第二章 基础知识

修贤超

https://xianchaoxiu.github.io

目录

向量范数的定义

- 令记号 $\|\cdot\|: \mathbb{R}^n \to \mathbb{R}^+$ 是一种非负函数, 如果它满足
  - **正定性** 对于 $\forall v \in \mathbb{R}^n$, 有 $\|v\| \geqslant 0$, 且 $\|v\| = 0 \to v = 0_{n\times 1}$
  - **齐次性** 对于 $\forall v \in \mathbb{R}^n$ 和 $\alpha \in \mathbb{R}$, 有 $\|\alpha v\| = |\alpha|\|v\|$
  - **三角不等式** 对于 $\forall v, w \in \mathbb{R}^n$, 均成立 $\|v + w\| \leqslant \|v\| + \|w\|$

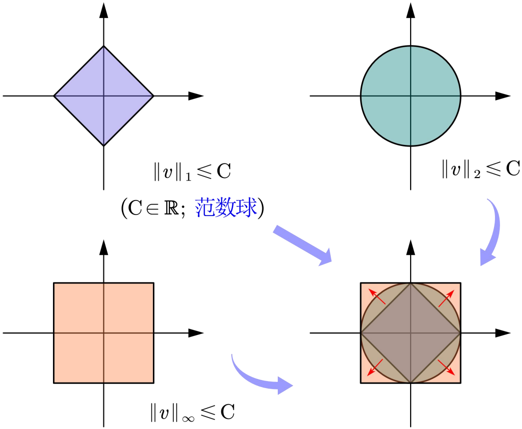
  则称 $\|\cdot\|$ 是定义在向量空间 $\mathbb{R}^n$ 上的**向量范数**

- 最常用的向量范数

$$\|v\|_p = (\sum_{i=1}^{n} |v_i|^p)^{\frac{1}{p}}, \quad \|v\|_\infty = \max_{1\leqslant j\leqslant n} |v_j|$$

向量范数的定义

- 不同范数所度量的距离分别具有怎样的特征



矩阵范数

- ℓ_1 范数 $\|A\|_1 = \sum_{i,j} |A_{ij}|$
- Frobenius 范数 $\|A\|_F = \sqrt{\sum_{i,j} A_{ij}^2} = \sqrt{\mathrm{Tr}(AA^\top)}$
- 算子范数是一类特殊的矩阵范数，由向量范数诱导得到

$$\|A\|_{(m,n)} = \max_{x\in\mathbb{R}^n, \|x\|_{(n)}=1} \|Ax\|_{(m)}$$

  - $p=1$ 时，$\|A\|_{p=1} = \max_{\|x\|_1=1} \|Ax\|_1 = \max_{1\leqslant j\leqslant n} \sum_{i=1}^m |a_{ij}|$
  - $p=2$ 时，$\|A\|_{p=2} = \max_{\|x\|_2=1} \|Ax\|_2 = \sqrt{\lambda_{\max}(A^\top A)}$，又称为 A 的谱范数
  - $p=\infty$ 时，$\|A\|_{p=\infty} = \max_{\|x\|_\infty=1} \|Ax\|_\infty = \max_{1\leqslant i\leqslant m} \sum_{j=1}^n |a_{ij}|$

矩阵范数

- 核范数

$$\|A\|_* = \sum_{i=1}^{r} \sigma_i$$

- 矩阵内积

$$< A, B >= \mathrm{Tr}(AB^\top) = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} b_{ij}$$

- 设 $A, B \in \mathbb{R}^{m \times n}$, 则

$$| < A, B > | \leqslant \|A\|_F \|B\|_F$$

  等号成立当且仅当 A 和 B 线性相关, 即柯西不等式

- 同一矩阵空间内, 矩阵范数彼此之间是相互等价的

目录

梯度

- 给定函数 $f:\mathbb{R}^n\to\mathbb{R}$，且 f 在点 x 的一个邻域内有意义，若存在向量 $g\in\mathbb{R}^n$ 满足

$$\lim_{p\to 0}\frac{f(x+p)-f(x)-\langle g,p\rangle}{\|p\|}=0$$

  其中 $\|\cdot\|$ 是任意的向量范数，就称 f 在点 x 处可微（或 Fréchet 可微），g 为 f 在点 x 处的梯度，记作

$$\nabla f(x)=[\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n}]^\top$$

- 如果对区域 D 上的每一个点 x 都有 $\nabla f(x)$ 存在，则称 f 在 D 上可微

海瑟矩阵

- 如果函数 $f(x):\mathbb{R}^n\to\mathbb{R}$ 在点 x 处的二阶偏导数 $\frac{\partial^2 f(x)}{\partial x_i \partial x_j}$ $i,j=1,2,\cdots,n$ 都存在，则 f 在点 x 处的海瑟矩阵为

$$\nabla^2 f(x)=\begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_3} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \frac{\partial^2 f(x)}{\partial x_2 \partial x_3} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \frac{\partial^2 f(x)}{\partial x_n \partial x_3} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}$$

- 当 $\nabla^2 f(x)$ 在区域 D 上的每个点 x 处都存在时，称 f 在 D 上二阶可微. 若 $\nabla^2 f(x)$ 在 D 上还连续，则称 f 在 D 上二阶连续可微
- 海瑟矩阵是一个对称矩阵

矩阵变量函数的导数

- 对于以 $m \times n$ 矩阵 X 为自变量的函数 $f(X)$，若存在矩阵 $G \in \mathbb{R}^{m\times n}$ 满足

$$\lim_{V\to 0} \frac{f(X+V) - f(X) - \langle G, V\rangle}{\|V\|} = 0$$

其中 $\|\cdot\|$ 是任意矩阵范数，就称矩阵变量函数 f 在 X 处 Fréchet 可微，G 为 f 在 Fréchet 可微意义下的梯度，记为

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}} & \cdots & \frac{\partial f}{\partial x_{1n}} \\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}} & \cdots & \frac{\partial f}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f}{\partial x_{m1}} & \frac{\partial f}{\partial x_{m2}} & \cdots & \frac{\partial f}{\partial x_{mn}} \end{bmatrix}$$

矩阵变量函数的导数

- 设 $f(X)$ 为矩阵变量函数，如果对任意方向 $V \in \mathbb{R}^{m\times n}$，存在矩阵 $G \in \mathbb{R}^{m\times n}$ 满足

$$\lim_{V\to 0} \frac{f(X+V) - f(X) - \langle G, V\rangle}{\|V\|} = 0$$

$$\Downarrow$$

$$\lim_{t\to 0} \frac{f(X+tV) - f(X) - t\langle G, V\rangle}{t} = 0$$

则称 f 关于 X *Gâteaux* 可微，G 为 f 在 X 处 *Gâteaux* 可微意义下的梯度

- 当 f 是 Fréchet 可微函数时，f 也是 Gâteaux 可微的，且梯度相等

例子

- 线性函数 $f(X) = \mathrm{Tr}(AX^\top B)$

$$\begin{aligned}\lim_{t\to 0}\frac{f(X+tV)-f(X)}{t} &= \lim_{t\to 0}\frac{\mathrm{Tr}(A(X+tV)^\top B)-\mathrm{Tr}(AX^\top B)}{t}\\ &= \mathrm{Tr}(AV^\top B) = \langle BA, V\rangle\end{aligned}$$

$$\Rightarrow \quad \nabla f(X) = BA$$

- 二次函数 $f(X,Y) = \frac{1}{2}\|XY-A\|_F^2$

$$\begin{aligned}f(X,Y+tV)-f(X,Y) &= \frac{1}{2}\|X(Y+tV)-A\|_F^2 - \frac{1}{2}\|XY-A\|_F^2\\ &= \langle tXV, XY-A\rangle + \frac{1}{2}t^2\|XV\|_F^2\\ &= t\langle V, X^\top(XY-A)\rangle + \mathcal{O}(t^2)\end{aligned}$$

$$\Rightarrow \quad \frac{\partial f}{\partial Y} = X^\top(XY-A), \quad \frac{\partial f}{\partial X} = (XY-A)Y^\top$$

目录

- 2.1 范数
- 2.2 导数
- 2.3 广义实值函数
- 2.4 凸集
- 2.5 凸函数
- 2.6 共轭函数
- 2.7 次梯度

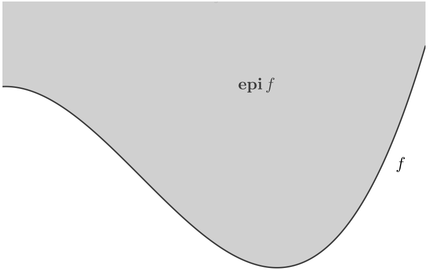
广义实值函数与适当函数

- 令$\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ 为广义实数空间，则映射 $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ 称为广义实值函数
- 给定广义实值函数 f 和非空集合 $\mathcal{X}$，如果存在 $x \in \mathcal{X}$ 使得 $f(x) < +\infty$，且对任意的 $x \in \mathcal{X}$ 都有 $f(x) > -\infty$，则称 f 是关于集合 $\mathcal{X}$ 的适当函数
  - 至少有一处取值不为正无穷
  - 处处取值不为负无穷
- 对于适当函数 f，规定其定义域

$$\mathrm{dom}\ f = \{x \mid f(x) < +\infty\}$$

闭函数

- 设f 为广义实值函数，称 $C_\alpha = \{x \mid f(x) \le \alpha\}$ 为α-下水平集
- 设 f 为广义实值函数，称 $\mathrm{epi}\ f = \{(x,t) \in \mathbb{R}^{n+1} \mid f(x) \le t\}$ 为上方图
- 设 f 为广义实值函数，若 $\mathrm{epi}\ f$ 为闭集，则称 f 为闭函数

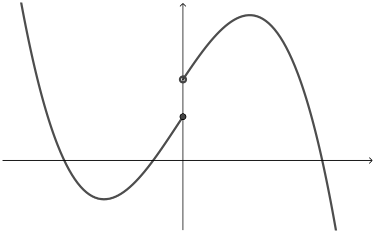


下半连续函数

- 设 f 为广义实值函数，若对任意的 $x \in \mathbb{R}^n$，有

$$\liminf_{y \to x} f(y) \geq f(x)$$

则 $f(x)$ 为**下半连续函数**



闭函数与下半连续函数

- 设广义实值函数 $f:\mathbb{R}^n \to \overline{\mathbb{R}}$，则以下命题等价
  - $f(x)$ 的任意 α-下水平集都是闭集
  - $f(x)$ 是下半连续的
  - $f(x)$ 是闭函数
- 闭（下半连续）函数的性质
  - 若 f 与 g 均为适当的闭（下半连续）函数，并且 $\mathrm{dom}\ f \cap \mathrm{dom}\ g \neq \emptyset$，则 $f+g$ 也是闭（下半连续）函数
  - 若 f 为闭（下半连续）函数, 则 $f(Ax+b)$ 也为闭（下半连续）函数
  - 若每一个函数 f_α 均为闭（下半连续）函数，则 $\sup_\alpha f_\alpha(x)$ 也为闭（下半连续）函数

目录

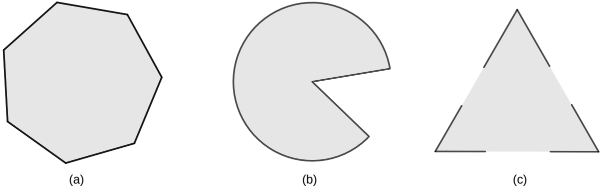
凸集的几何定义

- 若过集合 $\mathcal{C}$ 中的任意两点的直线都在 $\mathcal{C}$ 内, 则称 $\mathcal{C}$ 为仿射集, 即

$$x_1, x_2 \in \mathcal{C} \Rightarrow \theta x_1 + (1-\theta)x_2 \in \mathcal{C}, \forall \theta \in \mathbb{R}$$

- 若连接集合 $\mathcal{C}$ 中的任意两点的线段都在 $\mathcal{C}$ 内, 则称 $\mathcal{C}$ 为凸集, 即

$$x_1, x_2 \in \mathcal{C} \Rightarrow \theta x_1 + (1-\theta)x_2 \in \mathcal{C}, \forall 0 \leqslant \theta \leqslant 1$$



凸集的性质

- 若$\mathcal{S}$ 是凸集, 则 $k\mathcal{S} = \{ks \mid k \in \mathbb{R}, s \in \mathcal{S}\}$ 是凸集
- 若 $\mathcal{S}$ 和 $\mathcal{T}$ 均是凸集, 则 $\mathcal{S} + \mathcal{T} = \{s + t \mid s \in \mathcal{S}, t \in \mathcal{T}\}$ 是凸集
- 若 $\mathcal{S}$ 和 $\mathcal{T}$ 均是凸集, 则 $\mathcal{S} \cap \mathcal{T}$ 是凸集

  证明 设 $x, y \in \mathcal{S} \cap \mathcal{T}$ 且 $\theta \in [0, 1]$. 由于 $\mathcal{S}$ 和 $\mathcal{T}$ 均为凸集, 则

$$\theta x + (1 - \theta) y \in \mathcal{S} \cap \mathcal{T}$$

- 凸集的内部和闭包都是凸集

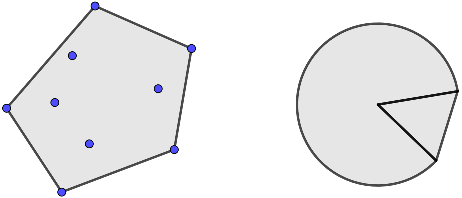
凸组合和凸包

- 形如

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$
$$\theta_1 + \cdots + \theta_k = 1, \theta_i \geqslant 0, i = 1, \cdots, k$$

  的点称为 $x_1, \cdots, x_k$ 的凸组合

- 集合 $\mathcal{S}$ 的所有点的凸组合构成的点集为 $\mathcal{S}$ 的凸包, 记为 $\mathbf{conv}\mathcal{S}$



- $\mathbf{conv}\mathcal{S}$ 是包含 $\mathcal{S}$ 的最小凸集

仿射组合和仿射包

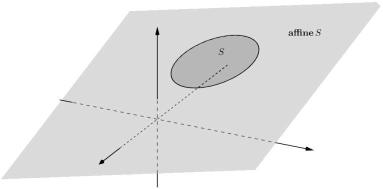
- 形如

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k$$

$$\theta_1 + \cdots + \theta_k = 1, \theta_i \in \mathbb{R}, i = 1, \cdots, k$$

的点称为 $x_1, \cdots, x_k$ 的仿射组合

- 集合 $\mathcal{S}$ 的所有点的仿射组合构成的点集为 $\mathcal{S}$ 的仿射包, 记为 affine$\mathcal{S}$



- affine$\mathcal{S}$ 是包含 $\mathcal{S}$ 的最小仿射集

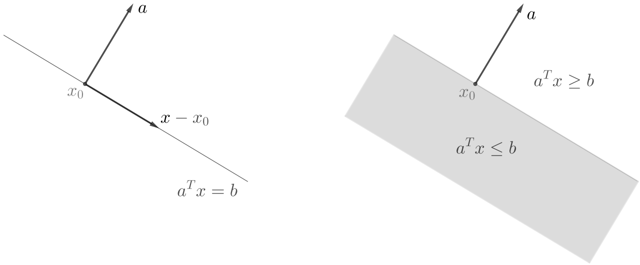
锥组合和凸锥

- 形如
$$x=\theta_1 x_1+\cdots+\theta_k x_k, \theta_i>0(i=1,\cdots,k)$$
的点称为 $x_1,\cdots,x_k$ 的**锥组合**
- 若集合 $\mathcal{S}$ 中任意点的锥组合都在 $\mathcal{S}$ 中, 则称 $\mathcal{S}$ 为**凸锥**

- 锥组合不要求系数的和为 1, 因此一般而言锥组合都是开放的

超平面和半空间

- 任取非零向量 $a \in \mathbb{R}^n$, 称 $\{x \mid a^\top x = b\}$ 为超平面, $\{x \mid a^\top x \leqslant b\}$ 为半空间
- 满足线性等式和不等式组的点的集合 $\{x \mid Ax \leqslant b, Cx = d\}$ 称为多面体



- 超平面是仿射集和凸集，半空间是凸集但不是仿射集，多面体是有限个半空间和超平面的交

范数球和椭球

- 设空间中到某一定点 x_c 的距离小于等于定值 r 的点的集合为(范数) 球, 即

$$\mathrm{B}(x_c, r) = \{x \mid \|x - x_c\| \leqslant r\} = \{x_c + ru \mid \|u\| \leqslant 1\}$$

- 设形如

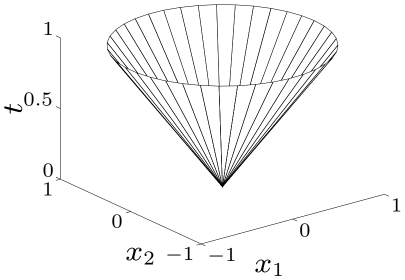
$$\{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leqslant 1\} = \{x_c + Au \mid \|u\|_2 \leqslant 1\}$$

  的集合为椭球, 其中 x_c 为椭球中心, P 对称正定, 且 A 非奇异
- 球和椭球的范围取决于 x 的范围

范数锥

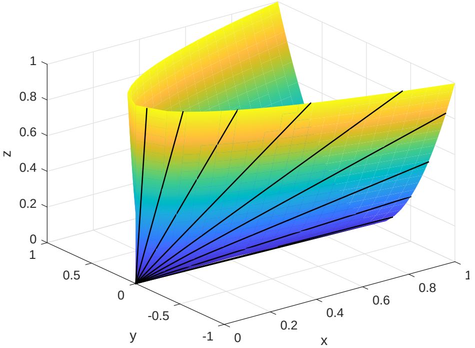
- 形如
$$\{(x,t) \mid \|x\| \leqslant t\}$$
的集合为**范数锥**

- 使用 $\|\cdot\|_2$ 度量距离的锥为二次锥，也称冰淇淋锥



(半) 正定锥

- 记$\mathcal{S}^n$ 为对称矩阵的集合, 即 $\mathcal{S}^n = \{X \in \mathbb{R}^{n\times n} \mid X^\top = X\}$
- 记 $\mathcal{S}^n_+$ 为半正定矩阵的集合, 即 $\mathcal{S}^n_+ = \{X \in \mathcal{S}^n \mid X \succeq 0\}$
- 记 $\mathcal{S}^n_{++}$ 为正定矩阵的集合, 即 $\mathcal{S}^n_{++} = \{X \in \mathcal{S}^n \mid X \succ 0\}$



对于矩阵 $\begin{pmatrix} x & y \\ y & z \end{pmatrix}$, 其特征值应全部大于等于 0

$$\Downarrow$$

$$\{(x, y, z) \mid x \geqslant 0, z \geqslant 0, xz \geqslant y^2\}$$

仿射变换的保凸性

- 设$f:\mathbb{R}^n \to \mathbb{R}^m$ 是仿射变换, 即 $f(x) = Ax + b,\ A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$, 则

$$S \subseteq \mathbb{R}^n\text{是凸集} \Rightarrow f(\mathcal{S}) = \{f(x) \mid x \in \mathcal{S}\}\text{是凸集}$$

$$\mathcal{C} \subseteq \mathbb{R}^m\text{是凸集} \Rightarrow f^{-1}(\mathcal{C}) = \{x \mid f(x) \in \mathcal{C}\}\text{是凸集}$$

- 线性矩阵不等式的解集 $\{x \mid x_1 A_1 + \cdots + x_m A_m \preceq B\}$ 是凸集
- 双曲锥 $\{x \mid x^\top P x \leqslant (c^\top x)^2, c^\top x \geqslant 0, P \in \mathcal{S}_+^n\}$ 是凸集

证明 双曲锥可以转化为二阶锥

$$\{x \mid \|Ax\|_2 \leqslant c^\top x, c^\top x \geqslant 0, A^\top A = P\}$$

而二阶锥可由二次锥 $\{(x,t) \mid \|x\|_2 \leqslant t, t \geqslant 0\}$ 经过仿射变换得到

分离超平面定理

- 如果 $\mathcal{C}$ 和 $\mathcal{D}$ 是不相交的凸集, 则存在非零向量 a 和常数 b, 使得

  $$a^\top x \leqslant b, \forall x \in \mathcal{C} \quad 且 \quad a^\top x \geqslant b, \forall x \in \mathcal{D}$$

  即超平面 $\{x \mid a^\top x = b\}$ 分离了 $\mathcal{C}$ 和 $\mathcal{D}$

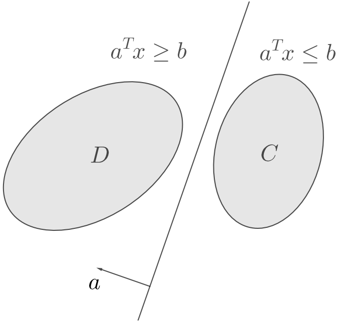
- 如果 $\mathcal{C}$ 和 $\mathcal{D}$ 是不相交的凸集, 且 $\mathcal{C}$ 是闭集, $\mathcal{D}$ 是紧集, 则存在非零向量 a 和常数 b, 使得

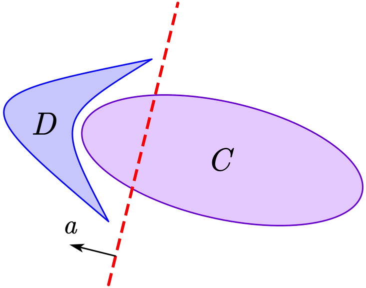
  $$a^\top x < b, \forall x \in \mathcal{C} \quad 且 \quad a^\top x > b, \forall x \in \mathcal{D},$$

  即超平面 $\{x \mid a^\top x = b\}$**严格**分离了 $\mathcal{C}$ 和 $\mathcal{D}$

分离超平面的示意

- 在$\mathbb{R}^2$ 中的 2 个凸集使用超平面即可轻松划分, 但遇到非凸集合就必须使用更加复杂的平面





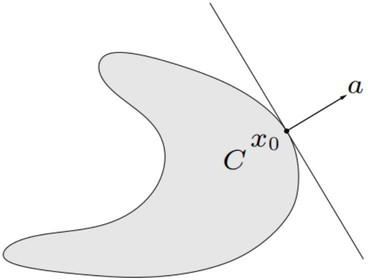
支撑超平面

- 给定集合 $\mathcal{C}$ 以及边界上的点 x_0, 如果 $a \neq 0$ 满足 $a^\top x \leqslant a^\top x_0, \forall x \in \mathcal{C}$, 则称

$$\{x \mid a^\top x = a^0\}$$

  为 $\mathcal{C}$ 在边界点 x_0 处的**支撑超平面**

- 若 $\mathcal{C}$ 是凸集, 则 $\mathcal{C}$ 的任意边界点处都存在支撑超平面



目录

凸函数的定义

- 设$f : \mathbb{R}^n \to \mathbb{R}$ 为适当函数，如果 $\mathrm{dom}\ f$ 是凸集，且

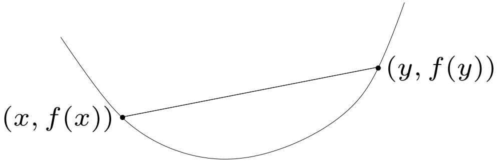
$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta)f(y)$$

  对所有 $x, y \in \mathrm{dom}\ f$, $0 \le \theta \le 1$ 都成立，则称 f 是**凸函数**

- 若对所有 $x, y \in \mathrm{dom}\ f$，$x \neq y$，$0 < \theta < 1$，有

$$f(\theta x + (1-\theta)y) < \theta f(x) + (1-\theta)f(y)$$

  则称 f 是**严格凸函数**



一元凸函数的例子

- 仿射函数　对任意 $a, b \in$，$ax + b$ 是 $\mathbb{R}$ 上的凸函数
- 指数函数　对任意 $a \in$，e^{ax} 是 $\mathbb{R}$ 上的凸函数
- 幂函数　对 $\alpha \geq 1$ 或 $\alpha \leq 0$，x^α 是 $\mathbb{R}_{++}$ 上的凸函数
- 绝对值的幂　对 $p \geq 1$，$|x|^p$ 是 $\mathbb{R}$ 上的凸函数
- 负熵　$x \log x$ 是 $\mathbb{R}_{++}$ 上的凸函数
- 仿射函数　对任意 $a, b \in \mathbb{R}$，$ax + b$ 是 $\mathbb{R}$ 上的凹函数
- 幂函数　对 $0 \leq \alpha \leq 1$，x^α 是 $\mathbb{R}_{++}$ 上的凹函数
- 对数函数　$\log x$ 是 $\mathbb{R}_{++}$ 上的凹函数

多元凸函数的例子

- 所有的仿射函数既是凸函数，又是凹函数

$$f(x) = a^\top x + b$$

$$f(X) = \mathrm{Tr}(A^\top X) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b$$

- 所有的范数都是凸函数

$$f(x) = \|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p} \quad (p \geq 1)$$

$$f(X) = \|X\|_2 = \sigma_{\max}(X) = (\lambda_{\max}(X^\top X))^{1/2}$$

强凸函数

- 若存在常数 $m>0$，使得

$$g(x)=f(x)-\frac{m}{2}\|x\|^2$$

为凸函数，则称 $f(x)$ 为**强凸函数**，其中 m 为强凸参数

- 若存在常数 $m>0$，使得对任意 $x,y\in\mathrm{dom}\ f$ 以及 $\theta\in(0,1)$，有

$$f(\theta x+(1-\theta y))\le\theta f(x)+(1-\theta)f(y)-\frac{m}{2}\theta(1-\theta)\|x-y\|^2,$$

则称 $f(x)$ 为**强凸函数**，其中 m 为强凸参数

- 为了方便也称 $f(x)$ 为 m-强凸函数
- 设 f 为强凸函数且存在最小值，则 f 的最小值点唯一

凸函数判定定理

- 设$f:\mathbb{R}^n \to \mathbb{R}$ 是凸函数，当且仅当对每个 $x \in \mathrm{dom}\ f v \in \mathbb{R}^n$，函数 $g:\mathbb{R} \to \mathbb{R}$ 是关于 t 的凸函数

$$g(t) = f(x+tv), \quad \mathrm{dom}\ g = \{t \mid x+tv \in \mathrm{dom}\ f\}$$

- $f(X) = -\log\det X$ 是凸函数，其中 $\mathrm{dom}\ f = \mathcal{S}^n_{++}$

  证明 任取 $X \succ 0$ 以及方向 $V \in \mathcal{S}^n$，将 f 限制在直线 $X+tV$（t 满足 $X+tV \succ 0$）上，那么

$$\begin{aligned} g(t) = -\log\det(X+tV) &= -\log\det X - \log\det(I+tX^{-1/2}VX^{-1/2}) \\ &= -\log\det X - \sum_{i=1}^{n}\log(1+t\lambda_i) \end{aligned}$$

  其中 λ_i 是 $X^{-1/2}VX^{-1/2}$ 第 i 个特征值. 对每个 $X \succ 0$ 以及方向 V，g 关于 t 是凸的，因此 f 是凸的

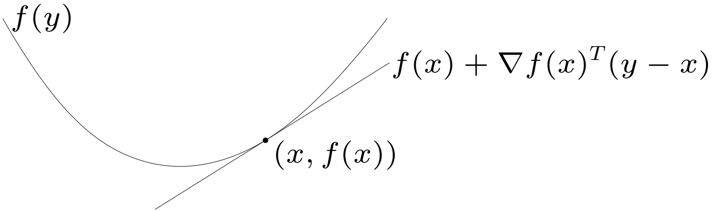
一阶条件

- 凸集上的可微函数 f 是凸函数当且仅当

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) \quad \forall x, y \in \operatorname{dom} f$$

- 设 f 为可微函数，则 f 为凸函数当且仅当 $\operatorname{dom} f$ 为凸集且 ∇f 为单调映射

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq 0, \quad \forall\, x, y \in \operatorname{dom} f$$



二阶条件

- 设f 为定义在凸集上的二阶连续可微函数，f 是凸函数当且仅当

$$\nabla^2 f(x) \succeq 0 \quad \forall x \in \mathrm{dom}\ f$$

如果 $\nabla^2 f(x) \succ 0\ \forall x \in \mathrm{dom}\ f$，则 f 是严格凸函数

- 最小二乘函数 $f(x) = \|Ax - b\|_2^2$

$$\nabla f(x) = 2A^\top(Ax - b), \quad \nabla^2 f(x) = 2A^\top A$$

对任意 A，函数 f 都是凸函数

- 二次函数 $f(x) = (1/2)x^\top Px + q^\top x + r$ (其中 $P \in \mathcal{S}^n$)

$$\nabla f(x) = Px + q, \qquad \nabla^2 f(x) = P$$

f 是凸函数当且仅当 $P \succeq 0$

上方图

- 函数 $f(x)$ 为凸函数当且仅当其上方图 $\mathrm{epi}f$ 是凸集

  必要性 若 f 为凸函数，则对任意 $(x_1, y_1), (x_2, y_2) \in \mathrm{epi}f, t \in [0, 1]$,

$$ty_1 + (1-t)y_2 \geq tf(x_1) + (1-t)f(x_2) \geq f(tx_1 + (1-t)x_2),$$

  故 $(tx_1 + (1-t)x_2, ty_1 + (1-t)y_2) \in \mathrm{epi}f, t \in [0, 1]$

  充分性 若 $\mathrm{epi}f$ 是凸集，则对任意 $x_1, x_2 \in \mathrm{dom}\ f, t \in [0, 1]$,

$$\begin{aligned}&(tx_1 + (1-t)x_2, tf(x_1) + (1-t)f(x_2)) \in \mathrm{epi}f\\ \Rightarrow\ &f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2).\end{aligned}$$

凸函数的判断方法

- 用定义验证（通常将函数限制在一条直线上）
- 利用一阶条件、二阶条件
- 直接研究 f 的上方图 $\mathrm{epi}\ f$
- 说明 f 可由简单的凸函数通过一些保凸的运算得到
  - 非负加权和
  - 与仿射函数的复合
  - 逐点取最大值
  - 与标量、向量函数的复合

非负加权和与仿射函数的复合

- 若 f 是凸函数，则 αf 是凸函数，其中 $\alpha \geq 0$
- 若 f_1, f_2 是凸函数，则 $f_1 + f_2$ 是凸函数
- 若 f 是凸函数，则 $f(Ax+b)$ 是凸函数
- 线性不等式的对数障碍函数

$$f(x) = -\sum_{i=1}^{m} \log(b_i - a_i^\top x), \quad \mathrm{dom}\ f = \{x \mid a_i^\top x < b_i, i = 1, ..., m\}$$

- 仿射函数的（任意）范数 $f(x) = \|Ax + b\|$

逐点取最大值

- 若 $f_1, \cdots, f_m$ 是凸函数，则 $f(x) = \max\{f_1(x), \cdots, f_m(x)\}$ 是凸函数
- 分段线性函数

$$f(x) = \max_{i=1,\cdots,m}(a_i^\top x + b_i)$$

- $x \in \mathbb{R}^n$ 的前 r 个最大分量之和

$$f(x) = x_{[1]} + x_{[2]} + \cdots + x_{[r]}$$

事实上，$f(x)$ 可以写成如下多个线性函数取最大值的形式

$$f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \leq i_1 < i_2 < \cdots < i_r \leq n\}$$

逐点取上界

- 若对每个 $y \in \mathcal{A}$，$f(x,y)$ 是关于 x 的凸函数，则

$$g(x) = \sup_{y \in \mathcal{A}} f(x,y)$$

  是凸函数

- 集合 C 的支撑函数

$$S_C(x) = \sup\nolimits_{y \in C} y^\top x$$

- 集合 C 点到给定点 x 的最远距离

$$f(x) = \sup_{y \in C} \|x - y\|$$

- 对称矩阵 $X \in \mathcal{S}^n$ 的最大特征值

$$\lambda_{\max}(X) = \sup_{\|y\|_2 = 1} y^\top X y$$

与标量函数的复合

- 给定函数 $g:\mathbb{R}^n\to\mathbb{R}$ 和 $h:\mathbb{R}\to\mathbb{R}$,

$$f(x)=h(g(x))$$

若 $\begin{array}{l} g \text{ 是凸函数}, h \text{ 是凸函数}, \tilde{h} \text{ 单调不减} \\ g \text{ 是凹函数}, h \text{ 是凸函数}, \tilde{h} \text{ 单调不增} \end{array}$，那么 f 是凸函数

证明 对 $n=1$，g,h 均可微的情形

$$f''(x)=h''(g(x))g'(x)^2+h'(g(x))g''(x)$$

- 如果 g 是凸函数，则 $\exp g(x)$ 是凸函数
- 如果 g 是正值凹函数，则 $1/g(x)$ 是凸函数

与向量函数的复合

- 给定函数 $g : \mathbb{R}^n \to \mathbb{R}^k$ 和 $h : \mathbb{R}^k \to \mathbb{R}$:

$$f(x) = h(g(x)) = h(g_1(x), g_2(x), ..., g_k(x))$$

若 $\begin{array}{l} g_i \text{ 是凸函数}, h \text{ 是凸函数}, \tilde{h} \text{ 关于每个分量单调不减} \\ g_i \text{ 是凹函数}, h \text{ 是凸函数}, \tilde{h} \text{ 关于每个分量单调不增} \end{array}$，那么 f 是凸函数

证明 对 $n = 1$，g, h 均可微的情形

$$f''(x) = g'(x)^\top \nabla^2 h(g(x)) g'(x) + \nabla h(g(x))^T g''(x)$$

- 如果 g_i 是正值凹函数，则 $\sum_{i=1}^m \log g_i(x)$ 是凹函数
- 如果 g_i 是凸函数，则 $\log \sum_{i=1}^m \exp g_i(x)$ 是凸函数

取下确界

- 若 $f(x,y)$ 关于 (x,y) 整体是凸函数，C 是凸集，则
$$g(x) = \inf_{y\in C} f(x,y)$$
是凸函数
- 考虑函数 $f(x,y) = x^\top Ax + 2x^\top By + y^\top Cy$，海瑟矩阵满足
$$\begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \succeq 0, \quad C \succ 0,$$
则 $f(x,y)$ 为凸函数. 对 y 求最小值得
$$g(x) = \inf_y f(x,y) = x^\top (A - BC^{-1}B^\top)x,$$
因此 g 是凸函数. 进一步地，A 的 Schur 补 $A - BC^{-1}B^\top \succeq 0$
- 点 x 到凸集 S 的距离 $\mathrm{dist}(x,S) = \inf_{y\in S}\|x-y\|$ 是凸函数

透视函数

- 定义 $f:\mathbb{R}^n\to\mathbb{R}$ 的透视函数 $g:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$,

$$g(x,t)=tf(x/t),\quad \mathrm{dom}\ g=\{(x,t)|x/t\in\mathrm{dom}\ f,t>0\}$$

  若 f 是凸函数，则 g 是凸函数

- $f(x)=x^\top x$ 是凸函数，则 $g(x,t)=x^\top x/t$ 是区域 $\{(x,t)\mid t>0\}$ 上的凸函数
- $f(x)=-\log x$ 是凸函数，则 $g(x,t)=t\log t-t\log x$ 是 $\mathbb{R}^2_{++}$ 上的凸函数
- 若 f 是凸函数，则

$$g(x)=(c^\top x+d)f((Ax+b)/(c^\top x+d))$$

  是区域 $\{x\mid c^\top x+d>0,(Ax+b)/(c^\top x+d)\in\mathrm{dom}\ f\}$ 上的凸函数

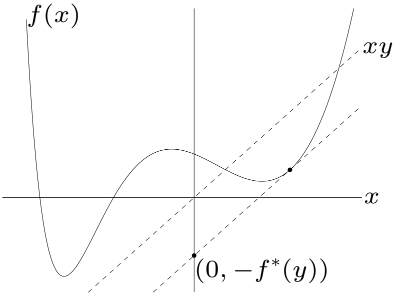
目录

共轭函数

- 适当函数 f 的共轭函数定义为

$$f^*(y) = \sup_{x \in \mathrm{dom}\ f} (y^\top x - f(x))$$

- f^* 恒为凸函数，无论 f 是否是凸函数



例子

- **负对数** $f(x) = -\log x$

$$\begin{aligned} f^*(y) &= \sup_{x>0}(xy + \log x) \\ &= \begin{cases} -1 - \log(-y) & y < 0 \\ \infty & \text{其他} \end{cases} \end{aligned}$$

- **强凸二次函数** $f(x) = (1/2)x^\top Qx, \ Q \in \mathcal{S}^n_{++}$

$$\begin{aligned} f^*(y) &= \sup_{x}(y^\top x - (1/2)x^\top Qx) \\ &= \frac{1}{2}y^\top Q^{-1}y \end{aligned}$$

目录

次梯度

- 可微凸函数 f 的一阶条件

$$f(y) \geq f(x) + \nabla f(x)^\top (y-x)$$

- 设 f 为适当凸函数，$x \in \mathrm{dom}\ f$, 若向量 $g \in \mathbb{R}^n$ 满足

$$f(y) \geq f(x) + g^\top (y-x), \quad \forall y \in \mathrm{dom}\ f$$

则称 g 为函数 f 在点 x 处的一个次梯度

- 进一步地，称集合

$$\partial f(x) = \{g \mid g \in \mathbb{R}^n, f(y) \geq f(x) + g^\top (y-x), \forall y \in \mathrm{dom}\ f\}$$

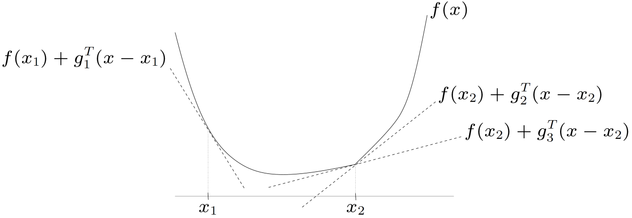
为 f 在点 x 处的次微分

次梯度

- 知$f(x) + g^\top(y - x)$ 是 $f(y)$ 的一个全局下界
- g 可以诱导出上方图 $\mathrm{epi}\ f$ 在点 $(x, f(x))$ 处的一个支撑超平面

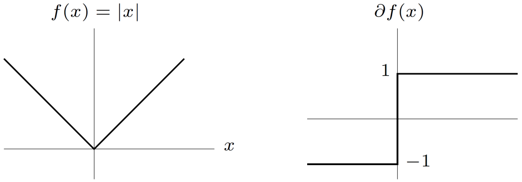
$$\begin{bmatrix} g \\ -1 \end{bmatrix}\left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix}\right) \le 0 \quad \forall\ (y, t) \in \mathrm{epi}\ f$$

- 如果 f 是可微凸函数, 那么 $\nabla f(x)$ 是 f 在点 x 处的一个次梯度
- g_2, g_3 是点 x_2 处的次梯度, g_1 是点 x_1 处的次梯度



次梯度存在性

- 设f 为凸函数，$\mathrm{dom}\ f$ 为其定义域．如果 $x \in \mathrm{int\ dom}\ f$，则 $\partial f(x)$ 是非空的，其中 $\mathrm{int\ dom}\ f$ 的含义是集合 $\mathrm{dom}\ f$ 的所有内点.
- 绝对值函数 $f(x) = |x|$



- 欧几里得范数 $f(x) = \|x\|_2$

$$\text{如果 } x \neq 0, \partial f(x) = \frac{1}{\|x\|_2} x, \quad \text{如果 } x = 0, \partial f(x) = \{g \mid \|g\|_2 \leq 1\}$$

次梯度的性质

- 对任何 $x \in \operatorname{dom} f$，$\partial f(x)$ 是一个闭凸集（可能为空集）
- 如果 $x \in \operatorname{int} \operatorname{dom} f$，则 $\partial f(x)$ 非空有界集
- 设凸函数 $f(x)$ 在 $x_0 \in \operatorname{int} \operatorname{dom} f$ 处可微，则 $\partial f(x_0) = \{\nabla f(x_0)\}$
- 设 $f: \mathbb{R}^n \to \mathbb{R}$ 为凸函数，$x, y \in \operatorname{dom} f$，则 $(u - v)^\top (x - y) \geq 0$, 其中 $u \in \partial f(x)$，$v \in \partial f(y)$
- 设 $f(x)$ 是闭凸函数且 ∂f 在点 $\bar{x}$ 附近存在且非空．若序列 $x^k \to \bar{x}$，$g^k \in \partial f(x^k)$ 为 $f(x)$ 在点 x^k 处的次梯度，且 $g^k \to \bar{g}$，则 $\bar{g} \in \partial f(\bar{x})$

方向导数

- 设 f 为适当函数，给定点 x_0 以及方向 $d \in \mathbb{R}^n$，方向导数（若存在）定义为

$$\lim_{t \downarrow 0} \phi(t) = \lim_{t \downarrow 0} \frac{f(x_0 + td) - f(x_0)}{t}$$

  其中 $t \downarrow 0$ 表示 t 单调下降趋于 0

- 若 f 是凸函数，则 $\phi(t)$ 在 $(0, +\infty)$ 上是单调不减的，$\lim$ 可替换为 $\inf$
- 对于凸函数 f，给定点 $x_0 \in \mathrm{dom}\ f$ 以及方向 $d \in \mathbb{R}^n$，其方向导数定义为

$$\partial f(x_0; d) = \inf_{t > 0} \frac{f(x_0 + td) - f(x_0)}{t}.$$

方向导数有限

- 设$f(x)$ 为凸函数，$x_0 \in \operatorname{int} \operatorname{dom} f$，则对任意 $d \in \mathbb{R}^n$，$\partial f(x_0; d)$ 有限

证明 首先 $\partial f(x_0; d)$ 不为正无穷是显然的. 由于 $x_0 \in \operatorname{int} \operatorname{dom} f$，根据次梯度的存在性定理可知 $f(x)$ 在点 x_0 处存在次梯度 g. 根据方向导数的定义，有

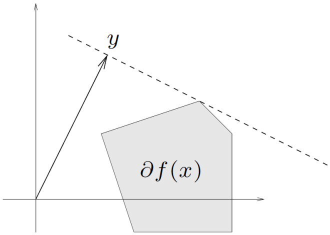
$$
\begin{aligned}
\partial f(x_0; d) &= \inf_{t>0} \frac{f(x_0 + td) - f(x_0)}{t} \\
&\geq \inf_{t>0} \frac{tg^\top d}{t} = g^\top d
\end{aligned}
$$

其中的不等式利用了次梯度的定义. 这说明 $\partial f(x_0; d)$ 不为负无穷

方向导数和次梯度

- 设 $f : \mathbb{R}^n \to (-\infty, +\infty]$ 为凸函数，$x_0 \in \mathrm{int\ dom}\ f$，$d$ 为 $\mathbb{R}^n$ 中任一方向，则

$$\partial f(x_0; d) = \max_{g \in \partial f(x_0)} g^\top d$$



- $\partial f(x; y)$ 是 $\partial f(x)$ 的支撑函数
- $\partial f(x_0; d) = \nabla f(x_0)^\top d$, 切对所有的 $x_0 \in \mathrm{int\ dom}\ f$ 以及所有的 d 都存在

次梯度的计算规则

- 若凸函数 f 在点 x 处可微，则 $\partial f(x) = \{\nabla f(x)\}$
- 设凸函数 f_1, f_2 满足 $\text{int dom } f_1 \cap \text{dom } f_2 \neq \emptyset$，而 $x \in \text{dom } f_1 \cap \text{dom } f_2$. 若

$$f(x) = \alpha_1 f_1(x) + \alpha_2 f_2(x), \quad \alpha_1, \alpha_2 \geq 0$$

则 $f(x)$ 的次微分

$$\partial f(x) = \alpha_1 \partial f_1(x) + \alpha_2 \partial f_2(x)$$

- 设 h 为适当凸函数，f 满足 $f(x) = h(Ax + b)$. 若存在 $x^\sharp \in \mathbb{R}^m$，使得 $Ax^\sharp + b \in \text{int dom } h$，则

$$\partial f(x) = A^\top \partial h(Ax + b), \quad \forall\, x \in \text{int dom } f$$

两个函数之和的次梯度

- 设 $f_1, f_2 : \mathbb{R}^n \to (-\infty, +\infty]$ 是两个凸函数，则对任意的 $x_0 \in \mathbb{R}^n$，

$$\partial f_1(x_0) + \partial f_2(x_0) \subseteq \partial(f_1 + f_2)(x_0).$$

进一步地，若 $\mathrm{int}\ \mathrm{dom}\ f_1 \cap \mathrm{dom}\ f_2 \neq \emptyset$，则对任意的 $x_0 \in \mathbb{R}^n$，

$$\partial(f_1 + f_2)(x_0) = \partial f_1(x_0) + \partial f_2(x_0).$$

证明 对于任意给定的 x_0，设 $g \in \partial(f_1 + f_2)(x_0)$．如果 $f_1(x_0) = +\infty$，则 $(f_1 + f_2)(x_0) = +\infty$．由次梯度的定义，我们有

$$(f_1 + f_2)(x) \geq (f_1 + f_2)(x_0) + g^\top (x - x_0)$$

对任意 $x \in \mathbb{R}^n$ 成立，故 $f_1 + f_2 \equiv +\infty$．这与 $\mathrm{int}\ \mathrm{dom}\ f_1 \cap \mathrm{dom}\ f_2 \neq \emptyset$ 矛盾，因此以下我们假设 $f_1(x_0), f_2(x_0) < +\infty$

函数族的上确界

- 设 $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ 均为凸函数，令

$$f(x) = \max\{f_1(x), f_2(x), \cdots, f_m(x)\}, \quad \forall x \in \mathbb{R}^n$$

对 $x_0 \in \bigcap_{i=1}^{m} \operatorname{int} \operatorname{dom} f_i$，定义 $I(x_0) = \{i \mid f_i(x_0) = f(x_0)\}$，则

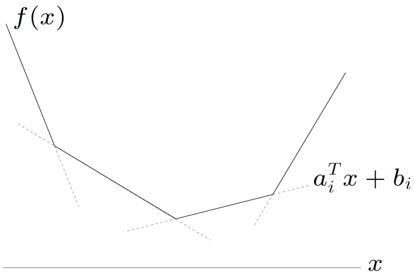
$$\partial f(x_0) = \operatorname{conv} \bigcup_{i \in I(x_0)} \partial f_i(x_0)$$

- $I(x_0)$ 表示点 x_0 处“有效”函数的指标
- $\partial f(x_0)$ 是点 x_0 处“有效”函数的次微分并集的凸包
- 如果 f_i 可微, $\partial f(x_0) = \operatorname{conv}\{\nabla f_i(x_0) \mid i \in I(x_0)\}$

例子

- **分段线性函数**

$$f(x) = \max_{i=1,2,\cdots,m} \{a_i^\top x + b_i\}$$



- **点 x 处的次微分是一个多面体**

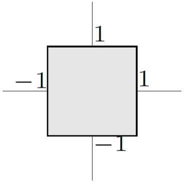
$$\partial f(x) = \mathrm{conv}\{a_i \mid i \in I(x)\}$$

其中 $I(x) = \{i \mid a_i^\top x + b_i = f(x)\}$

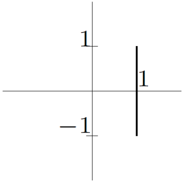
- ℓ_1-范数

$$f(x) = \|x\|_1 = \max_{s \in \{-1,1\}^n} s^\top x$$

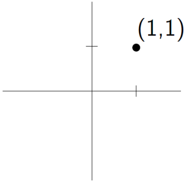
$$\partial f(x) = J_1 \times \cdots \times J_n, \quad J_k = \begin{cases} [-1,1], & x_k = 0 \\ \{1\}, & x_k > 0. \\ \{-1\}, & x_k < 0 \end{cases}$$



$\partial f(0,0) = [-1,1] \times [-1,1]$



$\partial f(1,0) = \{1\} \times [-1,1]$



$\partial f(1,1) = \{(1,1)\}$

复合函数

- 设 $f_1, f_2, \cdots, f_m : \mathbb{R}^n \to (-\infty, +\infty]$ 为 m 个凸函数，$h : \mathbb{R}^m \to (-\infty, +\infty]$ 为关于各分量单调递增的凸函数，令

$$f(x) = h(f_1(x), f_2(x), \cdots, f_m(x))$$

- $z = (z_1, z_2, \cdots, z_m) \in \partial h(f_1(\hat{x}), f_2(\hat{x}), \cdots, f_m(\hat{x}))$ 以及 $g_i \in \partial f_i(\hat{x})$
- $g z_1 g_1 + z_2 g_2 + \cdots + z_m g_m \in \partial f(\hat{x})$

证明

$$\begin{aligned} f(x) &\geq h(f_1(\hat{x}) + g_1^\top (x - \hat{x}), f_2(\hat{x}) + g_2^\top (x - \hat{x}), \cdots, f_m(\hat{x}) + g_m^\top (x - \hat{x})) \\ &\geq h(f_1(\hat{x}), f_2(\hat{x}), \cdots, f_m(\hat{x})) + \sum_{i=1}^m z_i g_i^\top (x - \hat{x}) \\ &= f(\hat{x}) + g^\top (x - \hat{x}) \end{aligned}$$

Q&A

Thank you!

感谢您的聆听和反馈