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

Lecture 3: Matrix Norms; Singular Value Decomposition

Xiangmin Jiao

SUNY Stony Brook

Outline

Frobenius Norm

- One can define a norm by viewing $m \times n$ matrix as vectors in $\mathbb{R}^{mn}$
- One useful norm is Frobenius norm (a.k.a. Hilbert-Schmidt norm)

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

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

- Furthermore,

$$\|A\|_F = \sqrt{\operatorname{tr}(A^T A)}$$

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

- Note that for $A \in \mathbb{R}^{n \times \ell}$ and $B \in \mathbb{R}^{\ell \times m}$,

$$\|AB\|_F \leq \|A\|_F \|B\|_F$$

because

$$\|AB\|_F^2 = \sum_{i=1}^{n} \sum_{j=1}^{m} |a_i^T b_j|^2 \leq \sum_{i=1}^{n} \sum_{j=1}^{m} \left(\|a_i^T\|_2 \|b_j\|_2 \right)^2 = \|A\|_F^2 \|B\|_F^2$$

General Definition of Matrix Norms

- However, viewing $m \times n$ matrix as vectors in $\mathbb{R}^{mn}$ is not always useful, because matrix operations do not behave this way
- Similar to vector norms, *general matrix norms* has the following properties (for $A, B \in \mathbb{R}^{m\times n}$)

$$
\begin{aligned}
&(1)\|A\| \geq 0, \text{ and } \|A\| = 0 \text{ only if } A = 0,\\
&(2)\|A+B\| \leq \|A\| + \|B\|,\\
&(3)\|\alpha A\| = |\alpha|\|A\|.
\end{aligned}
$$

- In addition, a matrix norm for $A, B \in \mathbb{R}^{n\times n}$ typically satisfies

$$\|AB\| \leq \|A\|\|B\|, \qquad \text{(submultiplicativity)}$$

which is a generalization of Cauchy-Schwarz inequality

Norms Induced by Vector Norms

- Matrix norms can be *induced* from vector norms, which can better capture behaviors of matrix-vector multiplications

Definition

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

$$\|Ax\|_{(m)} \leq C\|x\|_{(n)}.$$

- In other words, it is supremum of $\|Ax\|_{(m)}/\|x\|_{(n)}$ for all $x \in \mathbb{R}^n \backslash \{0\}$
- Maximum factor by which A can "stretch" $x \in \mathbb{R}^n$

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

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

1-norm

- By definition

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

- What is it equal to?

1-norm

- By definition
$$\|A\|_1 = \sup_{x\in\mathbb{R}^n, \|x\|_1=1} \|Ax\|_1$$
- What is it equal to?
  - Maximum of 1-norm of column vectors of A
  - Or maximum of column sum of absolute values of A, "column-sum norm"
- To show it, note that for $x \in \mathbb{R}^n$ and $\|x\|_1 = 1$
$$\|Ax\|_1 = \left\|\sum_{j=1}^{n} x_j a_j\right\|_1 \le \sum_{j=1}^{n} |x_j| \|a_j\|_1 \le \max_{1\le j\le n} \|a_j\|_1 \|x\|_1$$
- Let $k = \arg\max_{1\le j\le n} \|a_j\|_1$, then $\|Ae_k\|_1 = \|a_k\|_1$, so $\max_{1\le j\le n} \|a_j\|_1$ is tight upper bound

∞-norm

- By definition

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

- What is $\|A\|_\infty$ equal to?

∞-norm

- By definition

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

- What is $\|A\|_\infty$ equal to?
  - Maximum of 1-norm of column vectors of A^T
  - Or maximum of row sum of absolute values of A, "row-sum norm"
- To show it, note that for $x \in \mathbb{R}^n$ and $\|x\|_\infty = 1$

$$\|Ax\|_\infty = \max_{1\le i\le m} |a_{i,:}x| \le \max_{1\le i\le m} \|a_{i,:}^T\|_1 \|x\|_\infty$$

where $a_{i,:}$ denotes ith row vector of A and $\|a_{i,:}^T\|_1 = \sum_{j=1}^n |a_{ij}|$

- Furthermore, $\|a_{i,:}^T\|_1$ is a tight bound.
- Which vector can we choose for x for equality to hold?

2-norm

- What is 2-norm of a matrix?

2-norm

- What is 2-norm of a matrix?
  Answer: Its largest singular value, which we will explain in later lectures
- What is 2-norm of a diagonal matrix D?

2-norm

- What is 2-norm of a matrix?
  Answer: Its largest singular value, which we will explain in later lectures
- What is 2-norm of a diagonal matrix D?
  Answer: $\|D\|_2 = \max_{i=1}^{n}\{|d_{ii}|\}$
- What is 2-norm of rank-one matrix uv^T? Hint: Use Cauchy-Schwarz inequality.

2-norm

- What is 2-norm of a matrix?
  Answer: Its largest singular value, which we will explain in later lectures
- What is 2-norm of a diagonal matrix D?
  Answer: $\|D\|_2 = \max_{i=1}^{n}\{|d_{ii}|\}$
- What is 2-norm of rank-one matrix uv^T? Hint: Use Cauchy-Schwarz inequality.
  Answer: $\|uv^T\|_2 = \|u\|_2\|v\|_2$.

Bounding Matrix-Matrix Multiplication

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

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

- To show it, note for $x \in \mathbb{R}^n$

$$\|ABx\|_{(l)} \leq \|A\|_{(l,m)}\|Bx\|_{(m)} \leq \|A\|_{(l,m)}\|B\|_{(m,n)}\|x\|_{(n)},$$

- In general, this inequality is not an equality
- In particular, $\|A^p\| \leq \|A\|^p$ but $\|A^p\| \neq \|A\|^p$ in general for $p \geq 2$

Invariance under Orthogonal Transformation

- Given matrix $Q \in \mathbb{R}^{\ell \times m}$ with $\ell \geq m$. If $Q^T Q = I$, then Qx for $x \in \mathbb{R}^m$ corresponds to orthogonal transformation to coordinate system in $\mathbb{R}^\ell$
- If $Q \in \mathbb{R}^{m \times m}$, then Q is said to be an *orthogonal* matrix

Theorem

For any $A \in \mathbb{R}^{m \times n}$ and $Q \in \mathbb{R}^{\ell \times m}$ with $Q^T Q = I$ and $\ell \geq m$, we have

$$\|QA\|_2 = \|A\|_2 \textit{ and } \|QA\|_F = \|A\|_F.$$

In other words, 2-norm and Frobenius norms are invariant under orthogonal transformation.

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

Proof for Frobenius norm:

$$\|QA\|_F^2 = \operatorname{tr}\left((QA)^T QA\right) = \operatorname{tr}\left(A^T Q^T QA\right) = \operatorname{tr}\left(A^T A\right) = \|A\|_F^2.$$

Outline

Singular Value Decomposition (SVD)

- Given $A \in \mathbb{R}^{m \times n}$, its *SVD* is

$$A = U \Sigma V^T$$

  where $U \in \mathbb{R}^{m \times m}$ and $V \in \mathbb{R}^{n \times n}$ are orthogonal, and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal
- If $A \in \mathbb{C}^{m \times n}$, then its SVD is $A = U \Sigma V^H$, where $U \in \mathbb{C}^{m \times m}$ and $V \in \mathbb{C}^{n \times n}$ are unitary, and $\Sigma \in \mathbb{R}^{m \times n}$ is diagonal
- *Singular values* are diagonal entries of Σ, with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$
- *Left singular vectors* of A are column vectors of U
- *Right singular vectors* of A are column vectors of V and are the preimages of the principal semiaxes of AS
- SVD plays prominent role in data analysis and matrix analysis

Geometric Observation

- Image of unit sphere under any $m \times n$ matrix is a *hyperellipse*
- Give unit sphere S in $\mathbb{R}^n$, AS denotes shape after transformation
- *Singular values* correspond to the principal semiaxes of hyerellipse
- *Left singular vectors* are oriented in directions of principal semiaxes of AS
- *Right singular vectors* are preimages of principal semiaxes of AS
- $Av_j = \sigma_j u_j$ for $1 \leq j \leq n$

Two Different Types of SVD

- **Full SVD**: $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ is

$$A = U \Sigma V^T$$

Two Different Types of SVD

- **Full SVD**: $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ is

$$A = U \Sigma V^T$$

- **Thin SVD (Reduced SVD)**: $\hat{U} \in \mathbb{R}^{m \times n}$, $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ (assume $m \geq n$)

$$A = \hat{U} \hat{\Sigma} V^T$$

Two Different Types of SVD

- **Full SVD**: $U \in \mathbb{R}^{m \times m}$, $\Sigma \in \mathbb{R}^{m \times n}$, and $V \in \mathbb{R}^{n \times n}$ is

$$A = U \Sigma V^T$$

- **Thin SVD (Reduced SVD)**: $\hat{U} \in \mathbb{R}^{m \times n}$, $\hat{\Sigma} \in \mathbb{R}^{n \times n}$ (assume $m \geq n$)

$$A = \hat{U} \hat{\Sigma} V^T$$

- Furthermore, notice that

$$A = \sum_{i=1}^{\min\{m,n\}} \sigma_i u_i v_i^T$$

so we can keep only entries of U and V corresponding to nonzero σ_i.