优化理论与算法

第五章

凸函数

郭加熠|助理教授

目录

凸函数（Convex Function）

定义函数 f 是**凸函数**当且仅当：

1. $\mathrm{dom}(f)$ 是一个凸集
2. 对任意 $x_1, x_2 \in \mathrm{dom}(f)$ 且 $\theta \in [0, 1]$，有：

$$f(\theta x_1 + (1-\theta)x_2) \leq \theta f(x_1) + (1-\theta)f(x_2)$$

问题：若上述不等式仅在 $\theta = 1/2$ 成立，函数是否为凸函数?

关于凸函数的进一步讨论（More on Convex Function）

定义如果 $\mathrm{dom}(f)$ 是凸集，且

$$f(\theta x_1 + (1-\theta)x_2) < \theta f(x_1) + (1-\theta)f(x_2)$$

对所有 $x_1, x_2 \in \mathrm{dom}(f)$ 且 $\theta \in [0,1]$ 成立，则称 f 是**严格凸函数**（**strictly convex**）。

定义若 $-f$ 是凸函数，则 f 是**凹函数**（**concave**）。

定义若 $-f$ 是严格凸函数，则 f 是**严格凹函数**（**strictly concave**）。

目录

实数域 $\mathbb{R}$ 上的例子（Examples on $\mathbb{R}$）

凸函数（Convex）：

- 仿射函数（Affine）：$ax+b$，定义在 $\mathbb{R}$，其中 $a,b\in\mathbb{R}$
- 指数函数（Exponential）：e^{ax}，定义在 $\mathbb{R}$，其中 $a\in\mathbb{R}$
- 幂函数（Power）：x^p，定义在 $(0,+\infty)$，当 $p\geq 1$ 或 $p\leq 0$
- 绝对值幂函数：$|x|^p$，定义在 $\mathbb{R}$，当 $p\geq 1$
- 负熵函数（Negative entropy）：$x\ln x$，定义在 $(0,+\infty)$

凹函数（Concave）：

- 仿射函数（Affine）：$ax+b$，定义在 $\mathbb{R}$，其中 $a,b\in\mathbb{R}$
- 幂函数（Power）：x^p，定义在 $(0,+\infty)$，当 $0\leq p\leq 1$
- 对数函数（Logarithm）：$\ln x$，定义在 $(0,+\infty)$

例子：仿射函数与范数（Examples: Affine Functions and Norms）

仿射函数既是凸的也是凹的。

例子：

- $f(x) = a^\top x + b$，其中 $a \in \mathbb{R}^n$, $b \in \mathbb{R}$
- $f(X) = \mathrm{tr}(A^\top X) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} x_{ij} + b$

范数是凸函数（Norms are convex）：

- 一般的 l_p 范数：$\|x\|_p = (\sum_{i=1}^{n} |x_i|^p)^{1/p}$
- 谱范数（最大奇异值范数）：

$$f(X) = \|X\|_2 = \sigma_{\max}(A) = \sqrt{\lambda_{\max}(A^\top A)}$$

目录

验证函数的凸性（Verifying Convexity of a Function）

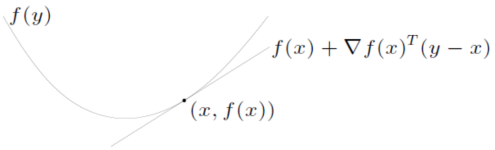
我们可以通过以下方式验证一个给定函数 f 是否为凸函数：

- 定义（Definition）
- **应用一些特殊判据（Applying some special criteria）**
  - **一阶条件（First-order conditions）**
  - 二阶条件（Second-order conditions）
  - 归约为标量函数（Reduction to a scalar function）
  - 上图函数法（Epigraph）
- 证明 f 是通过保持凸性的运算构造得到的

一阶条件（First-Order Condition）

一阶条件（1st-order condition）：可微函数 f 是凸的，当且仅当它的定义域是凸的，且满足：

$$f(x) + \nabla f(x)^\top (y - x) \le f(y) \quad \text{对所有} x, y \in \mathrm{dom}(f)$$



一阶近似是 f 的一个全局下界（**underestimate**）

对于无约束优化问题，有最优解 $x^* \in \{x : \nabla f(x) = 0\}$

一阶条件的证明（Proof of First-Order Condition）

微积分回顾（Review of Calculus）

导数： $f'(x) = \lim_{\epsilon\to 0} \dfrac{f(x+\epsilon)-f(x)}{\epsilon}$

梯度： $\nabla f = \left[\dfrac{\partial f(x)}{\partial x_1}, \ldots, \dfrac{\partial f(x)}{\partial x_n}\right]^\top$ 其中 $\dfrac{\partial f(x)}{\partial x_i} = \lim_{\epsilon\to 0} \frac{f(x+\epsilon e_i)-f(x)}{\epsilon}$

方向导数： $f'(x;d) = \lim_{\epsilon\to 0} \frac{f(x+\epsilon d)-f(x)}{\epsilon} = \sum_{i=1}^n \frac{\partial f(x)}{\partial x_i} d_i = \nabla f(x)^\top d$

证明： 假设 $f(x)$ 是凸函数。对任意 $x, y \in \operatorname{dom} f$，以及 $\theta \in [0,1]$，有：

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

等价于

$$\begin{aligned} f(y) &\ge \frac{1}{\theta}\left[f(x+\theta(y-x)) - (1-\theta)f(x)\right] = f(x) + \frac{f(x+\theta(y-x))-f(x)}{\theta} \\ &= f(x) + \nabla f(x)^\top (y-x) \quad \text{当 } \theta \to 0 \end{aligned}$$

一阶条件的证明（Proof of First-Order Condition）

证明： 假设一阶条件成立。

对任意 $x, y \in \mathrm{dom}\, f$ 且 $\theta \in [0,1]$，令 $z = \theta x + (1-\theta)y$，我们有：

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

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

现在考虑：

$$\theta f(x) + (1-\theta) f(y) \geq f(z) + \nabla f(z)^\top \left[\theta(x - z) + (1-\theta)(y - z)\right]$$

$$= f(z) + \nabla f(z)^\top \left[\theta x + (1-\theta)y - z\right] = f(z)$$

验证函数的凸性（Verifying Convexity of a Function）

我们可以通过以下方式验证一个函数 f 是否为凸函数：

- 定义（Definition）
- **应用一些特殊判据（Applying some special criteria）**
  - 一阶条件（First-order conditions）
  - **二阶条件（Second-order conditions）**
  - 归约为标量函数（Reduction to a scalar function）
  - 上图法（Epigraph）
- 证明 f 是通过保持凸性的操作构造得到的

二阶条件（Second-Order Condition）

函数 f 是**二阶可微**（twice differentiable），当 $\mathrm{dom}(f)$ 是开集，且 Hessian 矩阵 $\nabla^2 f(x) \in \mathbb{S}^n$，即

$$[\nabla^2 f(x)]_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j} \quad \text{对于} i,j = 1,\ldots,n$$

在每个 $x \in \mathrm{dom}(f)$ 上都存在。

二阶条件（2nd-order condition）： 对于定义域为凸集的二阶可微函数 f，有：

- f 是凸函数当且仅当：

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

- 若 $\nabla^2 f(x) \succ 0$，则 f 是严格凸函数。

二阶条件的证明（Proof of Second-Order Condition）

微积分回顾（Review of Calculus）

泰勒展开（单变量）：

$$f(x+a) = f(x) + f'(x)a + \frac{1}{2}f''(x)a^2 + \cdots = \sum_{k=0}^{\infty} \frac{f^{(k)}(x)}{k!}a^k$$

双变量的线性近似：

$$f(x_1+a, x_2+b) \approx f(x_1,x_2) + \frac{\partial f}{\partial x_1}a + \frac{\partial f}{\partial x_2}b = f(x_1,x_2) + \nabla f(x)^\top [a; b]$$

多变量泰勒展开：

$$f(x+a) = f(x) + \nabla f(x)^\top a + \frac{1}{2}a^\top \nabla^2 f(x) a + \ldots$$

证明： 假设二阶条件成立。对任意 $x, y \in \mathrm{dom} f$，有：

$$f(y) = f(x) + \nabla f(x)^\top (y-x) + \frac{1}{2}(y-x)^\top \nabla^2 f(x)(y-x) + \cdots \geq f(x) + \nabla f(x)^\top (y-x)$$

二阶条件的证明（Proof of Second-Order Condition）

证明： 假设 $f(x)$ 是凸函数。对任意 $x, y \in \mathrm{dom} f$，一阶条件给出：

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

将这两个不等式相加，得到：

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq 0$$

由广义中值定理（generalized mean-value theorem），存在 $\theta \in (0, 1)$，使得：

$$\nabla f(x) - \nabla f(y) = \nabla^2 f(\theta x + (1-\theta) y)(x - y)$$

因此，有：$(\nabla f(x) - \nabla f(y))^\top (x - y) = (x - y)^\top \nabla^2 f(\theta x + (1-\theta) y)(x - y) \geq 0$

说明当 $y \to x$ 时，$\nabla^2 f(x) \succeq 0$

例子

二次函数（Quadratic function）： $f(x) = \frac{1}{2}x^\top Px + q^\top x + r$，其中 P 是对称的 $n \times n$ 矩阵

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

当 $P \succeq 0$ 时函数是凸的。

问题： $f(x_1, x_2) = x_1^2 - 3x_1x_2 + x_2^2$ 是凸函数吗？令 $x_1 = u + v$, $x_2 = u - v$

最小二乘目标函数（Least-squares objective）： $f(x) = \|Ax - b\|^2$，其中 A 是 $m \times n$ 矩阵

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

对任意 A，f 是凸的。

线性分母下的二次函数（Quadratic-over-linear）： $f(x, y) = \frac{x^2}{y}$ （当 $y > 0$ 时是凸的）

验证函数的凸性（Verifying Convexity of a Function）

我们可以通过以下方式验证函数 f 是否为凸函数：

- 定义（Definition）
- **应用一些特殊判据（Applying some special criteria）**
  - 一阶条件（First-order conditions）
  - 二阶条件（Second-order conditions）
  - **化归为标量函数（Reduction to a scalar function）**
  - 上图法（Epigraph）
- 证明 f 是通过保持凸性的操作构造得到的

将凸函数限制在一条直线上的方法

化归为标量函数： $f : \mathbb{R}^n \to \mathbb{R}$ 是凸函数，当且仅当 $\operatorname{dom} f$ 是凸集，且如下定义的函数 $g : \mathbb{R} \to \mathbb{R}$ 是凸函数：

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

对于任意 $x \in \operatorname{dom} f$，$v \in \mathbb{R}^n$，$g(t)$ 是 t 上的凸函数。

通过一元函数的凸性验证多变量函数的凸性

例子： $f : \mathbb{S}^n \to \mathbb{R}$，$f(X) = -\ln\det X$，$\operatorname{dom} f = \mathbb{S}^n_{++}$

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

λ_i 是 $X^{-1/2}VX^{-1/2}$ 的特征值。由于 $g(t)$ 是凸的（对任意 V 且 $X \succ 0$），故 f 是凸函数。

验证函数的凸性（Verifying Convexity of a Function）

我们可以通过以下方式验证一个函数 f 是否为凸函数：

- 定义（Definition）
- **应用一些特殊判据（Applying some special criteria）**
  - 一阶条件（First-order conditions）
  - 二阶条件（Second-order conditions）
  - 化归为标量函数（Reduction to a scalar function）
  - **上图法（Epigraph）**
- 证明 f 是通过保持凸性的运算构造得到的

上图集与下水平集（Epigraph and Sublevel Set）

α-**下水平集（sublevel set）** 对于函数 $f : \mathbb{R}^n \to \mathbb{R}$ 定义为:

$$C_\alpha = \{x \in \mathrm{dom}\, f \mid f(x) \leq \alpha\}$$

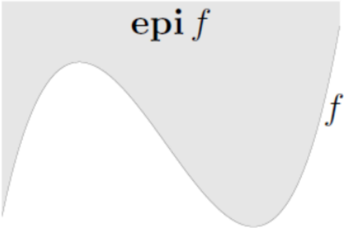
凸函数的下水平集是凸集（反过来不成立）。

上图集（Epigraph） 对于函数 $f : \mathbb{R}^n \to \mathbb{R}$

定义为:

$$\mathrm{epi}\, f = \{(x, t) \in \mathbb{R}^{n+1} \mid x \in \mathrm{dom}\, f,\ f(x) \leq t\}$$

函数 f 是凸的，当且仅当 $\mathrm{epi}\, f$ 是凸集。



目录

保持凸性的运算（Operations Preserving Convexity）

- 正数缩放（Positive Scaling）
- 求和（Sum）
- 与仿射函数的复合（Composition with affine function）
- 逐点最大值与上确界（Pointwise maximum and supremum）
- 函数复合（Composition）
- 最小化（Minimization）

缩放、求和、与仿射函数的复合

求和： 对于凸函数 f_1 和 f_2，它们的和 $f_1 + f_2$ 是凸的。该结论可扩展至积分形式：例如非负函数的卷积：$(f * g)(x) = \int_y f(y)g(x-y)dy$, 其中 $g(x) \geq 0$。

非负倍数： 若 f 是凸函数，$\lambda \geq 0$，则 λf 是凸函数。

与仿射函数的复合： 若 f 是凸函数，则 $f(Ax+b)$ 是凸函数。

例子：

- 线性不等式的对数障碍函数（Log-barrier for linear inequalities）：

$$f(x) = -\sum_{i=1}^{m} \ln(b_i - a_i^\top x), \quad \operatorname{dom} f = \{x \mid a_i^\top x < b_i,\ i = 1, \ldots, m\}$$

- 仿射函数的范数（Norm of affine function）：$f(x) = \|Ax + b\|$

逐点最大值（Pointwise Maximum）

若 $f_1, \ldots, f_m$ 是凸函数，则逐点最大值函数：

$$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$$

是凸函数。

证明提示： 使用定义或上图法（epigraph）

例子：

- 分段线性函数：$f(x) = \max_{i=1,\ldots,m}(a_i^\top x + b_i)$ 是凸函数。
- 向量 $x \in \mathbb{R}^n$ 中最大的 r 个分量之和：

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

  是凸的，其中 $x_{[i]}$ 表示 x 的第 i 大分量。

  证明提示： $f(x) = \max\{x_{i_1} + x_{i_2} + \cdots + x_{i_r} \mid 1 \le i_1 < \cdots < i_r \le n\}$

逐点上确界（Pointwise Supremum）

设 $f(x,y)$ 对于每个 $y \in \mathcal{A}$ 是关于 x 的凸函数。则定义：

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

是凸函数。

例子：

- 集合 $\mathcal{A}$ 的支撑函数（Support function）：

$$S_{\mathcal{C}}(x) = \sup_{z \in \mathcal{A}} z^\top x$$

- 到集合 $\mathcal{A}$ 中最远点的距离：

$$f(x) = \sup_{z \in \mathcal{A}} \|x - z\|$$

与标量函数的复合（Composition with Scalar Functions）

设 $g : \mathbb{R}^n \to \mathbb{R}$，$h : \mathbb{R} \to \mathbb{R}$，定义复合函数：

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

则 f 是凸函数，当满足以下条件之一：

- g 是凸函数，h 是**非减**且**凸**函数；
- g 是凹函数，h 是**非增**且**凸**函数。

例子：

- $e^{g(x)}$ 是凸函数，若 g 是凸函数；
- $\dfrac{1}{g(x)}$ 是凸函数，若 g 是凹函数且为正。

与向量函数的复合（Composition with Vector Functions）

设 $g:\mathbb{R}^n \to \mathbb{R}^k$，$h:\mathbb{R}^k \to \mathbb{R}$，定义复合函数：

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

则 f 是凸函数，当满足以下条件之一：

- g_i 是凸函数，h 在每个变量上都是**凸且非减**的；
- g_i 是凹函数，h 在每个变量上都是**凸且非增**的。

例子：

- $\log\left(\sum_{i=1}^{k} e^{g_i(x)}\right)$ 是凸函数，如果每个 g_i 是凸函数。

最小化（Minimization）

若 $f(x,y)$ 关于 (x,y) 是凸函数，且 C 是一个凸集合，则

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

是关于 x 的凸函数。定义域为：$\mathrm{dom}\, g = \{x \mid (x,y) \in \mathrm{dom}\, f \text{ for some } y \in C\}$

证明提示： 设 $\varepsilon > 0$ 是任意小的正数。对任意 $x_1, x_2 \in \mathbb{R}^n$ 和 $\theta \in (0,1)$，存在 $y_1, y_2 \in C$，使得：

$$f(x_1, y_1) \le g(x_1) + \varepsilon, \quad f(x_2, y_2) \le g(x_2) + \varepsilon$$

考虑：$(\theta x_1 + (1-\theta)x_2,\, \theta y_1 + (1-\theta)y_2)$

利用 f 和 C 的凸性完成证明。

另一种视角： 将 g 的上图集写作：

$$\mathrm{epi}\, g = \{(x,t) : (x,y,t) \in \mathrm{epi}\, f \text{ for some } y\}$$

例子（Example，集合的距离）： $\mathrm{dist}(x,C) = \inf_{y \in C} \|x - y\|$，当 C 是凸集时，$\mathrm{dist}(x,C)$ 是凸函数。

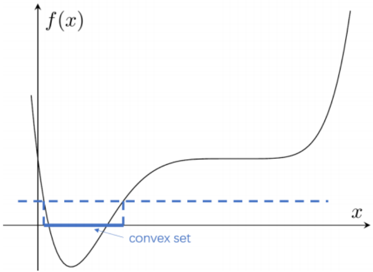
目录

拟凸性（Quasi-Convexity）

函数 $f : \mathbb{R}^n \to \mathbb{R}$ 是**拟凸函数（quasi-convex）**，当且仅当 $\mathrm{dom}\, f$ 是凸集，且如下的下水平集：

$$S_\alpha = \{x \in \mathrm{dom}\, f : f(x) \le \alpha\}$$

对所有 α 都是凸集。



另一种拟凸函数定义： 当 $\mathrm{dom}\, f$ 是凸的，且对于任意 $x, y \in \mathrm{dom}\, f$ 与 $\theta \in [0, 1]$，有：

$$f(\theta x + (1-\theta)y) \le \max\{f(x), f(y)\}$$

例子（Examples）

- $\sqrt{|x|}$ 在 $\mathbb{R}$ 上是拟凸的（quasi-convex）
- $\log(x)$ 在 $\mathbb{R}_+$ 上是拟凸的
- $f(x_1, x_2) = -x_1 x_2$ 在 $\mathbb{R}^2_{++}$ 上是拟凸的
- 距离比（Distance ratio）函数：

$$f(x) = \frac{\|x-a\|_2}{\|x-b\|_2}$$

在定义域 $\operatorname{dom} f = \{x : \|x-a\|_2 \leq \|x-b\|_2\}$ 上是拟凸的

可以构造凸函数 $\phi_\alpha(x) = \operatorname{dist}(x, \{z : f(z) \leq \alpha\})$，有 $f(x) \leq \alpha \Leftrightarrow \phi_\alpha(x) \leq 0$，其中 $f(x)$ 为拟凸函数。

共轭函数（The Conjugate Function）

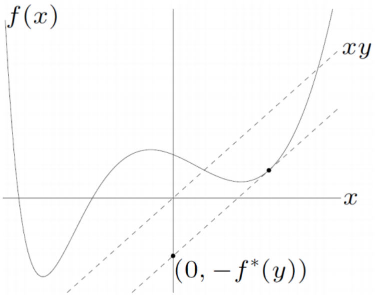
函数 f 的**共轭函数（conjugate）**定义为：

$$f^*(y) = \sup_{x \in \mathrm{dom}\, f} \left(y^T x - f(x) \right)$$

- f^* 是凸函数（即使 f 不是凸的）
- **Fenchel 不等式（Fenchel's inequality）：**

$$f(x) + f^*(y) \geq y^T x, \quad \forall x, y$$

- 如果 f 是凸函数且其上图集（epigraph）是闭集，则 $f^{**} = f$；否则，凸化非凸函数 f 将得到 f^{**}。
- 共轭函数在拉格朗日对偶问题（Lagrange dual）中非常有用



例子（Examples）

- 仿射函数（Affine function）： $f(x) = a^T x + b$

$$f^*(y) = \sup_x \left(y^T x - a^T x - b \right) = \begin{cases} -b & \text{若} y = a \\ \infty & \text{否则} \end{cases}$$

- 负对数函数（Negative logarithm）： $f(x) = -\log x$

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

- 严格凸二次函数（Strictly convex quadratic）： $f(x) = \frac{1}{2} x^T Q x$，其中 $Q \in \mathbb{S}^n_{++}$

$$f^*(y) = \sup_x \left(y^T x - \frac{1}{2} x^T Q x \right) = \frac{1}{2} y^T Q^{-1} y$$