第 2 章：数学基础

(3)：凸函数

授课教师：曹语

课程主页：`https://yucaoyc.github.io/math3806`

背景和目标

在学习微积分时，有个概念叫做“曲率”（向上弯/向下弯），在一维可以用二阶导数 f'' 来刻画

维数	条件
1	$f''(x) \geq 0$
n	$\nabla^2 f(\boldsymbol{x}) \succeq 0$

$\Longrightarrow$ 凸函数

本节课的学习目标

- 什么是凸集?（上节课）
- 什么是凸函数?
- 如果 Ω 为凸集，f_0 为凸函数，最优解有什么性质?

判断题: 假设定义域 $\Omega = \mathbb{R}^n$，且 $\nabla^2 f$ 存在并且连续。请判断下列说法真伪。

T/F 若 $\boldsymbol{x}^\star$ 是局部极小点，则 $\nabla^2 f(\boldsymbol{x}^\star) \succeq 0$

T/F 若 $\nabla f(\boldsymbol{x}) = \mathbf{0}$, 且 $\nabla^2 f(\boldsymbol{x}) \succeq 0$，则 $\boldsymbol{x}$ 是局部极小点

T/F 若 $\nabla f(\boldsymbol{x}) = \mathbf{0}$, 且 $\nabla^2 f(\boldsymbol{x}) \succ 0$，则 $\boldsymbol{x}$ 是局部极小点

提示: 二阶导数/ Hessian 给的是“曲率信息”。

目录

定义

定义 (凸函数)

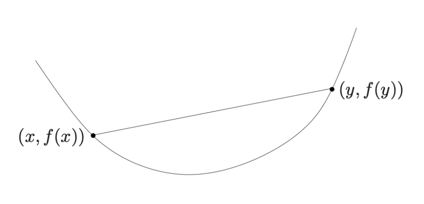
考虑函数 $f : \Omega \to \mathbb{R}$。首先要求定义域 $\Omega \subset \mathbb{R}^n$ 是凸集（保证线段 $\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}$ 仍在 Ω 内）。对任意 $\boldsymbol{x}, \boldsymbol{y} \in \Omega$ 和 $0 \leqslant \theta \leqslant 1$ 有

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leqslant \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y}), \qquad \text{(Jensen 不等式)}$$

则称该函数 f 是凸函数。

几何理解：

图像在两点连线（弦）的下方



课堂练习：利用定义验证

练习： 请利用凸函数的定义（Jensen 不等式），证明：一维仿射函数 $f(x) = ax + b$ （$a, b \in \mathbb{R}$）是凸函数。

练习： 如果我们要证明 $f(x) = x^2$ 是凸函数，根据定义我们需要证明怎样的一个代数不等式成立?

严格凸 如果当 $\boldsymbol{x} \neq \boldsymbol{y}$，$0 < \theta < 1$ 时，该不等式严格成立，则称函数 f 是严格凸；

凹函数 如果函数 $-f$ 是凸的，则称函数 f 是凹函数；

严格凹 如果 $-f$ 严格凸，则称函数 f 严格凹。

例子：所有的仿射函数既是凸函数，也是凹函数。

零阶条件

定理 (凸函数的零阶条件)

考虑定义域为 Ω 的函数 f，f 是凸的当且仅当对于任意 $\boldsymbol{x} \in \Omega$ 和任意向量 $\boldsymbol{v}$，函数 $g(t) = f(\boldsymbol{x} + t\boldsymbol{v})$ 是凸的（其定义域为 $\{t \mid \boldsymbol{x} + t\boldsymbol{v} \in \Omega\}$）。

大致思路:

- $\Rightarrow$：若 f 在 Ω 上凸，则可验证限制在任意直线 $\boldsymbol{x} + t\boldsymbol{v}$ 上仍满足 Jensen 不等式，所以 $g(t)$ 凸。
- $\Leftarrow$：任取 $\boldsymbol{x}, \boldsymbol{y} \in \Omega$，令 $\boldsymbol{v} = \boldsymbol{y} - \boldsymbol{x}$，并取 $g(t) = f\big(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x})\big)$，则 $t \in [0, 1]$ 对应 $\boldsymbol{x}$ 和 $\boldsymbol{y}$ 连线的线段。由于 g 凸，有 $g(\theta) \leq \theta g(1) + (1 - \theta)g(0)$，即 Jensen 不等式。

一阶条件

定理 (凸函数的一阶条件)

若 $f : \Omega \to \mathbb{R}$ 可微，

- 则函数 f 是凸函数的充要条件是 Ω 是凸集，且

$$f(\boldsymbol{y}) \geq f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}), \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \Omega$$

- 函数 f 严格凸的充要条件是 Ω 是凸集，且

$$f(\boldsymbol{y}) > f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^\top (\boldsymbol{y} - \boldsymbol{x}), \qquad \forall \boldsymbol{x}, \boldsymbol{y} \in \Omega, \boldsymbol{x} \neq \boldsymbol{y}$$

直观（碗与小桌板）：在一张“碗”状的函数图像下方垫一块“小桌板（切平面）”，这个桌子永远只能在碗的下方“稳稳托住”它，绝不会刺穿到碗的内部，**即切线或切平面总是给出凸函数的一个全局下界估计。**

课堂小测：一阶条件

练习： 考虑一维连续可微函数 $f(x) = x^2$，其定义域为 $\mathbb{R}$。请写出它在 $x_0 = 1$ 处的一阶条件不等式：$f(x) \geq f(1) + f'(1)(x-1), \ \forall x \in \mathbb{R}$ 并画草图验证这条切线是否“始终在函数图像的下方”。

练习： 考虑函数 $f(x) = x^3$。请问它在定义域 $\mathbb{R}$ 上是凸函数吗？可以利用一阶条件，或者在草稿纸上画图说明。

一阶条件的推论

定理 (一阶条件的推论)

若 f 为凸函数且可微，且在某点 x 处 $\nabla f(x) = \mathbf{0}$，则 x 为全局极小点。

"局部即大局"：为什么大家这么喜欢凸函数？这个推论道出了核心：**一旦你在谷底找到了一个平坦的地方（导数为零、局部极小），就不需要再四处寻觅，它就是整个世界的最低点(全局最优)！**

二阶条件

定理 (二阶条件)

若 $f : \Omega \to \mathbb{R}$ 二阶可微，则

- 函数 f 是凸函数的充要条件是 Ω 是凸集，且 $\nabla^2 f(\boldsymbol{x}) \succeq \mathbf{0}_{n\times n}$，$\forall \boldsymbol{x} \in \Omega$
- 函数 f 是凹函数的充要条件是 Ω 是凸集，且 $\nabla^2 f(\boldsymbol{x}) \preceq \mathbf{0}_{n\times n}$，$\forall \boldsymbol{x} \in \Omega$

二阶条件的推论

定理 (二阶条件的推论)

若 $f:\Omega\to\mathbb{R}$ 二阶可微，则

- 若 Ω 是凸集，且 $\nabla^2 f(\boldsymbol{x})\succ\boldsymbol{0}_{n\times n}$，$\forall\boldsymbol{x}\in\Omega$，则函数 f 是严格凸函数
- 若 Ω 是凸集，且 $\nabla^2 f(\boldsymbol{x})\prec\boldsymbol{0}_{n\times n}$，$\forall\boldsymbol{x}\in\Omega$，则函数 f 是严格凹函数

提示：该命题反过来不一定对；例如 $f(x)=x^4$ 在 $x=0$ 处二阶导为零，但它仍然是严格凸函数。

课堂小测：二阶条件

练习： 已知二维函数 $f(x,y) = 2x^2 + 2xy + y^2 - 4x + 6y$，定义域均为 $\mathbb{R}^2$。请问这个函数是（严格）凸函数吗?

练习： 如果函数改为 $f(x,y) = x^2 - y^2$，它是凸函数吗?

例子

例题 1: 某个函数既凸又凹 $\Longleftrightarrow$ 它是仿射函数

例题 2: 二次函数 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^\top \boldsymbol{x} + r$ 的凹凸形貌，完全由用来刻画“锅形”的矩阵 $\boldsymbol{P}$（即海森矩阵）来决定

例题 3: 对于任意的 $a \in \mathbb{R}$, $f(x) = e^{ax}$ 在 $\mathbb{R}$ 上是凸函数（增长极其夸张的陡坡，永远朝上弯）

例题 4: 对于任意的 $p \geq 1$, $f(x) = |x|^p$ 在 $\mathbb{R}$ 上是凸函数（各种不同平滑程度的底部开口向上的碗）

例题 5: 在定义域 $(0, \infty)$ 上，对数函数 $f(x) = \log(x)$ 是凹函数

例子

例题 6: 在定义域 $(0, \infty)$ 上，负熵函数 $f(x) = x\log(x)$ 是凸函数（信息论和热力学的常用函数）

例题 7: 范数函数 $f(\boldsymbol{x}) = \|\boldsymbol{x}\|$ 是凸函数

例题 8: 最大值函数 $f(\boldsymbol{x}) = \max\{x_1, x_2, \cdots, x_n\}$ 是凸函数

例题 9: 指数和的对数 $f(\boldsymbol{x}) = \log(e^{x_1} + e^{x_2} + \cdots + e^{x_n})$ 在 $\mathbb{R}^n$ 上是凸函数

目录

性质一：凸函数的下水平集为凸集

定义 (α-下水平集)

对于函数 $f:\Omega\to\mathbb{R}$，α-**下水平集**定义为

$$C_\alpha = \{\boldsymbol{x}\in\Omega \mid f(\boldsymbol{x})\leqslant\alpha\}$$

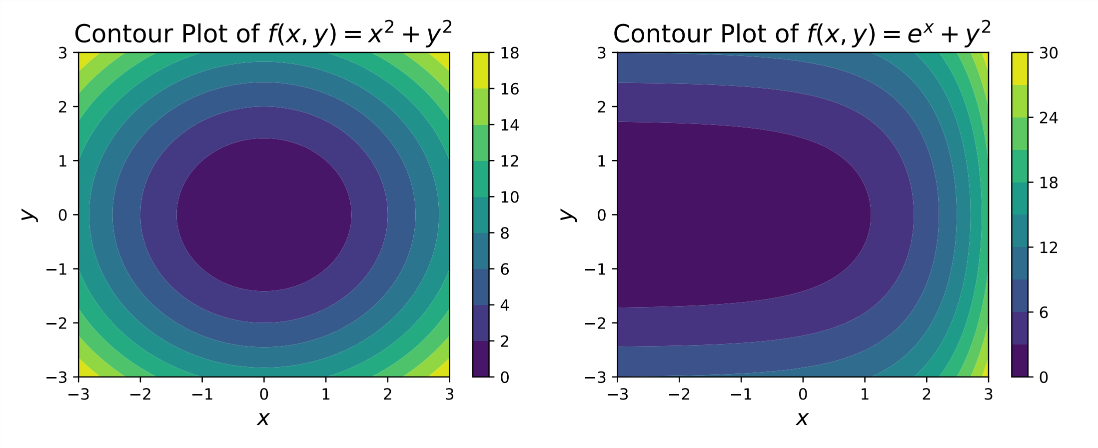
定理

若 f 为凸函数，则对于任意 $\alpha\in\mathbb{R}$, C_α 是凸集。

简要证明：

- 取任意 $\boldsymbol{x},\boldsymbol{y}\in C_\alpha$，则 $f(\boldsymbol{x})\le\alpha$，$f(\boldsymbol{y})\le\alpha$。
- 由凸性，$f(\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y})\le\theta f(\boldsymbol{x})+(1-\theta)f(\boldsymbol{y})\le\alpha$，因此 $\theta\boldsymbol{x}+(1-\theta)\boldsymbol{y}\in C_\alpha$，故而 C_α 是凸集。

探险家看地图的视角：如果你在用等高线地图分析一座山（即凸函数），你圈出“海拔低于某个高度 α ”的区域（此即“下水平集”），这个区域一定是**向外鼓的凸集形状**。



性质二：Jensen 不等式的拓展

Jensen 不等式

$$f(\theta \boldsymbol{x} + (1-\theta)\boldsymbol{y}) \leqslant \theta f(\boldsymbol{x}) + (1-\theta) f(\boldsymbol{y})$$

Jensen 不等式的拓展版： 若函数 f 是凸函数，$\boldsymbol{x}_1, \cdots, \boldsymbol{x}_k \in \Omega$, $\theta_1, \cdots, \theta_k \geqslant 0$ 且 $\theta_1 + \cdots + \theta_k = 1$ ，则

$$f\left(\theta_1 \boldsymbol{x}_1 + \cdots + \theta_k \boldsymbol{x}_k\right) \leqslant \theta_1 f\left(\boldsymbol{x}_1\right) + \cdots + \theta_k f\left(\boldsymbol{x}_k\right)$$

概率视角： 如果把 θ_k 看成是某个事件发生的概率，即 $\boldsymbol{X} = \boldsymbol{x}_k$ 的概率为 θ_k，则 Jensen 不等式的拓展版可以被改写为

$$f\big(\mathbb{E}(\boldsymbol{X})\big) \leq \mathbb{E}\big(f(\boldsymbol{X})\big)$$

这里用到：$\mathbb{E}(\boldsymbol{X})$ 本质上就是“加权平均/凸组合”。

目录

(1) 非负加权

凸函数的非负加权： 若 $f_1, f_2, \cdots, f_k$ 为凸函数，$w_1, \cdots, w_k \geq 0$，则

$$f(\boldsymbol{x}) = \sum_{i=1}^{k} w_i f_i(\boldsymbol{x})$$

是凸函数。

即所有凸函数的集合构成一个凸锥

(2) 复合仿射映射

复合仿射映射： 假设函数 $f:\mathbb{R}^n \to \mathbb{R}$ 为凸函数，矩阵 $\boldsymbol{A} \in \mathbb{R}^{n\times m}$ ，向量 $\boldsymbol{b} \in \mathbb{R}^n$，则如下的函数 $g:\mathbb{R}^m \to \mathbb{R}$ 为凸函数:

$$g(\boldsymbol{x}) = f(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b})$$

(3) 逐点最大和逐点上确界

逐点最大： 如果函数 $f_1, f_2, \cdots, f_k$ 均为凸函数，则逐点最大函数 f

$$f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_k(\boldsymbol{x})\} \text{是凸函数。}$$

逐点上确界： 如果对于任意 $\boldsymbol{y} \in \mathcal{A}$，函数 $g(\boldsymbol{x}, \boldsymbol{y})$ 关于 $\boldsymbol{x}$ 都是凸的，则函数 $f(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in \mathcal{A}} g(\boldsymbol{x}, \boldsymbol{y})$ 是关于 $\boldsymbol{x}$ 的凸函数。

例子：

- 由于 $g(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|$ 关于 $\boldsymbol{x}$ 是凸函数，点 $\boldsymbol{x}$ 到集合 C 最远的距离是凸函数，即

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

- 对称矩阵的最大特征值 $f(\boldsymbol{X}) = \sup\left\{\boldsymbol{y}^\top \boldsymbol{X} \boldsymbol{y} \mid \|\boldsymbol{y}\|_2 = 1\right\}$ 是凸函数

(4) 复合函数（I）

复合函数： 给定函数 $h:\mathbb{R}^k\to\mathbb{R}$ 以及 $g:\mathbb{R}^n\to\mathbb{R}^k$，复合函数 $f=h\circ g:\mathbb{R}^n\to\mathbb{R}$ 为 $f(\boldsymbol{x})=h(g(\boldsymbol{x}))$

后续我们先以 $n=1$ 作为例子（方便直接看二阶导数的符号）。

进一步考虑 $k=1$ 的情况，

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

我们可以得到一些规律：

- 如果 h 是凸函数且非减，g 是凸函数，则 f 是凸函数；
- 如果 h 是凸函数且非增，g 是凹函数，则 f 是凸函数；
- $\cdots$

复合函数（II）

接下来考虑 $k \geq 2$ 的情况，$f(x) = h\big(g_1(x), g_2(x), \cdots, g_k(x)\big)$

练习： 请计算

$$f''(x) = \underline{\qquad\qquad\qquad ?? \qquad\qquad\qquad}$$

提示：先用链式法则写出 $f'(x) = \sum_{i=1}^{k} \partial_i h\big(g(x)\big)\, g_i'(x)$，再对 x 求导一次。

我们类似可知：如果 h 是凸函数且在每维分量上 h 非减，g_i 是凸函数，则 f 是凸函数。【其他类似的结论请见课本】

例子： $h(\boldsymbol{z}) = \log\left(\sum_{i=1}^{k} \mathrm{e}^{z_i}\right)$ 是凸函数，且在每个分量上非减，因此只要 g_i 是凸函数，$f(x) = \log\left(\sum_{i=1}^{k} \mathrm{e}^{g_i(x)}\right)$ 是凸函数。 【更多例子见课本例 3.14】

(5) 最小化

结论（对凸集的部分最小化保凸）： 若函数 $g(\boldsymbol{x}, \boldsymbol{y})$ 关于 $(\boldsymbol{x}, \boldsymbol{y})$ 是凸函数，集合 C 是非空凸集，定义

$$f(\boldsymbol{x}) = \inf_{\boldsymbol{y} \in C} g(\boldsymbol{x}, \boldsymbol{y})$$

假设函数 f 的定义域非空，则 f 是关于 $\boldsymbol{x}$ 是凸函数。

例子： 某点到一个凸集 C 的最小距离是凸函数，即若 C 是凸集，则

$$\mathrm{dist}(\boldsymbol{x}, C) = \inf_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|$$

是关于 $\boldsymbol{x}$ 的凸函数。

提示： 最大距离是凸函数这一结论对于集合 C 并不需要任何条件；对最小距离，我们需要集合 C 是凸集；请将该结论和第25页的结论对比。

练习： 考虑一维问题

- 说明 $C = [-2, -1] \cup [1, 2]$ 不是凸集；
- 说明 $\mathrm{dist}(\boldsymbol{x}, C) = \inf_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|$ 不是关于 $\boldsymbol{x}$ 的凸函数；
- 画出 $f(\boldsymbol{x}) = \sup_{\boldsymbol{y} \in C} \|\boldsymbol{x} - \boldsymbol{y}\|$，并说明该函数是凸函数。

目录

总结

- 掌握凸函数、凹函数的定义、几何解释
- 掌握零阶、一阶、二阶条件的结论
- 对于具体的例子能验证是否是凸/凹函数
- 知道保凸运算，并能对于具体例子证明保凸运算

阅读作业 & 参考资料：

- 课本第 3.1 - 3.2 章（课本 3.3 章将在后面再介绍）

参考答案

- 第4页答案：True, False, True
- 第11页答案：$y(x) = x^3$ 不是凸函数。当 $x > 0$ 时它是凸的；当 $x < 0$ 时它是凹的。在 $x = 0$ 处它的切线穿过了图像。
- 第15页答案：第一个函数是凸函数，第二个函数不是凸函数。
  1. 先求一阶梯度矩阵 $\nabla f(x, y)$；
  2. 再求二阶海森矩阵（Hessian）$\nabla^2 f(x, y)$；
  3. 判断海森矩阵是否半正定（或者正定）。

$$\nabla^2 f = \left[\begin{smallmatrix} 4 & 2 \\ 2 & 2 \end{smallmatrix}\right], \quad \det\left(\nabla^2 f\right) = 8 - 4 > 0, \quad \text{正定，所以严格凸。}$$