凸函数

文再文

北京大学北京国际数学研究中心

教材《最优化：建模、算法与理论》配套电子教案

http://bicmr.pku.edu.cn/~wenzw/optbook.html

致谢：本教案由华奕轩协助准备

提纲

1 基础知识

2 凸函数的定义与性质

3 保凸的运算

4 凸函数的推广

梯度

定义 (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

$$\lim_{p\to 0}\frac{f(x+p)-f(x)-g^{\mathrm{T}}p}{\|p\|}=0,$$

其中$\|\cdot\|$ 是任意的向量范数，就称f 在点x 处**可微**（或Fréchet 可微）．此时g 称为f 在点x 处的**梯度**，记作$\nabla f(x)$．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．

若f 在点x 处的梯度存在，在定义式中令$p=\varepsilon e_i$，e_i是第i个分量为1的单位向量，可知$\nabla f(x)$ 的第i 个分量为$\dfrac{\partial f(x)}{\partial x_i}$．因此，

$$\nabla f(x)=\left[\frac{\partial f(x)}{\partial x_1},\frac{\partial f(x)}{\partial x_2},\cdots,\frac{\partial f(x)}{\partial x_n}\right]^{\mathrm{T}}.$$

海瑟矩阵

定义 (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:

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

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，可以证明此时海瑟矩阵是一个对称矩阵.

矩阵变量函数的导数

多元函数梯度的定义可以推广到变量是矩阵的情形．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，称G为f 在Fréchet 可微意义下的梯度．令$\frac{\partial f}{\partial x_{ij}}$ 表示f 关于x_{ij} 的偏导数．矩阵变量函数$f(X)$的梯度为

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

在实际应用中，矩阵Fréchet 可微的定义和使用往往比较繁琐，为此我们需要介绍另一种定义——Gâteaux 可微.

定义 (Gâteaux 可微)

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

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

则称f 关于X 是Gâteaux 可微的．满足上式的G称为f在X处在Gâteaux 可微意义下的梯度.

可以证明，当f是Fréchet 可微函数时，f也是Gâteaux 可微的，且这两种意义下的梯度相等.

- 线性函数：$f(X)=\mathrm{tr}(AX^{\mathrm{T}}B)$，其中$A\in\mathbb{R}^{p\times n}, B\in\mathbb{R}^{m\times p}, X\in\mathbb{R}^{m\times n}$ 对任意方向$V\in\mathbb{R}^{m\times n}$以及$t\in\mathbb{R}$，有

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

  因此，$\nabla f(X)=BA$．
- 二次函数：$f(X,Y)=\frac{1}{2}\|XY-A\|_F^2$，其中$(X,Y)\in\mathbb{R}^{m\times p}\times\mathbb{R}^{p\times n}$ 对变量Y，取任意方向V以及充分小的$t\in\mathbb{R}$，有

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

  由定义可知$\frac{\partial f}{\partial Y}=X^{\mathrm{T}}(XY-A)$．
  对变量X，同理可得$\frac{\partial f}{\partial X}=(XY-A)Y^{\mathrm{T}}$．

矩阵变量函数的导数

- *ln-det* 函数：$f(X)=\ln(\det(X))$, $X\in\mathcal{S}^n_{++}$，给定$X\succ 0$，对任意方向$V\in\mathcal{S}^n$ 以及$t\in\mathbb{R}$，我们有

$$
\begin{aligned}
&f(X+tV)-f(X)=\ln(\det(X+tV))-\ln(\det(X))\\
=&\ln(\det(X^{1/2}(I+tX^{-1/2}VX^{-1/2})X^{1/2}))-\ln(\det(X))\\
=&\ln(\det(I+tX^{-1/2}VX^{-1/2})).
\end{aligned}
$$

由于$X^{-1/2}VX^{-1/2}$是对称矩阵，所以它可以正交对角化，不妨设它的特征值为$\lambda_1,\lambda_2,\cdots,\lambda_n$，则

$$
\begin{aligned}
&\ln(\det(I+tX^{-1/2}VX^{-1/2}))=\ln\prod_{i=1}^{n}(1+t\lambda_i)\\
=&\sum_{i=1}^{n}\ln(1+t\lambda_i)=\sum_{i=1}^{n}t\lambda_i+\mathcal{O}(t^2)=t\,\mathrm{tr}(X^{-1/2}VX^{-1/2})+\mathcal{O}(t^2)\\
=&t\left\langle (X^{-1})^{\mathrm{T}},V\right\rangle+\mathcal{O}(t^2).
\end{aligned}
$$

因此，我们得到结论$\nabla f(X)=(X^{-1})^{\mathrm{T}}$.

广义实值函数与适当函数

定义 (广义实值函数)

令$\overline{\mathbb{R}} \overset{\text{def}}{=\!=} \mathbb{R} \cup \{\pm\infty\}$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**.

和数学分析一样，我们规定

$$-\infty < a < +\infty, \quad \forall\, a \in \mathbb{R}$$

$$(+\infty) + (+\infty) = +\infty, \quad +\infty + a = +\infty, \ \forall\, a \in \mathbb{R}.$$

定义 (适当函数)

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$f(x) < +\infty$，并且对任意的$x \in \mathcal{X}$，都有$f(x) > -\infty$，那么称函数f关于集合$\mathcal{X}$是适当的.

概括来说，适当函数f的特点是“至少有一处取值不为正无穷”，以及“处处取值不为负无穷”.

下水平集与上方图

定义 (α-下水平集)

对于广义实值函数$f:\mathbb{R}^n \to \overline{\mathbb{R}}$,

$$C_\alpha = \{x \,|\, f(x) \le \alpha\}$$

称为f 的α**-下水平集**.

定义 (上方图)

对于广义实值函数$f:\mathbb{R}^n \to \overline{\mathbb{R}}$,

$$\operatorname{epi} f = \{\,(x,t) \in \mathbb{R}^{n+1} \,|\, f(x) \le t\}$$

称为f 的**上方图**.

闭函数

定义 (闭函数)

设$f:\mathbb{R}^n\to\overline{\mathbb{R}}$为广义实值函数，若$\mathrm{epi}\,f$为闭集，则称$f$为**闭函数**.

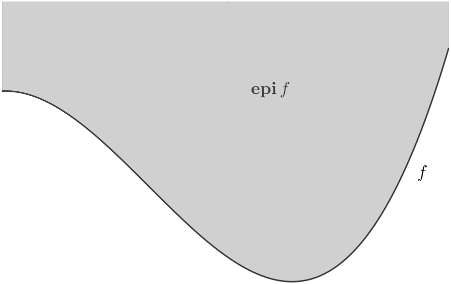


Figure: 函数f和其上方图$\mathrm{epi}\,f$

定义 (下半连续函数)

设广义实值函数$f: \mathbb{R}^n \to \overline{\mathbb{R}}$，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**.

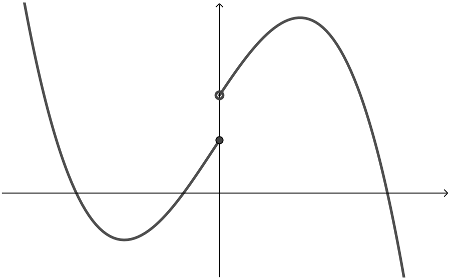


Figure: 下半连续函数$f(x)$

闭函数与下半连续函数

虽然表面上看这两种函数的定义方式截然不同，但闭函数和下半连续函数是等价的.

定理

设广义实值函数$f: \mathbb{R}^n \to \overline{\mathbb{R}}$，则以下命题等价：

1. $f(x)$的任意α-下水平集都是闭集；
2. $f(x)$是下半连续的；
3. $f(x)$是闭函数.

闭函数与下半连续函数

闭（下半连续）函数间的简单运算会保持原有性质:

- 加法: 若f 与g 均为适当的闭（下半连续）函数，并且$\mathbf{dom}\, f \cap \mathbf{dom}\, g \neq \emptyset$，则$f + g$ 也是闭（下半连续）函数. 其中适当函数的条件是为了避免出现未定式$(-\infty) + (+\infty)$ 的情况；
- 仿射映射的复合: 若f 为闭（下半连续）函数, 则$f(Ax + b)$ 也为闭（下半连续）函数；
- 取上确界: 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.

提纲

1 基础知识

2 凸函数的定义与性质

3 保凸的运算

4 凸函数的推广

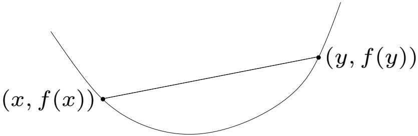
凸函数的定义

定义 (凸函数)

$f : \mathbb{R}^n \to \mathbb{R}$ 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f 是凸函数，则$-f$ 是凹函数
- 若对所有$x, y \in \mathrm{dom}\, f$ ，$x \ne y$, $0 < \theta < 1$，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

凸函数:

- 仿射函数: 对任意$a, b \in \mathbb{R}$，$ax + b$ 是$\mathbb{R}$ 上的凸函数
- 指数函数: 对任意$a \in \mathbb{R}$，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

凹函数:

- 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。所有的范数都是凸函数。

欧氏空间$\mathbb{R}^n$中的例子

- 仿射函数: $f(x) = a^T x + b$
- 范数: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$ $(p \geq 1)$；特别地，$\|x\|_\infty = \max_k |x_k|$

矩阵空间$\mathbb{R}^{m\times n}$中的例子

- 仿射函数:

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

- 谱范数:

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

强凸函数

- 定义1: 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**. 为了方便我们也称$f(x)$为m-强凸函数.

- 定义2: 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

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

则称$f(x)$为强凸函数，其中m为强凸参数.

- 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

将函数限制在任意直线上，然后判断对应的一维函数是否是凸的.

定理

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

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

例：$f(X) = -\log\det X$ 是凸函数，其中$\mathbf{dom}\, f = \mathbb{S}^n_{++}$. 任取$X \succ 0$ 以及方向$V \in \mathbb{S}^n$，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.

凸函数判定定理

Proof.

必要性:设$f(x)$是凸函数，要证$g(t)=f(x+tv)$是凸函数．先说明$\mathrm{dom}g$是凸集．对任意的$t_1,t_2\in\mathrm{dom}g$以及$\theta\in(0,1)$，

$$x+t_1v\in\mathrm{dom}f, x+t_2v\in\mathrm{dom}f$$

由$\mathrm{dom}f$是凸集可知$x+(\theta t_1+(1-\theta)t_2)v\in\mathrm{dom}f$,
这说明$\theta t_1+(1-\theta)t_2\in\mathrm{dom}g$，即$\mathrm{dom}g$是凸集．此外，我们有

$$\begin{aligned}
g(\theta t_1+(1-\theta)t_2) &= f(x+(\theta t_1+(1-\theta)t_2)v)\\
&= f(\theta(x+t_1v)+(1-\theta)(x+t_2v))\\
&\le \theta f(x+t_1v)+(1-\theta)f(x+t_2v)\\
&= \theta g(t_1)+(1-\theta)g(t_2).
\end{aligned}$$

结合以上两点得到函数$g(t)$是凸函数．

Proof.

充分性:先说明dom$f$是凸集，取$v = y - x$，以及$t_1 = 0, t_2 = 1$，由dom$g$是凸集可知$\theta \cdot 0 + (1-\theta) \cdot 1 \in \mathrm{dom}g$，即$\theta x + (1-\theta) y \in \mathrm{dom}f$，这说明dom$f$是凸集．再根据$g(t) = f(x + tv)$的凸性，我们有

$$\begin{aligned} g(1-\theta) &= g(\theta t_1 + (1-\theta) t_2) \\ &\le \theta g(t_1) + (1-\theta) g(t_2) \\ &= \theta g(0) + (1-\theta) g(1) \\ &= \theta f(x) + (1-\theta) f(y). \end{aligned}$$

而等式左边有

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

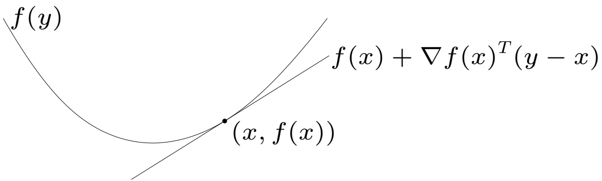
这说明$f(x)$是凸函数． □

一阶条件

定理

一阶条件：对于定义在凸集上的可微函数f，f 是凸函数当且仅当

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



几何直观：f的一阶逼近始终在f的图像下方

Proof.

必要性：设f是凸函数，则对于任意的$x, y \in \mathrm{dom} f$以及$t \in (0,1)$，有

$$tf(y) + (1-t)f(x) \geq f(x + t(y-x)).$$

将上式移项，两边同时除以t，注意$t > 0$，则

$$f(y) - f(x) \geq \frac{f(x + t(y-x)) - f(x)}{t}.$$

令$t \to 0$，由极限保号性可得

$$f(y) - f(x) \geq \lim_{t \to 0} \frac{f(x + t(y-x)) - f(x)}{t} = \nabla f(x)^{\mathrm{T}}(y-x).$$

这里最后一个等式成立是由于方向导数的性质.

Proof.

充分性：对任意的$x, y \in \mathrm{dom} f$以及任意的$t \in (0, 1)$，定义$z = tx + (1-t)y$，应用两次一阶条件我们有

$$f(x) \geq f(z) + \nabla f(z)^{\mathrm{T}}(x - z),$$
$$f(y) \geq f(z) + \nabla f(z)^{\mathrm{T}}(y - z).$$

将上述第一个不等式两边同时乘t，第二个不等式两边同时乘$1-t$，相加得

$$tf(x) + (1-t)f(y) \geq f(z) + 0.$$

这正是凸函数的定义，因此充分性成立. □

梯度单调性

定理

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

$$(\nabla f(x) - \nabla f(y))^{\mathrm{T}}(x - y) \geq 0, \quad \forall\, x, y \in \mathrm{dom} f.$$

Proof.

必要性：若f可微且为凸函数，根据一阶条件，我们有

$$\begin{aligned} f(y) &\geq f(x) + \nabla f(x)^{\mathrm{T}}(y - x), \\ f(x) &\geq f(y) + \nabla f(y)^{\mathrm{T}}(x - y). \end{aligned}$$

将两式不等号左右两边相加即可得到结论.

梯度单调性

Proof.

充分性：若∇f为单调映射，构造一元辅助函数

$$g(t)=f(x+t(y-x)),\quad g'(t)=\nabla f(x+t(y-x))^{\mathrm{T}}(y-x)$$

由∇f的单调性可知$g'(t)\geq g'(0),\forall\, t\geq 0$．因此

$$\begin{aligned} f(y)=g(1)&=g(0)+\int_0^1 g'(t)\mathrm{d}t \\ &\geq g(0)+g'(0)=f(x)+\nabla f(x)^{\mathrm{T}}(y-x). \end{aligned}$$

□

上方图

定理

函数$f(x)$为凸函数当且仅当其上方图epi$f$是凸集.

Proof.

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

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

$\square$

二阶条件

定理

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

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

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

例：二次函数$f(x) = (1/2)x^T Px + q^T x + r$ (其中$P \in \mathbb{S}^n$)

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

f 是凸函数当且仅当$P \succeq 0$

二阶条件

Proof.

必要性：反设$f(x)$在点x处的海瑟矩阵$\nabla^2 f(x) \not\succeq 0$，即存在非零向量$v \in \mathbb{R}^n$使得$v^{\mathrm{T}}\nabla^2 f(x)v < 0$. 根据佩亚诺（Peano）余项的泰勒展开，

$$f(x+tv) = f(x) + t\nabla f(x)^{\mathrm{T}}v + \frac{t^2}{2}v^{\mathrm{T}}\nabla^2 f(x)v + o(t^2).$$

移项后等式两边同时除以t^2，

$$\frac{f(x+tv) - f(x) - t\nabla f(x)^{\mathrm{T}}v}{t^2} = \frac{1}{2}v^{\mathrm{T}}\nabla^2 f(x)v + o(1).$$

当t充分小时，

$$\frac{f(x+tv) - f(x) - t\nabla f(x)^{\mathrm{T}}v}{t^2} < 0,$$

这显然和一阶条件矛盾，因此必有$\nabla^2 f(x) \succeq 0$成立.

Proof.

充分性：设$f(x)$满足二阶条件$\nabla^2 f(x) \succeq 0$，对任意$x, y \in \mathrm{dom} f$，根据泰勒展开，

$$f(y) = f(x) + \nabla f(x)^{\mathrm{T}}(y-x) + \frac{1}{2}(y-x)^{\mathrm{T}}\nabla^2 f(x+t(y-x))(y-x),$$

其中$t \in (0,1)$是和x, y有关的常数．由半正定性可知对任意$x, y \in \mathrm{dom} f$有

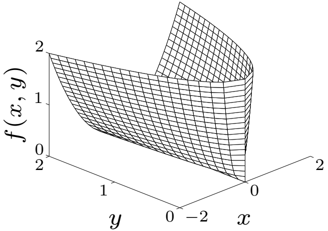
$$f(y) \geq f(x) + \nabla f(x)^{\mathrm{T}}(y-x).$$

由凸函数判定的一阶条件知f为凸函数．进一步，若$\nabla^2 f(x) \succ 0$，上式中不等号严格成立（$x \neq y$）．利用一阶条件的充分性的证明过程可得$f(x)$为严格凸函数． □

最小二乘函数：$f(x) = \|Ax - b\|_2^2$

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

对任意A，f都是凸函数



quadratic-over-linear函数: $f(x,y) = x^2/y$

$$\nabla^2 f(x,y) = \frac{2}{y^3}\begin{bmatrix} y \\ -x \end{bmatrix}\begin{bmatrix} y \\ -x \end{bmatrix}^T \succeq 0$$

是区域$\{(x,y) \mid y > 0\}$上的凸函数

log-sum-exp函数: $f(x) = \log \sum_{k=1}^{n} \exp x_k$ 是凸函数

$$\nabla^2 f(x) = \frac{1}{\mathbf{1}^T z}\operatorname{diag}(z) - \frac{1}{(\mathbf{1}^T z)^2} z z^T \quad (z_k = \exp x_k)$$

要证明$\nabla^2 f(x) \succeq 0$，我们只需证明对任意v, $v^T \nabla^2 f(x) v \geq 0$，即

$$v^T \nabla^2 f(x) v = \frac{(\sum_k z_k v_k^2)(\sum_k z_k) - (\sum_k v_k z_k)^2}{(\sum_k z_k)^2} \geq 0$$

由柯西不等式，得$(\sum_k v_k z_k)^2 \leq (\sum_k z_k v_k^2)(\sum_k z_k)$，因此$f$是凸函数

几何平均: $f(x) = (\prod_{k=1}^{n} x_k)^{1/n} (x \in \mathbb{R}_{++}^n)$ 是凹函数

Jensen不等式

基础Jensen不等式: 设f 是凸函数，则对于$0 \le \theta \le 1$,

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

概率Jensen 不等式: 设f 是凸函数，则对任意随机变量z

$$f(\mathbf{E}z) \le \mathbf{E}f(z)$$

基础Jensen不等式可以视为概率Jensen 不等式在两点分布下的特殊情况

$$\mathrm{prob}(z = x) = \theta, \quad \mathrm{prob}(z = y) = 1 - \theta$$

提纲

1 基础知识

2 凸函数的定义与性质

3 保凸的运算

4 凸函数的推广

保凸的运算

验证一个函数f 是凸函数的方法：

1. 用定义验证（通常将函数限制在一条直线上）
2. 利用一阶条件、二阶条件
3. 直接研究f 的上方图$\mathrm{epi}\, f$
4. 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
   - 取下确界
   - 透视函数

非负加权和与仿射函数的复合

非负数乘: 若f 是凸函数，则αf 是凸函数，其中$\alpha \geq 0$.

求和: 若f_1, f_2 是凸函数，则$f_1 + f_2$ 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

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

- 仿射函数的（任意）范数：$f(x) = \|Ax + b\|$

若$f_1, \ldots, f_m$ 是凸函数，则$f(x) = \max\{f_1(x), \ldots, f_m(x)\}$ 是凸函数

例子

- 分段线性函数：$f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$ 是凸函数
- $x \in \mathbb{R}^n$的前r 个最大分量之和：

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

是凸函数($x_{[i]}$ 为x 的从大到小排列的第i 个分量)

事实上，$f(x)$可以写成如下多个线性函数取最大值的形式：

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

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

例子

- 集合C的支撑函数：$S_C(x) = \sup_{y \in C} y^T x$ 是凸函数
- 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$X \in \mathbb{S}^n$的最大特征值

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

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

- 对$n = 1$，g, h均可微的情形，我们给出简证

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

- 注意：必须是$\tilde{h}$满足单调不减/不增的条件；如果仅是h满足单调不减/不增的条件，存在反例

推论

- 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

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

对$n=1$，g,h均可微的情形，我们给出简证

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

推论

- 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.

例子

- 考虑函数$f(x,y) = x^{\mathrm{T}}Ax + 2x^{\mathrm{T}}By + y^{\mathrm{T}}Cy$，海瑟矩阵满足

$$\begin{bmatrix} A & B \\ B^{\mathrm{T}} & C \end{bmatrix} \succeq 0, \quad C \succ 0,$$

则$f(x,y)$ 为凸函数.对y 求最小值得

$$g(x) = \inf_{y} f(x,y) = x^{\mathrm{T}}(A - BC^{-1}B^{\mathrm{T}})x,$$

因此g 是凸函数.进一步地，A 的Schur 补$A - BC^{-1}B^{\mathrm{T}} \succeq 0$

- 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** $g:\mathbb{R}^n\times\mathbb{R}\to\mathbb{R}$,

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

若f 是凸函数，则g 是凸函数.

例子

- $f(x)=x^Tx$是凸函数，因此$g(x,t)=x^Tx/t$ 是区域$\{(x,t)\mid t>0\}$上的凸函数
- $f(x)=-\log x$是凸函数，因此相对熵函数$g(x,t)=t\log t-t\log x$是$\mathbb{R}^2_{++}$上的凸函数
- 若f 是凸函数，那么

$$g(x)=(c^Tx+d)f\left((Ax+b)/(c^Tx+d)\right)$$

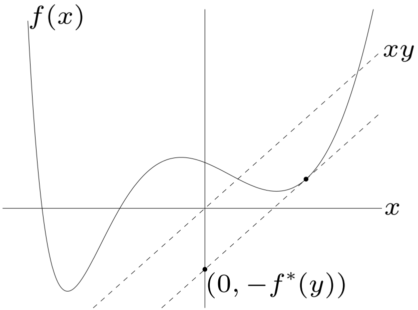
是区域$\{x|c^Tx+d>0,(Ax+b)/(c^Tx+d)\in\operatorname{dom} f\}$上的凸函数

共轭函数

适当函数f的**共轭函数**定义为

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

- 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$f(x) = (1/2)x^T Qx,\ Q \in \mathbb{S}^n_{++}$

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

提纲

1 基础知识

2 凸函数的定义与性质

3 保凸的运算

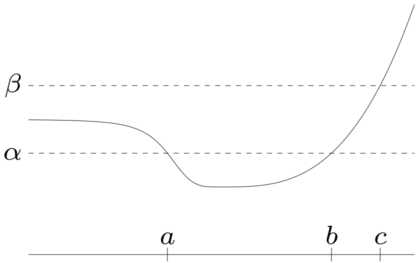
4 凸函数的推广

拟凸函数

$f : \mathbb{R}^n \to \mathbb{R}$ 称为拟凸的，如果dom f 是凸集，并且下水平集

$$S_\alpha = \{x \in \mathbf{dom}\, f | f(x) \leq \alpha\}$$

对任意α都是凸的



- 若f 是拟凸的，则称$-f$是拟凹的
- 若f既是拟凸的，又是拟凹的，则称f是拟线性的

- $\sqrt{|x|}$ 是$\mathbb{R}$上的拟凸函数
- $\mathrm{ceil}(x) = \inf\{z \in \mathbb{Z} | z \geq x\}$ 是拟线性的
- $\log x$ 是$\mathbb{R}_{++}$上的拟线性函数
- $f(x_1, x_2) = x_1 x_2$ 是$\mathbb{R}^2_{++}$上的拟凹函数
- 分式线性函数

$$f(x) = \frac{a^T x + b}{c^T x + d}, \quad \mathrm{dom}\, f = \{x | c^T x + d > 0\}$$

 是拟线性的

- 距离比值函数

$$f(x) = \frac{\|x - a\|_2}{\|x - b\|_2}, \quad \mathrm{dom}\, f = \{x | \ \|x - a\|_2 \leq \|x - b\|_2\}$$

 是拟凸的

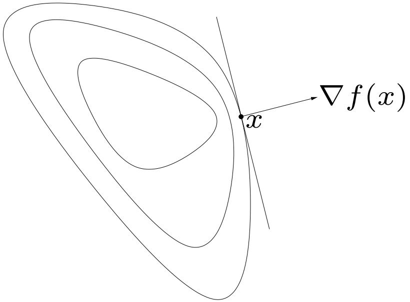
拟凸函数的性质

类Jensen不等式: 对拟凸函数f

$$0 \le \theta \le 1 \quad \Longrightarrow \quad f(\theta x + (1-\theta)y) \le \max\{f(x), f(y)\}$$

一阶条件: 定义在凸集上的可微函数f 是拟凸的，当且仅当

$$f(y) \le f(x) \quad \Longrightarrow \quad \nabla f(x)^T (y-x) \le 0$$



注：拟凸函数的和不一定是拟凸函数

对数凸函数

如果正值函数f满足$\log f$ 是凸函数，则f称为对数凸函数，即

$$f(\theta x+(1-\theta)y)\le f(x)^{\theta}f(y)^{1-\theta}\quad \text{for } 0\le\theta\le 1.$$

如果$\log f$ 是凹函数，则f 称为对数凹函数，

- 幂函数：当$a\le 0$时，x^a 是$\mathbb{R}_{++}$ 上的对数凸函数；当$a\ge 0$，x^a 是$\mathbb{R}_{++}$ 上的对数凹函数
- 许多常见的概率密度函数是对数凹函数，例如正态分布

$$f(x)=\frac{1}{\sqrt{(2\pi)^n\det\Sigma}}e^{-\frac{1}{2}(x-\bar{x})^T\Sigma^{-1}(x-\bar{x})}$$

- 高斯分布的累计分布函数Φ 是对数凹函数

$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}du$$

对数凸、凹函数的性质

- 定义在凸集上的二阶可微函数f是对数凹的，当且仅当

$$f(x)\nabla^2 f(x) \preceq \nabla f(x)\nabla f(x)^T$$

  对任意$x \in \mathrm{dom}\, f$成立

- 对数凹函数的乘积仍为对数凹函数
- 对数凹函数的和不一定为对数凹函数
- 若$f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ 是对数凹函数，那么

$$g(x) = \int f(x, y)dy$$

  是对数凹函数

广义不等式意义下的凸函数

$f : \mathbb{R}^n \to \mathbb{R}^m$ 称为K-凸函数: 如果$\mathbf{dom}\, f$ 是凸集，并且

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

对任意$x, y \in \mathbf{dom}\, f$, $0 \le \theta \le 1$成立

例子 $f : \mathbb{S}^m \to \mathbb{S}^m, f(X) = X^2$ 是$\mathbb{S}^m_+$-凸函数

证明：对固定的$z \in \mathbb{R}^m$, $z^T X^2 z = \|Xz\|_2^2$ 关于X是凸函数，即

$$z^T(\theta X + (1-\theta)Y)^2 z \le \theta z^T X^2 z + (1-\theta) z^T Y^2 z$$

对任意$X, Y \in \mathbb{S}^m$, $0 \le \theta \le 1$成立.

因此$(\theta X + (1-\theta)Y)^2 \preceq \theta X^2 + (1-\theta)Y^2$