Vv417 Lecture 19

Jing Liu

UM-SJTU Joint Institute

November 12, 2019

- The notion of vector space gives us a way to introduce some structure into arbitrary sets, however, we still lack a key concept that we have for $\mathbb{R}^n$.
- Recall for points in $\mathbb{R}^n$

$$A(a_1, a_2, \dots, a_n) \qquad \text{and} \qquad B(b_1, b_2, \dots, b_n)$$

we have the concept of distance between A and B

$$d = d(A, B) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

- From this definition of distance, we have countless indispensable theorems in geometry regarding relationships between subsets of $\mathbb{R}^n$.

Q: How can introduce similar structure into arbitrary sets?

- In general, a metric is a function that associate $x, y \in \mathcal{S}$ to a real number

$$d = d(x, y)$$

where the real number d is called the distance between x and y.

Definition

Let $\mathcal{S}$ be a nonempty set. A metric on $\mathcal{S}$ is a function

$$d\colon \mathcal{S} \times \mathcal{S} \to \mathbb{R}$$

such that for all x, y, $z \in \mathcal{S}$, the followings are true:

1. Nonnegative:
$$d(x, y) \geq 0$$
2. Unique:
$$d(x, y) = 0 \iff x = y$$
3. Symmetric:
$$d(x, y) = d(y, x)$$
4. Subadditive:
$$d(y, z) \leq d(x, y) + d(x, z)$$

A set $\mathcal{S}$ together with a metric d is called a metric space.

- Consider the set of continuous functions

$$\mathcal{C}[a,b]$$

and the following function

$$T(f,g) = \int_a^b \Big| f(x) - g(x) \Big| \, dx$$

Q: Can we use T as a metric for $\mathcal{C}[a,b]$?

- Given a set of a sequence of real or complex scalars

$$x = \Big\{ x_k \Big\}_{k=1}^{\infty}$$

- Let us use the following notation for the value of the following series

$$\|x\|_1 = \sum_{k=1}^{\infty} |x_k|$$

- The series is said to be absolutely convergent and x absolutely summable if

$$\|x\|_1 < \infty$$

- The set of all absolutely summable sequence is often denoted by

$$\ell_1 = \left\{ x = \{x_k\} : \ \|x\|_1 < \infty \right\}$$

Q: Is ℓ_1 a vector space?

Q: Is the following function a valid metric for ℓ_1,

$$d(x, y) = \sum_{k=1}^{\infty} |x_k - y_k|, \qquad x, y \in \ell_1$$

- This is known as the Manhattan distance, aka ℓ_1-distance.

Q: Is every metric space a vector space?

- If $\mathcal{S}$ a a generic metric space, then we often refer to the elements of $\mathcal{S}$ as

"points",

if $\mathcal{S}$ is also a vector space, then we usually refer to the elements of $\mathcal{S}$ as

"vectors".

- Having the notion of distance in a space is important, because we can now define the corresponding notion of convergence in the space.

Definition

Let $\mathcal{S}$ be a metric space. A sequence of points $\left\{a_k\right\}$ in $\mathcal{S}$ converges to $a \in \mathcal{S}$ if

$$\lim_{n\to\infty} d(a, a_n) = 0$$

This is, for every $\epsilon > 0$, there exists some integer $N > 0$ such that

$$d(a, a_n) < \epsilon \qquad \text{whenever} \qquad n \geq N$$

- Convergence implicitly depends on the choice of metric for $\mathcal{S}$, so if we want to emphasise that we are using a particular metric, we may say

$$a_n \to a \text{ with respect to the metric } d.$$

Definition

Let $\mathcal{S}$ be a metric space. A sequence of points $\left\{a_n\right\}$ in $\mathcal{S}$ is said to be a Cauchy sequence if for every $\epsilon > 0$ there exists an integer $N > 0$ such that

$$d(a_m, a_n) < \epsilon \qquad \text{whenever} \qquad m, n \geq N$$

Theorem

If $\left\{a_n\right\}$ is a convergent sequence in a metric space $\mathcal{S}$, then

$$\left\{a_n\right\}$$

is a Cauchy sequence in $\mathcal{S}$.

Proof

Let $a_n \to a$ as $n \to \infty$. For any $\epsilon > 0$, there exists an integer $N > 0$ such that

$$d(a, a_n) < \epsilon \qquad \text{whenever} \qquad n \geq N$$

Proof

Consequently, if $m, n > N$,

$$d(a_m, a_n) \le d(a, a_m) + d(a, a_n) < 2\epsilon$$

by the subadditive property of the metric space $\mathcal{S}$, therefore, it is Cauchy.

Q: Is the converse of this theorem true?

- Let $\mathcal{C}[-1,1]$ denote the space of all continuous functions $[-1,1] \to \mathbb{R}$, and

$$d(f,g) = \int_{-1}^{1} \Big| f(x) - g(x) \Big| \, dx$$

be the metric for $\mathcal{C}[-1,1]$. The sequence $\{y_n\}$ defined by

$$y_n(x) = \begin{cases} -1 & \text{if} \quad x \in [-1, -\frac{1}{n}], \\ nx & \text{if} \quad x \in (-\frac{1}{n}, \frac{1}{n}), \\ 1 & \text{if} \quad x \in [\frac{1}{n}, 1]. \end{cases}$$

is Cauchy but not convergent. Since the limit $\lim_{n\to\infty} y_n(x)$ is not continuous.

Q: Recall the connection between the notion of distance and length, is there a natural metric for a given vector space? What is missing in a vector space?

Definition

Let $\mathcal{V}$ be a vector space. A norm on $\mathcal{V}$ is a function

$$\|\cdot\| : \mathcal{V} \to \mathbb{R}$$

such that for all $\mathbf{u}, \mathbf{v} \in \mathcal{V}$, the followings are true:

1. Nonnegative:
$$\|\mathbf{v}\| \geq 0$$
2. Homogeneity:
$$\|\alpha \mathbf{v}\| = |\alpha| \|\mathbf{v}\| \qquad \text{for any scalar } \alpha.$$
3. Subadditive:
$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

A vector space $\mathcal{V}$ together with a norm $\|\cdot\|$ is called a normed vector space.

- Note the function that gives the magnitude/length of a vector $\mathbf{v} \in \mathbb{R}^n$

$$\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

satisfies all three requirements.

- In general, we refer to the number $\|\mathbf{v}\|$ as the length of $\mathbf{v} \in \mathcal{V}$, and

$$\|\mathbf{u} - \mathbf{v}\|$$

as the distance between the vectors $\mathbf{u}$ and $\mathbf{v}$, the metric

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

is called the metric on $\mathcal{V}$ induced from $\|\cdot\|$.

- A metric gives us a notion of the distance between points in a space, a norm gives us a notion of the length of an individual vector. A norm can only be defined on a vector space, while a metric can be defined on arbitrary sets.

- It can be shown for the space of absolutely summable sequence ℓ_1, the sum

$$\|x\|_1 = \sum_{k=1}^{\infty} |x_k|$$

can be used as the norm. It is known as the ℓ_1-norm for the vector space ℓ_1, which is thus a normed space, a metric space as well as being a vector space.

- The ℓ_1-norm can also be defined for other vector spaces, for example, in $\mathbb{R}^n$

$$\|\mathbf{v}\|_1 = \sum_{i=1}^{n} |v_i|$$

- More generally, we could define a ℓ_p-norm, aka p-norm on $\mathbb{R}^n$ by

$$\|\mathbf{v}\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{1/p}, \qquad \text{for any real number } p \geq 1.$$

- In particular, if $p = 2$, then ℓ_p norm is simply the usual length in $\mathbb{R}^n$

$$\|\mathbf{v}\|_2 = \left(\sum_{i=1}^{n} |v_i|^2\right)^{1/2} = \sqrt{\mathbf{v} \cdot \mathbf{v}} \qquad \text{where} \quad \mathbf{v} \in \mathbb{R}^n$$

- Frobenius norm is a norm on a matrix space $\mathbb{R}^{m \times n}$,

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

Q: Can you see why Frobenius norm is clearly a valid norm?

- Note the matrix space $\mathbb{R}^{m \times n}$ is isomorphic to the Euclidean space $\mathbb{R}^{mn}$, and

$$\|\mathbf{A}\|_F = \|\mathbf{s}\|_2 \qquad \text{where} \quad \mathbf{s} = [\mathbf{A}]_{\mathcal{S}} \in \mathbb{R}^{mn}$$

i.e. the coordinate vector of $\mathbf{A}$ with respect to the standard basis of $\mathbb{R}^{m \times n}$, that is, $\mathbf{s}$ is a vector contains all entries of $\mathbf{A}$ according to some fixed order.

- For square matrices $\mathbf{A} \in \mathbb{R}^{n \times n}$, the following two properties are relevant:

Definition

A matrix norm on $\mathbb{R}^{n \times n}$ is said to be compatible with a vector norm on $\mathbb{R}^n$ if

$$\|\mathbf{A}\mathbf{v}\| \leq \|\mathbf{A}\|\|\mathbf{v}\| \qquad \text{for all} \quad \mathbf{A} \in \mathbb{R}^{n \times n}, \mathbf{v} \in \mathbb{R}^n.$$

Definition

A matrix norm on $\mathbb{R}^{n \times n}$ is said to be sub-multiplicative if

$$\|\mathbf{A}\mathbf{B}\| \leq \|\mathbf{A}\|\|\mathbf{B}\| \qquad \text{for all} \quad \mathbf{A}, \mathbf{B} \in \mathbb{R}^{n \times n}.$$

Q: Let $\|\cdot\|$ be a vector norm on $\mathbb{R}^n$, is the following a matrix norm on $\mathbb{R}^{n \times n}$?

$$\|\mathbf{A}\|_o = \max_{\hat{\mathbf{x}}} \|\mathbf{A}\hat{\mathbf{x}}\|$$

Q: Is $\|\cdot\|_o$ compatible with the vector norm $\|\cdot\|$ on $\mathbb{R}^n$? Is it sub-multiplicative?

Definition

The matrix norm $\|\cdot\|_o \colon \mathbb{R}^{n\times n} \to \mathbb{R}$,

$$\|\mathbf{A}\|_o = \max_{\hat{\mathbf{x}}} \|\mathbf{A}\hat{\mathbf{x}}\|$$

is known as the operator norm induced by the vector norm $\|\cdot\|$ on $\mathbb{R}^n$.

Theorem

Let $\mathbf{A}$ be an $n \times n$ matrix with columns $\mathbf{a}_i$ and rows $\mathbf{A}_i$ for $i = 1, 2, \ldots, n$, then

$$\|\mathbf{A}\|_1 = \max_{\hat{\mathbf{x}}} \|\mathbf{A}\hat{\mathbf{x}}\|_1 = \max_j \left\{ \sum_{i=1}^{n} |a_{ij}| \right\}$$

$$\|\mathbf{A}\|_\infty = \max_{\hat{\mathbf{x}}} \|\mathbf{A}\hat{\mathbf{x}}\|_\infty = \max_i \left\{ \sum_{j=1}^{n} |a_{ij}| \right\}$$

Q: Do you remember strictly diagonally dominant and Jacobi iteration?

Definition

A square matrix $\mathbf{A}$ is said to be strictly diagonally dominant if

$$|a_{ii}| > \sum_{j=1}^{i-1} |a_{ij}| + \sum_{j=i+1}^{n} |a_{ij}| \qquad \text{for all} \quad i = 1, 2, \ldots n$$

- Recall Jacobi iteration is a iterative method for solving $\mathbf{Ax} = \mathbf{b}$

$$\mathbf{x}^{(k+1)} = \mathbf{D}^{-1}\Big(\mathbf{b} + (\mathbf{D} - \mathbf{A})\,\mathbf{x}^{(k)}\Big)$$

  where $\mathbf{D}$ is the diagonal matrix such that

$$d_{ij} = \begin{cases} a_{ij} & i = j \\ 0 & i \neq j \end{cases}$$

- Recall such iterative schemes are useful for large sparse systems in practice.

- However, earlier we have only proved the first half of the following theorem:

Theorem

If $\mathbf{A}$ is strictly diagonally dominant, then

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

has a unique solution, and for any choice of the initial guess $\mathbf{x}^{(0)}$, the sequence

$$\left\{\mathbf{x}^{(k)}\right\}$$

produced by the Jacobi or Gauss-Seidel iteration converge to the exact solution.

Proof

- Let us show the Jacobi iteration is convergent for a certain norm on $\mathbb{R}^n$

$$\mathbf{x}^{(k+1)} = \mathbf{D}^{-1}\left(\mathbf{b} + (\mathbf{D} - \mathbf{A})\,\mathbf{x}^{(k)}\right) = \underbrace{\mathbf{D}^{-1}\mathbf{b}}_{\mathbf{c}} + \underbrace{\mathbf{D}^{-1}(\mathbf{D} - \mathbf{A})}_{\mathbf{M}}\mathbf{x}^{(k)}$$

Proof

- Consider the iteration formula in this form at the exaction solution $\mathbf{x}^*$,

$$\begin{aligned}\mathbf{M}\mathbf{x}^* + \mathbf{c} &= \mathbf{D}^{-1}\left(\mathbf{D} - \mathbf{A}\right)\mathbf{x}^* + \mathbf{D}^{-1}\mathbf{b}\\ &= \mathbf{D}^{-1}\mathbf{D}\mathbf{x}^* - \mathbf{D}^{-1}\mathbf{A}\mathbf{x}^* + \mathbf{D}^{-1}\mathbf{b}\\ &= \mathbf{x}^*\end{aligned}$$

- Now if we subtract this identity from the $(k+1)$th iteration formula, we have

$$\mathbf{x}^{(k+1)} - \mathbf{x}^* = \mathbf{M}\mathbf{x}^{(k)} + \mathbf{c} - \left(\mathbf{M}\mathbf{x}^* + \mathbf{c}\right) = \mathbf{M}\left(\mathbf{x}^{(k)} - \mathbf{x}^*\right)$$

  which means the induced distance with respect to any norm on $\mathbb{R}^n$ is

$$\left\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\right\| = \left\|\mathbf{M}\left(\mathbf{x}^{(k)} - \mathbf{x}^*\right)\right\|$$

- Since the operator norm is compatible with the vector norm that induced it,

$$\left\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\right\| = \left\|\mathbf{M}\left(\mathbf{x}^{(k)} - \mathbf{x}^*\right)\right\| \leq \left\|\mathbf{M}\right\|_o \left\|\left(\mathbf{x}^{(k)} - \mathbf{x}^*\right)\right\|$$

Proof

- So if we can show $\|\mathbf{M}\|_o < 1$, then we will have the desired result

$$\lim_{k\to\infty} \left\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\right\| = 0$$

- The inverse of the diagonal matrix is simply the diagonal matrix of $1/a_{ii}$,

$$\begin{aligned}\mathbf{M} = \mathbf{D}^{-1}(\mathbf{D} - \mathbf{A}) &= \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}\\ &= \mathbf{I} - \mathbf{E}_{(1/a_{nn})n} \cdots \mathbf{E}_{(1/a_{22})2}\mathbf{E}_{(1/a_{11})1}\mathbf{A}\end{aligned}$$

  and multiplying a diagonal matrix is equivalent to n type-II operations.

- Hence the matrix has the following form

$$\mathbf{M} = \begin{bmatrix} 0 & -a_{12}/a_{11} & \cdots & -a_{1n}/a_{11} \\ -a_{21}/a_{22} & 0 & \cdots & -a_{2n}/a_{22} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1}/a_{nn} & -a_{n2}/a_{nn} & \cdots & 0 \end{bmatrix}$$

Proof

- Since strictly diagonally dominance

$$|a_{ii}| > \sum_{j=1}^{i-1} |a_{ij}| + \sum_{j=i+1}^{n} |a_{ij}| = \sum_{j \neq i} |a_{ij}|$$

is defined in terms of the absolute values of the entries in the rows, and

$$\|\mathbf{A}\|_\infty = \max_{\hat{\mathbf{x}}} \|\mathbf{A}\hat{\mathbf{x}}\|_\infty = \max_i \left\{ \sum_{j=1}^{n} |a_{ij}| \right\}$$

- Let us consider the ℓ_∞ norm on $\mathbb{R}^n$, and the operator norm induced by it.

$$\|\mathbf{M}\|_\infty = \max_i \left\{ \sum_{j=1}^{n} |m_{ij}| \right\} = \sum_{j \neq q} \left| -\frac{a_{qj}}{a_{qq}} \right| = \frac{1}{|a_{qq}|} \sum_{j \neq q} |a_{qj}| < 1$$

which completes the proof since $\mathbb{R}^n$ is a finite dimensional space.

- For every normed space $\mathcal{V}$, we have the induced metric on $\mathcal{V}$. Therefore all definitions made for metric spaces apply to $\mathcal{V}$, using the induced norm

$$d(\mathbf{u}, \mathbf{v}) = \|\mathbf{u} - \mathbf{v}\|$$

- Specifically, convergence in a normed space is defined by

$$\mathbf{v}_n \to \mathbf{v} \iff \lim_{n \to \infty} \|\mathbf{v} - \mathbf{v}_n\| = 0$$

- Every convergent sequence in a normed vector space must be Cauchy, but the converse does not hold in general. In some normed spaces it is true that every Cauchy sequence in the space is convergent.

Definition

A normed space $\mathcal{V}$ is a Banach space if every Cauchy sequence in $\mathcal{V}$ converges to an element of $\mathcal{V}$. This property is known as complete.