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

Lecture 3: Vector Norms; Matrix Norms

Xiangmin Jiao

SUNY Stony Brook

Outline

Definition of Norms

- Norm captures "size" of vector or "distance" between vectors
- There are many different measures for "sizes" but a norm must satisfy some requirements:

Definition

A *norm* is a function $\| \cdot \| : \mathbb{R}^n \to \mathbb{R}$ that assigns a real-valued length to each vector. It must satisfy the following conditions:

$$(1)\|x\| \geq 0, \text{ and } \|x\| = 0 \text{ only if } x = 0,$$
$$(2)\|x + y\| \leq \|x\| + \|y\|,$$
$$(3)\|\alpha x\| = |\alpha| \|x\|.$$

- An example is Euclidean length (i.e, $\|x\| = \sqrt{\sum_{i=1}^{n} |x_i|^2}$)

p-norms

- Euclidean length is a special case of p-norms, defined as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

  for $1 \leq p \leq \infty$
- Euclidean norm is 2-norm $\|x\|_2$ (i.e., $p = 2$)
- 1-norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
- ∞-norm: $\|x\|_\infty$. What is its value?

p-norms

- Euclidean length is a special case of p-norms, defined as

$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

  for $1 \leq p \leq \infty$
- Euclidean norm is 2-norm $\|x\|_2$ (i.e., $p = 2$)
- 1-norm: $\|x\|_1 = \sum_{i=1}^{n} |x_i|$
- ∞-norm: $\|x\|_\infty$. What is its value?
  - Answer: $\|x\|_\infty = \max_{1 \leq i \leq n} |x_i|$
- Why we require $p \geq 1$? What happens if $0 \leq p < 1$?

Some Properties of Vector Norms

- Hölder inequality: $|x^T y| \leq \|x\|_p \|y\|_q, \quad \frac{1}{p} + \frac{1}{q} = 1$
- In particular when $p = q = 2$: $|x^T y| \leq \|x\|_2 \|y\|_2$ (Cauchy-Schwarz inequality)
- All norms are *equivalent*, in that there exists c_1 and c_2 s.t.

$$c_1 \|x\|_\alpha \leq \|x\|_\beta \leq c_2 \|x\|_\alpha$$

  where c_1 and c_2 are constants and may depend on n. e.g.,

$$\|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2$$

- 2-norm is preserved under orthogonal transformation Qx

$$\|Qx\|_2 = \|x\|_2$$

Weighted p-norms

- A generalization of p-norm is *weighted p-norm*, which assigns different weights (priorities) to different components.
  - It is anisotropic instead of isotropic
- Algebraically, $\|x\|_W = \|Wx\|$, where W is diagonal matrix with ith diagonal entry $w_i \neq 0$ being weight for ith component
- In other words,

$$\|x\|_W = \left(\sum_{i=1}^{n} |w_i x_i|^p \right)^{1/p}$$

- What happens if we allow $w_i = 0$?

Weighted p-norms

- A generalization of p-norm is *weighted p-norm*, which assigns different weights (priorities) to different components.
  - It is anisotropic instead of isotropic
- Algebraically, $\|x\|_W = \|Wx\|$, where W is diagonal matrix with ith diagonal entry $w_i \neq 0$ being weight for ith component
- In other words,

$$\|x\|_W = \left(\sum_{i=1}^{n} |w_i x_i|^p \right)^{1/p}$$

- What happens if we allow $w_i = 0$?
- Can we further generalize it to allow W being arbitrary matrix?
  - No. But we can allow W to be arbitrary nonsingular matrix.

A-norm of Vectors

- Given a positive definite matrix $A \in \mathbb{R}^{n\times n}$, the A-norm on $\mathbb{R}^n$ is

$$\|x\|_A = \sqrt{x^T Ax}$$

- Note: Weighted 2-norm with W is A-norm with $A = W^2$.
- These conventions are somewhat inconsistent, but they are both commonly used in the literature

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,

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

  where $\mathrm{tr}(B)$ 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 $\|x\|_1 = 1$

$$\|Ax\|_1 = \left\| \sum_{j=1}^{n} x_j a_j \right\|_1 \leq \sum_{j=1}^{n} |x_j| \|a_j\|_1 \leq \max_{1 \leq j \leq n} \|a_j\|_1 \|x\|_1$$

- Let $k = \arg\max_{1 \leq j \leq n} \|a_j\|_1$, then $\|Ae_k\|_1 = \|a_k\|_1$, so $\max_{1 \leq j \leq n} \|a_j\|_1$ 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

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

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:

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

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