第 2 章：数学基础

(3)：凸函数

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，由 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

在优化中，求解凸函数的极小值具有很好的理论，因此常被作为测试、开发优化算法的基础

在本章，我们需要学习如下问题：

- 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；在这个前提下

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

目录

定义

凸函数 考虑函数 $f:\Omega\to\mathbb{R}$，如果定义域 $\Omega\subset\mathbb{R}^n$ 是凸集，且对于任意 $\boldsymbol{x},\boldsymbol{y}\in\Omega$ 和任意 $0\leqslant\theta\leqslant 1$ 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；

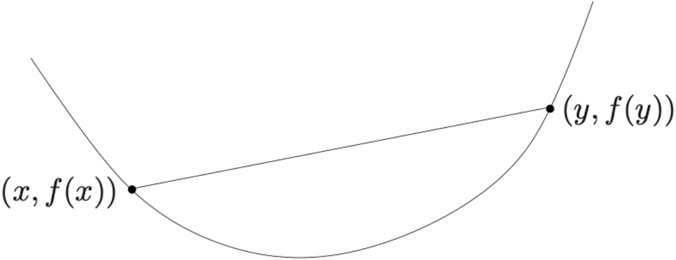
严格凸 如果当 $\boldsymbol{x}\neq\boldsymbol{y}$，$0<\theta<1$ 时，该不等式严格成立，则称函数 f 是严格凸；

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。

凸函数的几何解释



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 $\{t \mid \boldsymbol{x} + t\boldsymbol{v} \in \Omega\}$）。

（证明见课本或课内）

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

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

（证明见课本或课内）

定理 (一阶条件的推论)

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

用处：对于凸函数求极小值，只要梯度很小，就近似是全局最优解。

定理

若 $f:\Omega\to\mathbb{R}$ 可微，

- 函数 f 严格凸的充要条件是 Ω 是凸集，且

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

- 函数 f 是凹函数的充要条件是 Ω 是凸集，且

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

- 函数 f 是严格凹的充要条件是 Ω 是凸集，且

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

二阶条件

定理

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 $\nabla^2 f(\boldsymbol{x}) \preceq \mathbf{0}_{n\times n}$，$\forall \boldsymbol{x} \in \Omega$

定理

若 $f : \Omega \to \mathbb{R}$ 二阶可微，则

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

提示：该命题反过来不一定对；例如 $f(x) = x^4$

例子

例题 1: 某个函数是既凸又凹的等价于其是仿射函数

例题 2: 二次函数 $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{P}\boldsymbol{x} + \boldsymbol{q}^\top \boldsymbol{x} + r$ 的凹凸性由矩阵 $\boldsymbol{P}$ 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

例题 4: 对于任意的 $p \geq 1$, $f(x) = |x|^p$ 在 $\mathbb{R}$ 上是凸函数

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

例题 6: 在定义域 $(0, \infty)$ 上，负熵函数 $f(x) = x\log(x)$ 是凸函数

例子

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

例题 10: 定义域为 $\mathbf{S}^n_{++}$ 上，$f(\boldsymbol{X}) = \log\det \boldsymbol{X}$ 是凹函数

【提示：利用零阶条件来判定】

目录

性质一：凸函数的下水平集为凸集

定义 (α-下水平集)

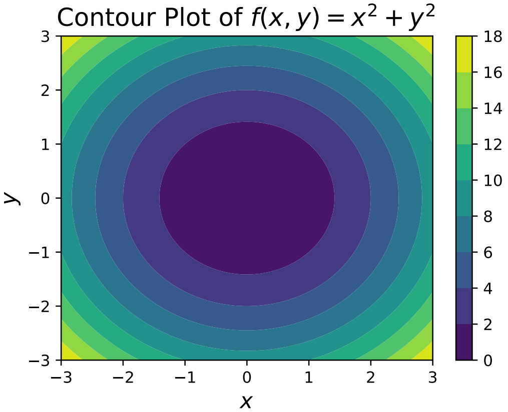
对于函数 $f : \Omega \to \mathbb{R}$，$\alpha$-**下水平集**定义为

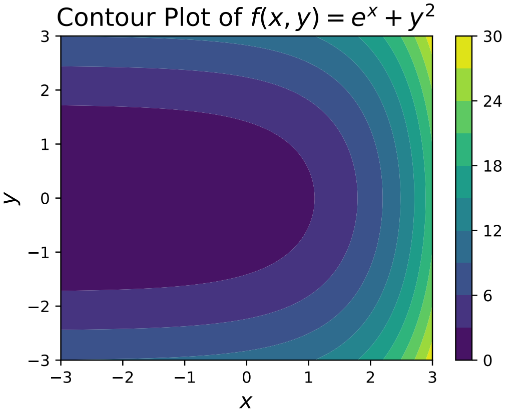
$$C_\alpha = \{\boldsymbol{x} \in \Omega \mid f(\boldsymbol{x}) \leqslant \alpha\}$$

定理

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

【几何含义：对凸函数画等高线地图可以形成一系列的凸集。】


Contour Plot of $f(x, y) = x^2 + y^2$
x
y
18
16
14
12
10
8
6
4
2
0


Contour Plot of $f(x, y) = e^x + y^2$
x
y
30
24
18
12
6
0

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

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

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

$$f(\boldsymbol{x}) = \max\{f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \cdots, f_k(\boldsymbol{x})\}$$

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 $f(\boldsymbol{X}) = \sup\{\boldsymbol{y}^T\boldsymbol{X}\boldsymbol{y} \mid \|\boldsymbol{y}\|_2 = 1\}$ 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 $f(\boldsymbol{x})=h(g(\boldsymbol{x}))$

后续我们将以 $n=1$ 作为例子。

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；
- ...

例子： 如果 g 是凸函数，则 $e^{g(x)}$ 是凸函数。更多例子请见课本例 3.13

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

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

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： 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： 某点到一个凸函数的最小距离是凸函数，即若 C 是凸集，则

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

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 并不需要任何附加条件；对于最小距离，我们需要；请将这页的结论和第20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
- 对于具体的例子能证明是否是凸/凹函数
- 了解 Jensen 不等式的几个变式
- 知道 5 个保凸运算的结论，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）

第3页答案：True, False, True；其他练习答案见课内