高级工程数学
2021-2022 (1)

沈超敏
计算机科学与技术学院
cmshen@cs.ecnu.edu.cn

教书院 219

Concepts from Geometry

4.1 Line Segments

The line segment between two points $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\mathbb{R}^n$ is the set of points on the straight line joining points $\boldsymbol{x}$ and $\boldsymbol{y}$. Note that if $\boldsymbol{z}$ lies on the line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$, then

$$\boldsymbol{z} - \boldsymbol{y} = \alpha(\boldsymbol{x} - \boldsymbol{y}),$$

where α is a real number from the interval [0,1]. The equation above can be rewritten as $\boldsymbol{z} = \alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y}$. Hence, the line segment between $\boldsymbol{x}$ and $\boldsymbol{y}$ can be represented as

$$\{\alpha\boldsymbol{x} + (1-\alpha)\boldsymbol{y} : \alpha \in [0,1]\}.$$

4.2 Hyperplanes and Linear Varieties

Let $u_1, u_2, ..., u_n \in \mathbb{R}$, where at least one of the u_i is nonzero. The set of all points $\boldsymbol{x} = [x_1, x_2, ..., x_n]^\top$ that satisfy the linear equation

$$u_1 x_1 + u_2 x_2 + \cdots + u_n x_n = v$$

is called *a hyperplane* of the space $\mathbb{R}^n$. i.e.,

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^\top \boldsymbol{x} = v\}.$$

A hyperplane is not necessarily a subspace of $\mathbb{R}^n$ since, in general, it does not contain the origin.

For $n = 2$, the equation of the hyperplane has the form $u_1 x_1 + u_2 x_2 = v$, which is the equation of a straight line. Thus, straight lines are hyperplanes in $\mathbb{R}^2$.

In $\mathbb{R}^3$ (three-dimensional space), hyperplanes are ordinary planes. By translating a hyperplane so that it contains the origin of $\mathbb{R}^n$, it becomes a subspace of $\mathbb{R}^n$. Because the dimension of this subspace is $n - 1$, we say that the hyperplane has dimension $n - 1$.

4.2 Hyperplanes and Linear Varieties

The hyperplane $H = \{\boldsymbol{x} : u_1x_1 + \cdots + u_nx_n = v\}$ divides $\mathbb{R}^n$ into two halfspaces. One of these half-spaces consists of the points satisfying the inequality $u_1x_1 + \cdots + u_nx_n > v$, denoted

$$H_+ = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^\top \boldsymbol{x} > v\}.$$

The other half-space consists of the points satisfying the inequality $u_1x_1 + \cdots + u_nx_n < v$, denoted

$$H_- = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{u}^\top \boldsymbol{x} < v\}.$$

The half-space H_+ is called the positive half-space
The half-space H_- is called the negative half-space.

4.2 Hyperplanes and Linear Varieties

Let $\boldsymbol{a} = [a_1, a_2, \ldots, a_n]^\top$ be an arbitrary point of the hyperplane H. Thus, $\boldsymbol{u}^\top \boldsymbol{a} - v = 0$. We can write

$$\begin{aligned}
\boldsymbol{u}^\top \boldsymbol{x} - v &= \boldsymbol{u}^\top \boldsymbol{x} - v - (\boldsymbol{u}^\top \boldsymbol{a} - v) \\
&= \boldsymbol{u}^\top (\boldsymbol{x} - \boldsymbol{a}) \\
&= u_1(x_1 - a_1) + u_2(x_2 - a_2) + \cdots + u_n(x_n - a_n) = 0.
\end{aligned}$$

The numbers $(x_i - a_i)$, $i = 1, \ldots, n$, are the components of the vector $\boldsymbol{x} - \boldsymbol{a}$. Therefore, the hyperplane H consists of the points $\boldsymbol{x}$ for which $\langle \boldsymbol{u}, \boldsymbol{x} - \boldsymbol{a} \rangle = 0$. In other words, the hyperplane H consists of the points $\boldsymbol{x}$ for which the vectors $\boldsymbol{u}$ and $\boldsymbol{x} - \boldsymbol{a}$ are orthogonal (see Figure 4.3). We call the vector $\boldsymbol{u}$ the *normal* to the hyperplane H. The set H_+ consists of those points $\boldsymbol{x}$ for which $\langle \boldsymbol{u}, \boldsymbol{x} - \boldsymbol{a} \rangle \geq 0$, and H_- consists of those points $\boldsymbol{x}$ for which $\langle \boldsymbol{u}, \boldsymbol{x} - \boldsymbol{a} \rangle \leq 0$.

4.2 Hyperplanes and Linear Varieties

A *linear variety* is a set of the form

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$$

for some matrix $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ and vector $\boldsymbol{b} \in \mathbb{R}^m$. If $\dim \mathcal{N}(\boldsymbol{A}) = r$, we say that the linear variety has dimension r. A linear variety is a subspace if and only if $\boldsymbol{b} = \mathbf{0}$. If $\boldsymbol{A} = \boldsymbol{O}$, the linear variety is $\mathbb{R}^n$. If the dimension of the linear variety is less than n, then it is the intersection of a finite number of hyperplanes.

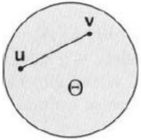
4.3 Convex Sets

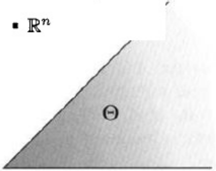
Recall that the line segment between two points $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ is the set $\{\boldsymbol{w} \in \mathbb{R}^n : \boldsymbol{w} = \alpha \boldsymbol{u} + (1-\alpha)\boldsymbol{v}, \alpha \in [0,1]\}$. A point $\boldsymbol{w} = \alpha \boldsymbol{u} + (1-\alpha)\boldsymbol{v}$ (where $\alpha \in [0,1]$) is called a *convex combination* of the points $\boldsymbol{u}$ and $\boldsymbol{v}$.

A set $\Theta \subset \mathbb{R}^n$ is *convex* if for all $\boldsymbol{u}, \boldsymbol{v} \in \Theta$, the line segment between $\boldsymbol{u}$ and $\boldsymbol{v}$ is in Θ. Figure 4.4 gives examples of convex sets, whereas Figure 4.5 gives examples of sets that are not convex. Note that Θ is convex if and only if $\alpha \boldsymbol{u} + (1-\alpha)\boldsymbol{v} \in \Theta$ for all $\boldsymbol{u}, \boldsymbol{v} \in \Theta$ and $\alpha \in (0,1)$.

Examples of convex sets include the following:

- The empty set
- A set consisting of a single point
- A line or a line segment
- A subspace
- A hyperplane
- A linear variety
- A half-space
- $\mathbb{R}^n$





4.3 Convex Sets

Theorem 4.1 *Convex subsets of* $\mathbb{R}^n$ *have the following properties:*

a. *If* Θ *is a convex set and* β *is a real number, then the set*

$$\beta\Theta = \{\boldsymbol{x} : \boldsymbol{x} = \beta\boldsymbol{v}, \boldsymbol{v} \in \Theta\}$$

is also convex.

b. *If* Θ_1 *and* Θ_2 *are convex sets, then the set*

$$\Theta_1 + \Theta_2 = \{\boldsymbol{x} : \boldsymbol{x} = \boldsymbol{v}_1 + \boldsymbol{v}_2, \boldsymbol{v}_1 \in \Theta_1, \boldsymbol{v}_2 \in \Theta_2\}$$

is also convex.

c. *The intersection of any collection of convex sets is convex (see Figure 4.6 for an illustration of this result for two sets).* □

Proof.

a. Let $\beta\boldsymbol{v}_1, \beta\boldsymbol{v}_2 \in \beta\Theta$, where $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \Theta$. Because Θ is convex, we have $\alpha\boldsymbol{v}_1 + (1-\alpha)\boldsymbol{v}_2 \in \Theta$ for any $\alpha \in (0,1)$. Hence,

$$\alpha\beta\boldsymbol{v}_1 + (1-\alpha)\beta\boldsymbol{v}_2 = \beta(\alpha\boldsymbol{v}_1 + (1-\alpha)\boldsymbol{v}_2) \in \beta\Theta,$$

and thus $\beta\Theta$ is convex.

4.3 Convex Sets

A point $\boldsymbol{x}$ in a convex set Θ is said to be an *extreme point* of Θ if there are no two distinct points $\boldsymbol{u}$ and $\boldsymbol{v}$ in Θ such that $\boldsymbol{x} = \alpha\boldsymbol{u} + (1-\alpha)\boldsymbol{v}$ for some $\alpha \in (0,1)$. For example, in Figure 4.4, any point on the boundary of the disk is an extreme point, the vertex (corner) of the set on the right is an extreme point, and the endpoint of the half-line is also an extreme point.

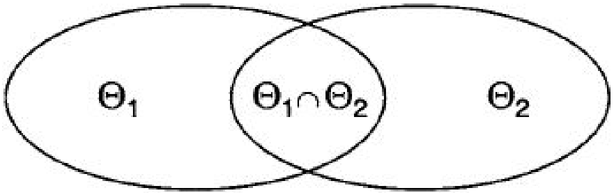


Figure 4.6 Intersection of two convex sets.

Video 20 结束

4.4 Neighborhoods

A *neighborhood* of a point $\boldsymbol{x} \in \mathbb{R}^n$ is the set

$$\{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon\},$$

where ε is some positive number. The neighborhood is also called a *ball* with radius ε and center $\boldsymbol{x}$.

In the plane $\mathbb{R}^2$, a neighborhood of $\boldsymbol{x} = [x_1, x_2]^\top$ consists of all the points inside a disk centered at $\boldsymbol{x}$. In $\mathbb{R}^3$, a neighborhood of $\boldsymbol{x} = [x_1, x_2, x_3]^\top$ consists of all the points inside a sphere centered at $\boldsymbol{x}$ (see Figure 4.7).

4.4 Neighborhoods

A point $\boldsymbol{x} \in S$ is said to be an *interior point* of the set S if the set S contains some neighborhood of $\boldsymbol{x}$; that is, if all points within some neighborhood of $\boldsymbol{x}$ are also in S (see Figure 4.8). The set of all the interior points of S is called the *interior* of S.

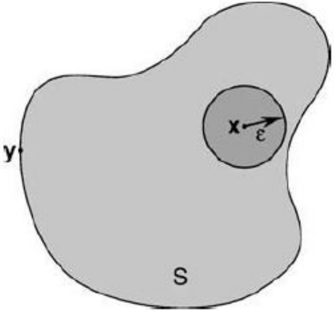


Figure 4.8 $\boldsymbol{x}$ is an interior point; $\boldsymbol{y}$ is a boundary point.

4.4 Neighborhoods

A point x is said to be a *boundary point* of the set S if every neighborhood of x contains a point in S and a point not in S (see Figure 4.8). Note that a boundary point of S may or may not be an element of S. The set of all boundary points of S is called the *boundary* of S.

A set S is said to be *open* if it contains a neighborhood of each of its points; that is, if each of its points is an interior point, or equivalently, if S contains no boundary points.

A set S is said to be *closed* if it contains its boundary (see Figure 4.9). We can show that a set is closed if and only if its complement is open.

A set that is contained in a ball of finite radius is said to be *bounded*. A set is *compact* if it is both closed and bounded. Compact sets are important in optimization problems for the following reason.

Theorem 4.2 ***Theorem of Weierstrass.*** *Let $f : \Omega \to \mathbb{R}$ be a continuous function, where $\Omega \subset \mathbb{R}^n$ is a compact set. Then, there exists a point $\boldsymbol{x}_0 \in \Omega$ such that $f(\boldsymbol{x}_0) \leq f(\boldsymbol{x})$ for all $\boldsymbol{x} \in \Omega$. In other words, f achieves its minimum on Ω.* □

Proof. See [112, p. 89] or [2, p. 154]. ∎

4.5 Polytopes (多面体) and Polyhedra

A set that can be expressed as the intersection of a finite number of half-spaces is called a *convex polytope* (see Figure 4.10).

A nonempty bounded polytope is called a *polyhedron* (see Figure 4.11).

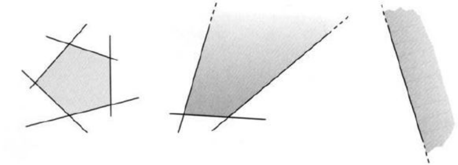


Figure 4.10 Polytopes.

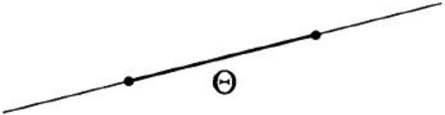


Figure 4.11 One-dimensional polyhedron.

Video 21 结束

第四章内容总结

超平面是 $n-1$ 维的，因为 $x_1, \cdots, x_n$ 是 n 维的，超平面给了一个限制条件，所以有 $n-1$ 个自由度.

超平面法线的理解：法线先记作 n (normal 的意思)，normal 应满足同超平面内任意向量垂直，那如何表示超平面内任何向量呢？超平面内找一个已知点 a，再找任意点 x，则 $x-a$ 可代表超平面内任意向量，slide 中已证 $\langle u, a-x \rangle = 0$，所以 u 即所求的法线.

超平面的交的理解方法，方程每一个都可写成 $a_1x_1 + \ldots + a_nx_n = b_1$，每一个都是超平面，所以解是超平面的交. 那么有几个超平面呢？并不是有几个方程就是几个超平面，因为有重复的

凸集概念在后面带限制条件的优化问题要用到，直观上说，优化迭代时，x_1 跳到 x_2，x_2 跳到 x_3，我至少要保证点都在规定的集合中. 跳的时候还有跳多远 (步长) 的事，如果两点间所有的点都在集合中，可以根据需要，有时跳近点.

Elements of Calculus

5.1 Sequences and Limits

A *sequence of real numbers* is a function whose domain is the set of natural numbers $1, 2, \ldots, k, \ldots$ and whose range is contained in $\mathbb{R}$. Thus, a sequence of real numbers can be viewed as a set of numbers $\{x_1, x_2, \ldots, x_k, \ldots\}$, which is often also denoted as $\{x_k\}$ (or sometimes as $\{x_k\}_{k=1}^{\infty}$, to indicate explicitly the range of values that k can take).

A sequence $\{x_k\}$ is *increasing* if $x_1 < x_2 < \cdots < x_k \cdots$; that is, $x_k < x_{k+1}$ for all k. If $x_k \leq x_{k+1}$, then we say that the sequence is *nondecreasing*. Similarly, we can define *decreasing* and *nonincreasing sequences*. Nonincreasing or nondecreasing sequences are called *monotone sequences*.

A number $x^* \in \mathbb{R}$ is called the *limit* of the sequence $\{x_k\}$ if for any positive ε there is a number K (which may depend on ε) such that for all $k > K$, $|x_k - x^*| < \varepsilon$; that is, x_k lies between $x^* - \varepsilon$ and $x^* + \varepsilon$ for all $k > K$. In this case we write

$$x^* = \lim_{k \to \infty} x_k$$

or

5.1 Sequences and Limits

The notion of a sequence can be extended to sequences with elements in $\mathbb{R}^n$. Specifically, a sequence in $\mathbb{R}^n$ is a function whose domain is the set of natural numbers $1, 2, \ldots, k, \ldots$ and whose range is contained in $\mathbb{R}^n$. We use the notation $\{\boldsymbol{x}^{(1)}, \boldsymbol{x}^{(2)}, \ldots\}$ or $\{\boldsymbol{x}^{(k)}\}$ for sequences in $\mathbb{R}^n$. For limits of sequences in $\mathbb{R}^n$, we need to replace absolute values with vector norms. In other words, $\boldsymbol{x}^*$ is the limit of $\{\boldsymbol{x}^{(k)}\}$ if for any positive ε there is a number K (which may depend on ε) such that for all $k > K$, $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\| < \varepsilon$. As before, if a sequence $\{\boldsymbol{x}^{(k)}\}$ is convergent, we write $\boldsymbol{x}^* = \lim_{k\to\infty} \boldsymbol{x}^{(k)}$ or $\boldsymbol{x}^{(k)} \to \boldsymbol{x}^*$.

5.1 Sequences and Limits

Theorem 5.1 *A convergent sequence has only one limit.* □

Proof. We prove this result by contradiction. Suppose that a sequence $\{\boldsymbol{x}^{(k)}\}$ has two different limits, say $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$. Then, we have $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| > 0$. Let

$$\varepsilon = \frac{1}{2}\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|.$$

From the definition of a limit, there exist K_1 and K_2 such that for $k > K_1$ we have $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_1\| < \varepsilon$, and for $k > K_2$ we have $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_2\| < \varepsilon$. Let $K = \max\{K_1, K_2\}$. Then, if $k > K$, we have $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_1\| < \varepsilon$ and $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_2\| < \varepsilon$. Adding $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_1\| < \varepsilon$ and $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_2\| < \varepsilon$ yields

$$\|\boldsymbol{x}^{(k)} - \boldsymbol{x}_1\| + \|\boldsymbol{x}^{(k)} - \boldsymbol{x}_2\| < 2\varepsilon.$$

Applying the triangle inequality gives

$$\begin{aligned}\|-\boldsymbol{x}_1 + \boldsymbol{x}_2\| &= \|\boldsymbol{x}^{(k)} - \boldsymbol{x}_1 - \boldsymbol{x}^{(k)} + \boldsymbol{x}_2\| \\ &= \|(\boldsymbol{x}^{(k)} - \boldsymbol{x}_1) - (\boldsymbol{x}^{(k)} - \boldsymbol{x}_2)\| \\ &\leq \|\boldsymbol{x}^{(k)} - \boldsymbol{x}_1\| + \|\boldsymbol{x}^{(k)} - \boldsymbol{x}_2\|.\end{aligned}$$

Therefore,

$$\|-\boldsymbol{x}_1 + \boldsymbol{x}_2\| = \|\boldsymbol{x}_1 - \boldsymbol{x}_2\| < 2\varepsilon.$$

However, this contradicts the assumption that $\|\boldsymbol{x}_1 - \boldsymbol{x}_2\| = 2\varepsilon$, which completes the proof. ∎

Video 22 结束

5.1 Sequences and Limits

A sequence $\{\boldsymbol{x}^{(k)}\}$ in $\mathbb{R}^n$ is *bounded* if there exists a number $B \geq 0$ such that $\|\boldsymbol{x}^{(k)}\| \leq B$ for all $k = 1, 2, \ldots$.

Theorem 5.2 *Every convergent sequence is bounded.* □

Proof. Let $\{\boldsymbol{x}^{(k)}\}$ be a convergent sequence with limit $\boldsymbol{x}^*$. Choose $\varepsilon = 1$. Then, by definition of the limit, there exists a natural number K such that for all $k > K$,

$$\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\| < 1.$$

By the result of Exercise 2.9, we get

$$\|\boldsymbol{x}^{(k)}\| - \|\boldsymbol{x}^*\| \leq \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\| < 1 \quad \text{for all } k > K.$$

Therefore,

$$\|\boldsymbol{x}^{(k)}\| < \|\boldsymbol{x}^*\| + 1 \quad \text{for all } k > K.$$

Letting

$$B = \max\left\{\|\boldsymbol{x}^{(1)}\|, \|\boldsymbol{x}^{(2)}\|, \ldots, \|\boldsymbol{x}^{(K)}\|, \|\boldsymbol{x}^*\| + 1\right\},$$

we have

$$B \geq \|\boldsymbol{x}^{(k)}\| \quad \text{for all } k,$$

which means that the sequence $\{\boldsymbol{x}^{(k)}\}$ is bounded. ∎

5.1 Sequences and Limits

For a sequence $\{x_k\}$ in $\mathbb{R}$, a number B is called an *upper bound* if $x_k \leq B$ for all $k = 1, 2, \ldots$. In this case, we say that $\{x_k\}$ is *bounded above*. Similarly, B is called a *lower bound* if $x_k \geq B$ for all $k = 1, 2, \ldots$. In this case, we say that $\{x_k\}$ is *bounded below*. Clearly, a sequence is bounded if it is both bounded above and bounded below.

Any sequence $\{x_k\}$ in $\mathbb{R}$ that has an upper bound has a *least upper bound* (also called the *supremum*), which is the smallest number B that is an upper bound of $\{x_k\}$. Similarly, any sequence $\{x_k\}$ in $\mathbb{R}$ that has a lower bound has a *greatest lower bound* (also called the *infimum*). If B is the least upper bound of the sequence $\{x_k\}$, then $x_k \leq B$ for all k, and for any $\varepsilon > 0$, there exists a number K such that $x_K > B - \varepsilon$. An analogous statement applies to the greatest lower bound: If B is the greatest lower bound of $\{x_k\}$, then $x_k \geq B$ for all k, and for any $\varepsilon > 0$, there exists a number K such that $x_K < B + \varepsilon$.

Video 23 结束

5.1 Sequences and Limits

Theorem 5.3 *Every monotone bounded sequence in $\mathbb{R}$ is convergent.* □

Proof. We prove the theorem for nondecreasing sequences. The proof for nonincreasing sequences is analogous.

Let $\{x_k\}$ be a bounded nondecreasing sequence in $\mathbb{R}$ and x^* the least upper bound. Fix a number $\varepsilon > 0$. Then, there exists a number K such that $x_K > x^* - \varepsilon$. Because $\{x_k\}$ is nondecreasing, for any $k \geq K$,

$$x_k \geq x_K > x^* - \varepsilon.$$

Also, because x^* is an upper bound of $\{x_k\}$, we have

$$x_k \leq x^* < x^* + \varepsilon.$$

Therefore, for any $k \geq K$,

$$|x_k - x^*| < \varepsilon,$$

which means that $x_k \to x^*$. ■

Video 24 结束

5.1 Sequences and Limits

Theorem 5.4 *Consider a convergent sequence $\{\boldsymbol{x}^{(k)}\}$ with limit $\boldsymbol{x}^*$. Then, any subsequence of $\{\boldsymbol{x}^{(k)}\}$ also converges to $\boldsymbol{x}^*$.* □

Proof. Let $\{\boldsymbol{x}^{(m_k)}\}$ be a subsequence of $\{\boldsymbol{x}^{(k)}\}$, where $\{m_k\}$ is an increasing sequence of natural numbers. Observe that $m_k \geq k$ for all $k = 1, 2, \ldots$. To show this, first note that $m_1 \geq 1$ because m_1 is a natural number. Next, we proceed by induction by assuming that $m_k \geq k$. Then, we have $m_{k+1} > m_k \geq k$, which implies that $m_{k+1} \geq k+1$. Therefore, we have shown that $m_k \geq k$ for all $k = 1, 2, \ldots$.

Let $\varepsilon > 0$ be given. Then, by definition of the limit, there exists K such that $\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\| < \varepsilon$ for any $k > K$. Because $m_k \geq k$, we also have $\|\boldsymbol{x}^{(m_k)} - \boldsymbol{x}^*\| < \varepsilon$ for any $k > K$. This means that

$$\lim_{k \to \infty} \boldsymbol{x}^{(m_k)} = \boldsymbol{x}^*.$$

■

It turns out that any bounded sequence contains a convergent subsequence. This result is called the *Bolzano-Weierstrass theorem* (see [2, p. 70]).

5.1 Sequences and Limits

Consider a function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ and a point $\boldsymbol{x}_0 \in \mathbb{R}^n$. Suppose that there exists $\boldsymbol{f}^*$ such that for any convergent sequence $\{\boldsymbol{x}^{(k)}\}$ with limit $\boldsymbol{x}_0$, we have

$$\lim_{k\to\infty} \boldsymbol{f}(\boldsymbol{x}^{(k)}) = \boldsymbol{f}^*.$$

Then, we use the notation

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \boldsymbol{f}(\boldsymbol{x})$$

to represent the limit $\boldsymbol{f}^*$.

It turns out that $\boldsymbol{f}$ is continuous at $\boldsymbol{x}_0$ if and only if for any convergent sequence $\{\boldsymbol{x}^{(k)}\}$ with limit $\boldsymbol{x}_0$, we have

$$\lim_{k\to\infty} \boldsymbol{f}(\boldsymbol{x}^{(k)}) = \boldsymbol{f}\left(\lim_{k\to\infty} \boldsymbol{x}^{(k)}\right) = \boldsymbol{f}(\boldsymbol{x}_0)$$

(see [2, p. 137]). Therefore, using the notation introduced above, the function $\boldsymbol{f}$ is continuous at $\boldsymbol{x}_0$ if and only if

$$\lim_{\boldsymbol{x}\to\boldsymbol{x}_0} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}_0).$$

5.1 Sequences and Limits

矩阵的极限

We say that a sequence $\{\boldsymbol{A}_k\}$ of $m \times n$ matrices converges to the $m \times n$ matrix $\boldsymbol{A}$ if

$$\lim_{k \to \infty} \|\boldsymbol{A} - \boldsymbol{A}_k\| = 0.$$

Lemma 5.1 *Let $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. Then, $\lim_{k \to \infty} \boldsymbol{A}^k = \boldsymbol{O}$ if and only if the eigenvalues of $\boldsymbol{A}$ satisfy $|\lambda_i(\boldsymbol{A})| < 1$, $i = 1, \ldots, n$.* □

Proof. To prove this theorem, we use the *Jordan form* (see, e.g., [47]). Specifically, it is well known that any square matrix is similar to the Jordan form: There exists a nonsingular $\boldsymbol{T}$ such that

$$\boldsymbol{T}\boldsymbol{A}\boldsymbol{T}^{-1} = \operatorname{diag}\left[\boldsymbol{J}_{m_1}(\lambda_1), \ldots, \boldsymbol{J}_{m_s}(\lambda_1), \boldsymbol{J}_{n_1}(\lambda_2), \ldots, \boldsymbol{J}_{t_v}(\lambda_q)\right] \triangleq \boldsymbol{J},$$

Video 25 结束

5.2 Differentiability

Differential calculus is based on the idea of approximating an arbitrary function by an *affine function*. A function $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^m$ is *affine* if there exists a *linear function* $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^m$ and a vector $\boldsymbol{y} \in \mathbb{R}^m$ such that

$$\mathcal{A}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}) + \boldsymbol{y}$$

for every $\boldsymbol{x} \in \mathbb{R}^n$. Consider a function $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ and a point $\boldsymbol{x}_0 \in \mathbb{R}^n$. We wish to find an affine function $\mathcal{A}$ that approximates $\boldsymbol{f}$ near the point $\boldsymbol{x}_0$. First, it is natural to impose the condition

$$\mathcal{A}(\boldsymbol{x}_0) = \boldsymbol{f}(\boldsymbol{x}_0).$$

Because $\mathcal{A}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}) + \boldsymbol{y}$, we obtain $\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}_0) - \mathcal{L}(\boldsymbol{x}_0)$. By the linearity of $\mathcal{L}$,

$$\mathcal{L}(\boldsymbol{x}) + \boldsymbol{y} = \mathcal{L}(\boldsymbol{x}) - \mathcal{L}(\boldsymbol{x}_0) + \boldsymbol{f}(\boldsymbol{x}_0) = \mathcal{L}(\boldsymbol{x} - \boldsymbol{x}_0) + \boldsymbol{f}(\boldsymbol{x}_0).$$

Hence, we may write

$$\mathcal{A}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x} - \boldsymbol{x}_0) + \boldsymbol{f}(\boldsymbol{x}_0).$$

Next, we require that $\mathcal{A}(\boldsymbol{x})$ approaches $\boldsymbol{f}(\boldsymbol{x})$ faster than $\boldsymbol{x}$ approaches $\boldsymbol{x}_0$; that is,

$$\lim_{\boldsymbol{x} \to \boldsymbol{x}_0, \boldsymbol{x} \in \Omega} \frac{\|\boldsymbol{f}(\boldsymbol{x}) - \mathcal{A}(\boldsymbol{x})\|}{\|\boldsymbol{x} - \boldsymbol{x}_0\|} = 0.$$

The conditions above on $\mathcal{A}$ ensure that $\mathcal{A}$ approximates $\boldsymbol{f}$ near $\boldsymbol{x}_0$ in the sense that the error in the approximation at a given point is "small" compared with the distance of the point from $\boldsymbol{x}_0$.

5.3 The Derivative Matrix

Any linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$, and in particular the derivative $\mathcal{L}$ of $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$, can be represented by an $m \times n$ matrix. To find the matrix representation $\boldsymbol{L}$ of the derivative $\mathcal{L}$ of a differentiable function $\boldsymbol{f}: \mathbb{R}^n \to \mathbb{R}^m$, we use the natural basis $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n\}$ for $\mathbb{R}^n$. Consider the vectors

$$\boldsymbol{x}_j = \boldsymbol{x}_0 + t\boldsymbol{e}_j, \quad j = 1, \ldots, n.$$

By the definition of the derivative, we have

$$\lim_{t \to 0} \frac{\boldsymbol{f}(\boldsymbol{x}_j) - (t\boldsymbol{L}\boldsymbol{e}_j + \boldsymbol{f}(\boldsymbol{x}_0))}{t} = \boldsymbol{0}$$

for $j = 1, \ldots, n$. This means that

$$\lim_{t \to 0} \frac{\boldsymbol{f}(\boldsymbol{x}_j) - \boldsymbol{f}(\boldsymbol{x}_0)}{t} = \boldsymbol{L}\boldsymbol{e}_j$$

for $j = 1, \ldots, n$. But $\boldsymbol{L}\boldsymbol{e}_j$ is the jth column of the matrix $\boldsymbol{L}$. On the other hand, the vector $\boldsymbol{x}_j$ differs from $\boldsymbol{x}_0$ only in the jth coordinate, and in that coordinate the difference is just the number t. Therefore, the left side of the preceding equation is the partial derivative

$$\frac{\partial \boldsymbol{f}}{\partial x_j}(\boldsymbol{x}_0).$$

Because vector limits are computed by taking the limit of each coordinate function, it follows that if

$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} f_1(\boldsymbol{x}) \\ \vdots \\ f_m(\boldsymbol{x}) \end{bmatrix},$$

then

$$\frac{\partial \boldsymbol{f}}{\partial x_j}(\boldsymbol{x}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_j}(\boldsymbol{x}_0) \\ \vdots \\ \frac{\partial f_m}{\partial x_j}(\boldsymbol{x}_0) \end{bmatrix},$$

5.3 The Derivative Matrix

Theorem 5.5 *If a function* $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ *is differentiable at* $\boldsymbol{x}_0$*, then the derivative of* $\boldsymbol{f}$ *at* $\boldsymbol{x}_0$ *is determined uniquely and is represented by the* $m \times n$ *derivative matrix* $D\boldsymbol{f}(\boldsymbol{x}_0)$*. The best affine approximation to* $\boldsymbol{f}$ *near* $\boldsymbol{x}_0$ *is then given by*

$$\mathcal{A}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x}_0) + D\boldsymbol{f}(\boldsymbol{x}_0)(\boldsymbol{x} - \boldsymbol{x}_0),$$

in the sense that

$$\boldsymbol{f}(\boldsymbol{x}) = \mathcal{A}(\boldsymbol{x}) + \boldsymbol{r}(\boldsymbol{x})$$

and $\lim_{\boldsymbol{x} \to \boldsymbol{x}_0} \|\boldsymbol{r}(\boldsymbol{x})\| / \|\boldsymbol{x} - \boldsymbol{x}_0\| = 0$*. The columns of the derivative matrix* $D\boldsymbol{f}(\boldsymbol{x}_0)$ *are vector partial derivatives. The vector*

$$\frac{\partial \boldsymbol{f}}{\partial x_j}(\boldsymbol{x}_0)$$

is a tangent vector at $\boldsymbol{x}_0$ *to the curve* $\boldsymbol{f}$ *obtained by varying only the* j*th coordinate of* $\boldsymbol{x}$*.* $\square$

5.3 The Derivative Matrix

If $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then the function ∇f defined by

$$\nabla f(\boldsymbol{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\boldsymbol{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\boldsymbol{x}) \end{bmatrix} = Df(\boldsymbol{x})^\top$$

is called the *gradient* of f. The gradient is a function from $\mathbb{R}^n$ to $\mathbb{R}^n$, and can be pictured as a *vector field*, by drawing the arrow representing $\nabla f(\boldsymbol{x})$ so that its tail starts at $\boldsymbol{x}$.

Given $f : \mathbb{R}^n \to \mathbb{R}$, if ∇f is differentiable, we say that f is *twice differentiable*, and we write the derivative of ∇f as

$$D^2 f = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_2 \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_1} \\ \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_n \partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_n} & \frac{\partial^2 f}{\partial x_2 \partial x_n} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}.$$

(The notation $\frac{\partial^2 f}{\partial x_i \partial x_j}$ represents taking the partial derivative of f with respect to x_j first, then with respect to x_i.) The matrix $D^2 f(\boldsymbol{x})$ is called the *Hessian matrix* of f at $\boldsymbol{x}$, and is often also denoted $\boldsymbol{F}(\boldsymbol{x})$.

5.3 The Derivative Matrix

Note that the Hessian matrix of a function $f : \mathbb{R}^n \to \mathbb{R}$ at $\boldsymbol{x}$ is symmetric if f is twice continuously differentiable at $\boldsymbol{x}$. This is a well-known result from calculus called *Clairaut's theorem* or *Schwarz's theorem*. However, if the second partial derivatives of f are not continuous, then there is no guarantee that the Hessian is symmetric, as shown in the following well-known example.

Example 5.1 Consider the function

$$f(\boldsymbol{x}) = \begin{cases} x_1 x_2 (x_1^2 - x_2^2)/(x_1^2 + x_2^2) & \text{if } \boldsymbol{x} \neq \boldsymbol{0} \\ 0 & \text{if } \boldsymbol{x} = \boldsymbol{0}. \end{cases}$$