}\|\|\mathbf{B}\|\|\mathbf{x}\|$$

- In general, this inequality is not an equality
- In particular, $\|\mathbf{A}^n\| \leqslant \|\mathbf{A}\|^n$ but $\|\mathbf{A}^n\| \neq \|\mathbf{A}\|^n$ in general for $n \geqslant 2$.

General Matrix Norms

- One can view $m \times n$ matrices as mn-dimensional vectors and obtain general matrix norms, which satisfy (for $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{m\times n}$)
  - $\|\mathbf{A}\| \geqslant 0$, and $\|\mathbf{A}\| = 0$ only if $\mathbf{A} = \mathbf{0}$.
  - $\|\mathbf{A} + \mathbf{B}\| \leqslant \|\mathbf{A}\| + \|\mathbf{B}\|$
  - $\|\alpha\mathbf{A}\| = |\alpha|\|\mathbf{A}\|$

Frobenius Norm

- One useful norm is Frobenius norm (a.k.a. Hilbert-Schmidt norm)

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} = \sqrt{\sum_{j=1}^{n}\|\mathbf{a}_j\|_2^2}$$

  i.e., 2-norm of mn-vector

- Furthermore,

$$\|\mathbf{A}\|_F = \sqrt{tr(\mathbf{A}^*\mathbf{A})}$$

  where $tr(\mathbf{B})$ denotes trace of $\mathbf{B}$, the sum of its diagonal entries

Frobenius Norm

- One useful norm is Frobenius norm (a.k.a. Hilbert-Schmidt norm)

$$\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|^2} = \sqrt{\sum_{j=1}^{n}\|\mathbf{a}_j\|_2^2}$$

  i.e., 2-norm of mn-vector
- Furthermore,

$$\|\mathbf{A}\|_F = \sqrt{tr(\mathbf{A}^*\mathbf{A})}$$

  where $tr(\mathbf{B})$ denotes trace of $\mathbf{B}$, the sum of its diagonal entries
- Note that

$$\|\mathbf{AB}\|_F \leqslant \|\mathbf{A}\|_F\|\mathbf{B}\|_F$$

  because

$$\|\mathbf{AB}\|_F = \sum_{i=1}^{n}\sum_{j=1}^{n}|\mathbf{a}_i^*\mathbf{b}_j|^2 \leqslant \sum_{i=1}^{n}\sum_{j=1}^{n}(\|\mathbf{a}_i^*\|_2\|\mathbf{b}_j\|_2)^2 \leqslant \|\mathbf{A}\|_F^2\|\mathbf{B}\|_F^2$$

Invariance under Unitary Multiplication

Theorem

For any $\mathbf{A} \in \mathbb{C}^{m\times n}$ *and unitary* $\mathbf{Q} \in \mathbb{C}^{m\times m}$, *we have*

$$\|\mathbf{QA}\|_2 = \|\mathbf{A}\|_2 \text{ and } \|\mathbf{QA}\|_F = \|\mathbf{A}\|_F$$

In other words, 2-norm and Frobenius norms are invariant under unitary multiplication.

Invariance under Unitary Multiplication

Theorem

For any $\mathbf{A} \in \mathbb{C}^{m\times n}$ *and unitary* $\mathbf{Q} \in \mathbb{C}^{m\times m}$, *we have*

$$\|\mathbf{QA}\|_2 = \|\mathbf{A}\|_2 \text{ and } \|\mathbf{QA}\|_F = \|\mathbf{A}\|_F$$

In other words, 2-norm and Frobenius norms are invariant under unitary multiplication.

- Proof for 2-norm: $\|\mathbf{Qy}\|_2 = \|\mathbf{y}\|_2$ for $\mathbf{y} \in \mathbb{C}^m$ and therefore $\|\mathbf{QAx}\|_2 = \|\mathbf{Ax}\|_2$ for $\mathbf{x} \in \mathbb{C}^n$. It then follows from definition of 2-norm.

Invariance under Unitary Multiplication

Theorem

For any $\mathbf{A} \in \mathbb{C}^{m \times n}$ *and unitary* $\mathbf{Q} \in \mathbb{C}^{m \times m}$, *we have*

$$\|\mathbf{QA}\|_2 = \|\mathbf{A}\|_2 \text{ and } \|\mathbf{QA}\|_F = \|\mathbf{A}\|_F$$

In other words, 2-norm and Frobenius norms are invariant under unitary multiplication.

- Proof for 2-norm: $\|\mathbf{Qy}\|_2 = \|\mathbf{y}\|_2$ for $\mathbf{y} \in \mathbb{C}^m$ and therefore $\|\mathbf{QAx}\|_2 = \|\mathbf{Ax}\|_2$ for $\mathbf{x} \in \mathbb{C}^n$. It then follows from definition of 2-norm.
- Proof for Frobenius norm:

$$\|\mathbf{QA}\|_F^2 = tr((\mathbf{QA})^*(\mathbf{QA})) = tr(\mathbf{A}^*\mathbf{Q}^*\mathbf{QA}) = tr(\mathbf{A}^*\mathbf{A}) = \|\mathbf{A}\|_F^2$$

Outline

1. Matrix Norms

2. Singular Value Decomposition

Geometric Observation

- The image of unit sphere under any $m \times n$ matrix is a hyperellipse
- Given a unit sphere $S \in \mathbb{R}^n$, let $\mathbf{A}S$ denote the shape after transformation
- SVD is
$$\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$$
where $\mathbf{U} \in \mathbb{C}^{m\times m}$ and $\mathbf{V} \in \mathbb{C}^{n\times n}$ and $\mathbf{\Sigma} \in \mathbf{R}^{m\times n}$ is diagonal.
- Singular values are diagonal entries of $\mathbf{\Sigma}$, correspond to the principal semiaxes, with entries $\sigma_1 \geqslant \sigma_2 \geqslant \cdots \geqslant \sigma_n \geqslant 0$.
- Left singular vectors of $\mathbf{A}$ are column vectors of $\mathbf{U}$ and are oriented in the directions of the principal semiaxes of $\mathbf{A}S$
- Right singular vectors of $\mathbf{A}$ are column vectors of $\mathbf{V}$ and are the preimages of the principal semiaxes of $\mathbf{A}S$
- $\mathbf{A}\mathbf{v}_j = \sigma_j \mathbf{u}_j$ for $1 \leqslant j \leqslant n$