凸函数

王正强

重庆邮电大学通信学院

提纲

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，

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

Hessian矩阵

定义 (Hessian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*Hessian*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，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.

矩阵变量函数的导数

- 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".

下水平集与上镜图

定义 (α-下水平集)

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

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

称为f 的α-下水平集.

定义 (上镜图)

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

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

称为f 的上镜图.

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．

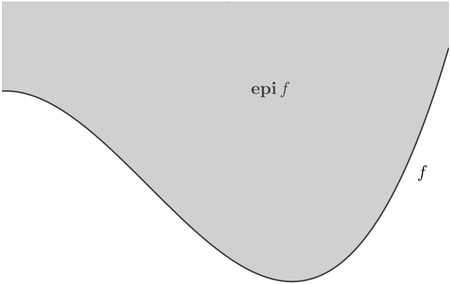


Figure: 函数f和其上镜图$\mathrm{epi}\, f$

R上的例子

- 函数$f: \mathbf{R} \to \mathbf{R}$，$f(x) = x\log x$，$\mathrm{dom} f = \mathbf{R}_{++}$不是闭函数。
- 函数$f: \mathbf{R} \to \mathbf{R}$，

$$f(x) = \begin{cases} x\log x & x > 0 \\ 0 & x = 0 \end{cases}$$

在$\mathrm{dom} f = \mathbf{R}_{+}$是闭函数。

- 函数$f(x) = -\log x$，$\mathrm{dom} f = \mathbf{R}_{++}$是闭函数。

下半连续函数

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

$$\liminf_{y\to x} f(y)\ge f(x),$$

则$f(x)$为下半连续函数.

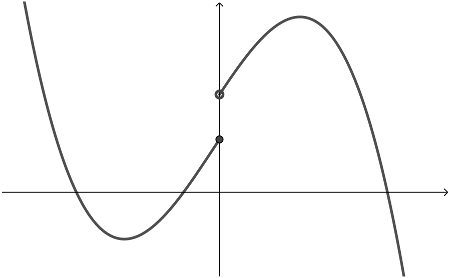


Figure: 下半连续函数$f(x)$

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.

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

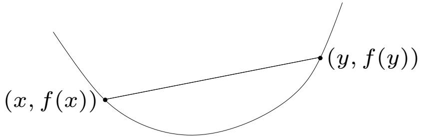
凸函数的定义

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

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

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

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

  则称f 是严格凸函数

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.

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

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

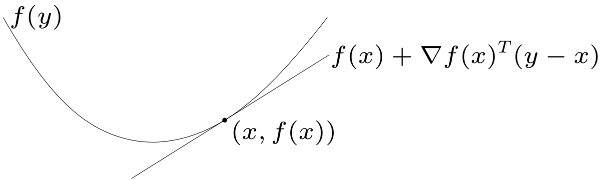
这说明$f(x)$是凸函数． □

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

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



几何直观：f的一阶逼近始终在f的图像下方

一阶条件说明从一个凸函数的局部信息（即它在某点的函数值及导数），可以得到它的全局下估计（优化问题中SCA应用）。

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$x, y \in \mathbf{dom} f$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

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

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. $\square$

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.

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

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

□

例如：$f(x)=x^2, f(x,y)=x^2+4y^2$.

上境图

定理

函数$f(x)$为凸函数当且仅当其上镜图epi$f$是凸集.

Proof.

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

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

故$(tx_1 + (1-t)x_2, ty_1 + (1-t)y_2) \in \text{epi} f, t \in [0, 1]$.
充分性：若epi$f$是凸集，则对任意$x_1, x_2 \in \text{dom}\, f, t \in [0, 1]$,

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

□

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

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

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

例：二次函数$f(x) = (1/2)x^T P x + q^T x + r$ (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$P \succeq 0$
注：在判断函数凸性和凹性时，利用一阶或二阶条件时，$domf$是凸集这个前提条件必须满足。例如$f(x) = 1/x^2$, $domf = \{x \in R | x \neq 0\}$

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． □

功率控制问题

如下优化问题是否为凸优化问题(由于约束是凸集，只需要判断目标函数是否为凹函数)?

$$\text{maximize} \sum_{i=1}^{n} \frac{L w_i h_i p_i}{\sum_{k \neq i} h_k p_k + L h_i p_i + \sigma^2} \quad (1)$$

$$\text{subject to} \sum_{j=1}^{n} g_j p_j \leq T, \quad (2)$$

$$p_j \geq 0, j = 1, \cdots, n. \quad (3)$$

问题(1)-(3) 中的优化变量为$p_1, \cdots, p_n$，其中$L > 1$, $T > 0$, $\sigma^2 > 0$和$h_i > 0$, $g_i > 0$都是常数，n为$n \geq 2$的正整数。

问题来源：

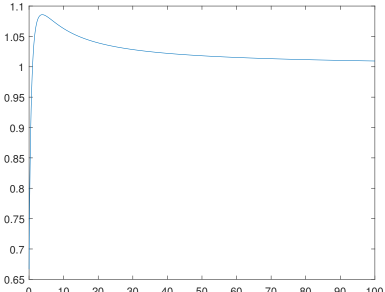
Zhengqiang Wang, Lingge Jiang, Chen He. Optimal Price-Based Power Control Algorithm in Cognitive Radio Networks. IEEE Transactions on Wireless Communications, 2014, 13(11):5909–5920, Nov. 2014.

非凸优化问题！因为目标函数不是凹函数。

$$f(p_1,\cdots,p_n)=\sum_{i=1}^{n}\frac{Lw_ih_ip_i}{\sum_{k\neq i}h_kp_k+Lh_ip_i+\sigma^2}$$

目标函数非凹函数，取一个反例。取$n=2,L=2,h_1=1,h_2=1,$ $w_1=1,w_2=1,\sigma^2=1,$ 此时目标函数为：$f(p_1,p_2)=\frac{2p_1}{p_2+2p_1+1}+\frac{2p_2}{p_1+2p_2+1},$

固定取$p_2=1,f(p_1,1)=\frac{2p_1}{2+2p_1}+\frac{2}{p_1+3}$



变量替换将非凸优化优化问题转化为凸优化问题

Lemma

Let $a_i = \frac{h_i p_i}{\sum\limits_{j=1}^{n} h_j p_j + \sigma^2} (i = 1, \cdots, n)$, *the revenue of the PU satisfies the following equation:*

$$\sum_{i=1}^{n} \frac{L w_i h_i p_i}{\sum_{k \neq i} h_k p_k + \sigma^2 + L h_i p_i} = \sum_{i=1}^{n} \frac{L w_i a_i}{((L-1)a_i + 1)} \tag{4}$$

and the transmit power of i*-th SU satisfies the following equation:*

$$p_i = \frac{\sigma^2 a_i}{\left(1 - \sum\limits_{i=1}^{n} a_i\right) h_i}, \forall i \in \{1, \ldots, n\}. \tag{5}$$

等价凸优化问题

Using lemma 1 and $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} h_i p_i / \left(\sum_{i=1}^{n} h_i p_i + \sigma^2 \right) < 1$, (1)-(3) are equivalent to the following problem:

$$\text{maximize} \sum_{i=1}^{n} \frac{L w_i a_i}{((L-1)a_i + 1)} \tag{6}$$

$$\text{subject to } \sigma^2 \sum_{i=1}^{n} \frac{g_i a_i}{h_i} + T \sum_{i=1}^{n} a_i \leq T, \tag{7}$$

$$a_i \geq 0, i = 1, \cdots, n. \tag{8}$$

The optimization variables for (6)-(8) are $a_1, \cdots, a_n$. The second derivative of the above optimization problem with respect to variable a_i is given by $-\frac{2Lw_i(L-1)}{(a_iL-a_i+1)^3}$. $-\frac{2Lw_i(L-1)}{(a_iL-a_i+1)^3} < 0$ is valid because $L > 1$ and $a_i \geq 0$ are satisfied. The hessian matrix of the objective function is positive definite matrix. Therefore, (6)-(8) is a convex programming.

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，$\|x\|_\infty = \max_k |x_k|$
- 最大值函数: $f(x) = \max\{x_1, \cdots, x_n\}$在$\mathbf{R^n}$上是凸的。

矩阵空间$\mathbb{R}^{m \times n}$中的例子

- 仿射函数:

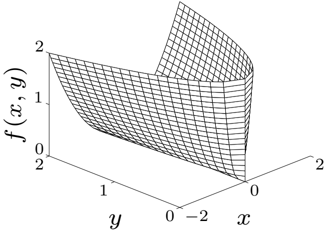
$$f(X) = \mathrm{tr}(A^T X) + b = \sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij} X_{ij} + b$$

二阶条件的应用

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: $f(x) = \log(e^{x_1} + \cdots + e^{x_n}), x \in R^n$是凸函数。
证明：观察f对各个分量的一二阶偏导数，

知 $\frac{\partial f}{\partial x_i} = \frac{e^{x_i}}{e^{x_1}+\cdots+e^{x_n}} \Rightarrow \frac{\partial^2 f}{\partial x_i \partial x_j} \begin{cases} \frac{-e^{x_i} \cdot e^{x_j}}{(e^{x_1}+\cdots+e^{x_n})^2}, & i \neq j \\ \frac{-e^{x_i} \cdot e^{x_i} + e^{x_i} \cdot (e^{x_1}+\cdots+e^{x_n})}{(e^{x_1}+\cdots+e^{x_n})^2}, & i = j \end{cases}$

令$z = [e^{x_1}, \cdots, e^{x_n}]^T$，则$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{-z_i z_j}{(1^T z)^2}(i \neq j), \frac{\partial^2 f}{\partial^2 x_i} = \frac{-z_i z_j + z_i 1^T z}{(1^T z)^2}$

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

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

几何平均函数$f(x)=\left(\prod_{i=1}^n x_i\right)^{1/n}$在定义域$\mathrm{dom}\, f=\mathbf{R}_{++}$是凹的。
其Hessian矩阵$\nabla^2 f(x)$可以通过下面两个式子给出：

$$\frac{\partial^2 f(x)}{\partial x_k^2}=-(n-1)\frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{n^2x_k^2}\qquad \frac{\partial^2 f(x)}{\partial x_k\partial x_l}=\frac{\left(\prod_{i=1}^n x_i\right)^{1/n}}{n^2x_kx_l}\quad \forall k\neq l$$

因此$\nabla^2 f(x)$具有如下表达式：

$$\nabla^2 f(x)=-\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\,\mathrm{diag}\left(1/x_1^2,\cdots,1/x_n^2\right)-qq^T\right)$$

其中，$q_i=1/x_i$。我们需要证明$\nabla^2 f(x)\preceq 0$，即对任意向量v，有

$$v^T\nabla^2 f(x)v=-\frac{\prod_{i=1}^n x_i^{1/n}}{n^2}\left(n\sum_{i=1}^n v_i^2/x_i^2-\left(\sum_{i=1}^n v_i/x_i\right)^2\right)\leqslant 0$$

这同样可以应用Cauchy-Schwarz不等式$\left(a^Ta\right)\left(b^Tb\right)\geqslant\left(a^Tb\right)^2$得到，
只需令向量$a=1$，向量b的分量$b_i=v_i/x_i$。

下水平集(sublevel set)

定义

定义　集合

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

称为f的α下水平集。

定理

凸函数的任一下水平集均为凸集。

- 任一下水平集均为凸集的函数不一定为凸函数。

定理

凹函数的任一上水平集均为凸集。

定义

定义　函数f的α-上水平集：$\{x \in \mathrm{dom} f \mid f(x) \ge \alpha\}$

例3.3 $x \in \mathbf{R}^n_+$的几何均值和算术均值分别为:

$$G(x) = \left(\prod_{i=1}^{n} x_i\right)^{1/n}, \quad A(x) = \frac{1}{n}\sum_{i=1}^{n} x_i$$

在G中定义$0^{1/n} = 0$。算术几何平均不等式表明$G(x) \leq A(x)$。
假设$0 \leq \alpha \leq 1$,并考虑该集合:

$$\{x \in \mathbf{R}^n_+ \mid G(x) \geq \alpha A(x)\}$$

几何平均值大于等于因子α乘以算术平均值的向量的集合。这个集合是凸的,因为它是凹函数$G(x) - \alpha A(x)$的0-上水平集。事实上,这个集合是正齐次的,所以它是一个凸锥。

函数上境图(Epigraph)

定义

集合$\mathrm{epi}f = \{(x, t) | x \in \mathrm{dom}\, f, f(x) \le t\}$为函数$f$的上镜图。

- 定理：函数f为凸函数当且仅当f的上境图为凸集。
- 定理：函数f为凹函数当且仅当f的亚图为凸集

$$\mathrm{hypo}f = \{(x, t) | x \in \mathrm{dom}\, f, f(x) \ge t\}$$

上述定理将凸函数与凸集联系起来。

例3.4 矩阵分式函数。函数$f: \mathbf{R}^n \times \mathbf{S}^n \to \mathbf{R}$定义为：

$$f(x,Y) = x^T Y^{-1} x$$

该函数在定义域$f = \mathbf{R}^n \times \mathbf{S}^n_{++}$上是凸的。（这推广了线性函数上的二次函数$f(x,y) = x^2/y$，其定义域为$f = \mathbf{R} \times \mathbf{R}_{++}$）确定$f$的凸性的一个简单方法是通过其上镜图：

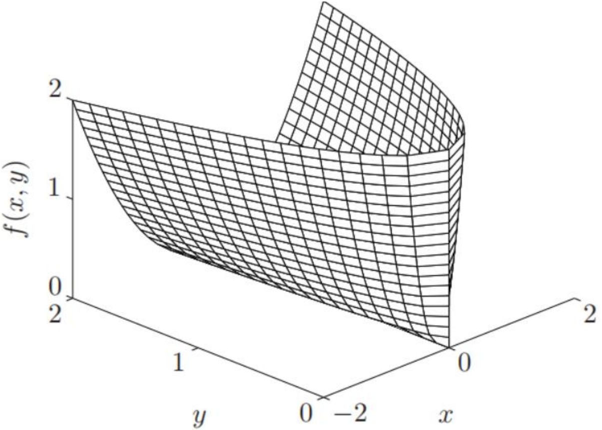
$$\begin{aligned} \text{epi} f &= \left\{(x,Y,t) \mid Y \succ 0, x^T Y^{-1} x \leq t\right\} \\ &= \left\{(x,Y,t) \mid \begin{bmatrix} Y & x \\ x^T & t \end{bmatrix} \succeq 0, Y \succ 0\right\} \end{aligned}$$

通过舒尔补条件判断分块矩阵的半正定性（见附录A.5.5）。最后一个条件关于(x,Y,t)中的线性矩阵不等式，因此epi$f$是凸的。
对于特殊情况$n=1$，矩阵分数函数化简为二次复线性函数x^2/y，相应的线性矩阵不等式表示为：

$$\begin{bmatrix} y & x \\ x & t \end{bmatrix} \succeq 0, \quad y > 0$$

函数上境图(Epigraph)

$f(x,y)=x^2/y$的图像



关于凸函数的很多结果可以从几何的角度利用上镜图并结合凸集的一些结论来证明（或理解），作为一个例子，考虑凸函数的一阶条件：

$$f(y) \geqslant f(x) + \nabla f(x)^T (y - x)$$

其中函数f是凸的，$x, y \in \textbf{dom}\, f$。我们可以利用$\textbf{epi}\, f$从几何角度理解上述基本不等式。如果$(y, t) \in \textbf{epi}\, f$，有

$$t \geqslant f(y) \geqslant f(x) + \nabla f(x)^T (y - x).$$

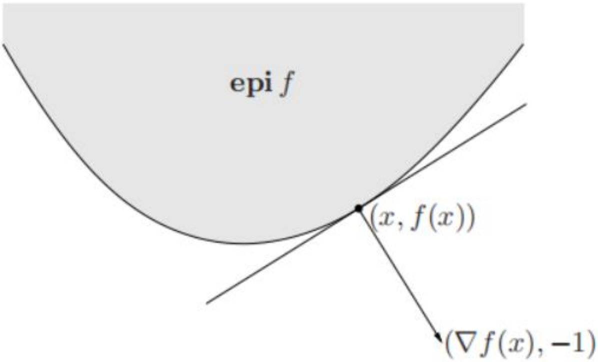
上式可以描述为

$$(y, t) \in \textbf{epi}\, f \Longrightarrow \begin{bmatrix} \nabla f(x) \\ -1 \end{bmatrix}^T \left(\begin{bmatrix} y \\ t \end{bmatrix} - \begin{bmatrix} x \\ f(x) \end{bmatrix} \right) \leq 0.$$

这意味着法向量为$(\nabla f(x), -1)$的超平面在边界点$(x, f(x))$支撑着$\textbf{epi}\, f$。

函数上境图(Epigraph)

对于一个可微凸函数f，向量$(\nabla f(x), -1)$定义了一个对于$f(x)$在x处的上境图的支撑超平面。



Jensen不等式极其扩展

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

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

概率**Jensen** 不等式: 设f 是凸函数，则对任意随机变量z

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

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

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

Jensen不等式扩展至更多点的凸组合。

- f为凸函数，则有

$$f\left(\theta_1 x_1+\ldots+\theta_n x_n\right) \leq \theta_1 f\left(x_1\right)+\ldots+\theta_n f\left(x_n\right)$$

其中$0 \leq \theta_i \leq 1, \theta_1+\ldots+\theta_n=1$。

- Jensen不等式的积分形式：

$$f(\mathrm{Ex})=f\left(\int_S p(x) x d x\right) \leq \int_S p(x) f(x) d x=\mathrm{E} f(\mathrm{x})$$

其中$\int_S p(x) d x=1, p(x) \geq 0.$

- 算术几何平均不等式:

$$\sqrt{ab} \le \frac{a+b}{2}, a, b \ge 0.$$

- Hölder不等式：

$$\sum_{i=1}^{n} x_i y_i \le \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |y_i|^q\right)^{\frac{1}{q}}, \frac{1}{p}+\frac{1}{q} = 1, p > 1$$

提纲

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 的上镜图$\text{epi}\, f$
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

- 凸函数的非负加权和

$$f(x)=\omega_1 f_1(x)+\cdots+\omega_n f_n(x), \quad \omega_i \geq 0, i=1,\cdots,n.$$

- 凸函数与仿射变换的复合

$$g(x)=f(Ax+b), \quad A \in R^{n\times m}, b \in R^n,$$
$$\mathrm{dom f} = \{x \mid Ax+b \in \mathrm{dom f}\}$$

- 凸函数的逐点最大值

$$f(x) = \max\left(f_1(x), \cdots, f_n(x)\right)$$

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

例3.5 分段线性函数。函数

$$f(x) = \max\left\{a_1^T x + b_1, \ldots, a_L^T x + b_L\right\}$$

定义一个分段线性(或者实际上是仿射)函数(有L个或更少的区域)。该函数是凸的，因为它是仿射函数的点向最大值。
反过来也可以证明:任何具有L或更少区域的分段线性凸函数都可以用这种形式表示。

例3.6: 最大r个分量之和。
对于任意$x \in R^n$，用$x_{[i]}$表示x中第i大的分量，即将x的分量按照非升序进行排列得到下式:

$$x_{[1]} \geq \cdots \geq x_{[n]}$$

则对x的最大r个分量进行求和所得的函数

$$f(x) = \sum_{i=1}^{r} x_{[\mathrm{i}]}$$

是凸函数。

$$f(x) = \sum_{i=1}^{r} x_{[\mathrm{i}]} = \max\left\{x_{i_1} + \cdots + x_{i_r} \mid 1 \leq i_1 \leq \cdots \leq i_r\right\}$$

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

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

- 实对称矩阵的最大特征值$f(X) = \lambda_{\max}(X), X \in S^m$ 是一个凸函数。利用:$f(X) = \sup \{ \boldsymbol{y}^T \boldsymbol{X} \boldsymbol{y} \mid \|\boldsymbol{y}\|_2 = 1 \}$

例3.11 矩阵范数。考虑函数$f(X)=\|X\|_2$，其定义域为$\operatorname{dom} f=\mathbf{R}^{p\times q}$，其中$\|\cdot\|_2$表示谱范数或者最大奇值。函数$f$可以表述为

$$f(X)=\sup\left\{u^TXv \mid \|u\|_2=1,\|v\|_2=1\right\}$$

由于它是X的一族线性函数的逐点上确界，所以是凸函数。

作为一个推广，假设$\|\cdot\|_a$和$\|\cdot\|_b$分别是$\mathbf{R}^p$和$\mathbf{R}^q$上的范数，定义矩阵$X\in\mathbf{R}^{p\times q}$的诱导范数为

$$\|X\|_{a,b}=\sup_{v\neq 0}\frac{\|Xv\|_a}{\|v\|_b}.$$

（当两个范数都取Euclid 范数时，上述定义范数即为谱范数。）

诱导范数可以写成

$$\|X\|_{a,b}=\sup\left\{\|Xv\|_a \mid \|v\|_b=1\right\}=\sup\left\{u^TXv \mid \|u\|_{a*}=1,\|v\|_b=1\right\}$$

其中$\|\cdot\|_{a*}$是$\|\cdot\|_a$的对偶范数，在此我们利用了

$$\|z\|_a=\sup\left\{u^Tz \mid \|u\|_{a*}=1\right\}.$$

因此$\|X\|_{a,b}$可以表示成X的一系列线性函数的上确界，它是凸函数。

保凸运算

- 复合运算

  $g : \mathcal{R}^n \to \mathcal{R}^k, h : \mathcal{R}^k \to \mathcal{R}$, 定义复合函数$f = h \circ g : \mathcal{R}^n \to \mathcal{R}$

  $f(x) = h(g(x)), \operatorname{dom} f = \{x \in \operatorname{dom} g | g(x) \in \operatorname{dom} h\}$

- 最小值算子

  $g(x) = \inf_{y \in C} f(x, y)$，其中$f$ 关于(x, y) 是凸的。

保凸运算（标量复合）

标量函数$k=1$情况, 利用函数f的二阶导数:

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

标量复合

如果$n > L$, 同时不再假设函数h 和g 可微或者dom $g = \mathrm{dom}\, \mathrm{g} = R^n$, $\mathrm{dom}\, \mathrm{h} = R$，类似如下结论成立:

- 如果h 是凸函数且$\tilde{h}$ 非减, g 是凸函数, 则f 是凸函数
- 如果h 是凸函数且$\tilde{h}$ 非增, g 是凹函数, 则f 是凸函数
- 如果h 是凹函数且$\tilde{h}$ 非减, g 是凹函数, 则f 是凹函数
- 如果h 是凹函数且$\tilde{h}$ 非增, g 是凸函数, 则f 是凹函数

简单函数复合例子

- 如果$\mathbf{g}$ 是凸函数, 则$\exp(\mathrm{g}(\mathrm{x}))$ 是凸函数。
- 如果g 是凹函数且大于零, 则$\log(g(x))$ 是凹函数。
- 如果$\mathbf{g}$ 是凹函数且大于零, 则$\mathbf{1}/\mathbf{g}(\mathrm{x})$ 是凸函数。
- 如果$\mathbf{g}$ 是凸函数且不小于零, $\mathbf{p} \geq \mathbf{1}, g^p(x)$ 是凸函数。
- 如果$\mathbf{g}$ 是凸函数, 则$-\log(-\mathbf{g}(\mathbf{x}))$ 在$\{\mathbf{x}|\mathbf{g}(\mathbf{x}) < 0\}$ 上是凸函。

标量复合

例子，复合定理中h需要满足的条件

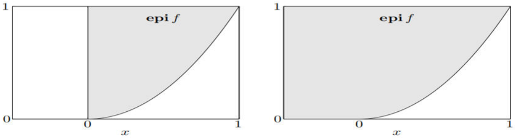


图3.7 左。函数x^2, 其定义域为$\mathbf{R}_+$, 在其定义域内是凸且非减的, 但是其扩展值延伸不是非减的。右。函数$\max\{x,0\}^2$, 其定义域为$\mathbf{R}$, 函数是凸函数, 且其扩展值延伸是非减的。

- 函数$h(x)=x^{1/2}$, 定义域为$\mathbf{dom}\, h=\mathbf{R}_+$, 其为凹函数且$\tilde{h}$ 非减。
- 函数$h(x)=x^{3/2}$, 定义域为$\mathbf{dom}\, h=\mathbf{R}_+$, 其为凸函数, 但是不满足$\tilde{h}$ 非减的条件。例如, $\tilde{h}(-1)=\infty$ 但$\tilde{h}(1)=1$ 。
- 当$x\geqslant 0$ 时, $h(x)=x^{3/2}$, 当$x<0$ 时$h(x)=0$, 定义域为$\mathbf{dom}\, h=\mathbf{R}$, h 是凸函数且满足$\tilde{h}$ 非减的条件。

矢量复合

复合$g : \mathbf{R}^n \to \mathbf{R}^k$ 和$h : \mathbf{R}^k \to \mathbf{R}$:

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

f 是凸的

如果g_i 凸面的h 凸面的$\tilde{h}$ 在每个自变量中不减少

如果g_i 凹面的h 凸面的$\tilde{h}$在每个自变量中不增加

证明（对于$n = 1$，可微分g, h）

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

示例

- $\sum_{i=1}^m \log g_i(x)$ 如果g_i 是凹的和正的
- $\log \sum_{i=1}^m \exp g_i(x)$ 是凸的，如果g_i 是凸的

矢量复合的例子

- 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;
- 如果h 是凸函数且在每维分量上h 非增, g_i 是凹函数, 则f 是凸函数;
- 如果h 是凹函数且在每维分量上h 非减, g_i 是凹函数, 则f 是凹函数。

例3.14 矢量复合的例子。

- 令$h(z) = z_{[1]} + \cdots + z_{[r]}$, 即对$z \in \mathbf{R}^k$ 的前r 大分量进行求和。则h 是凸函数且在每一维分量上非减。假设$g_1, \cdots, g_k$ 是$\mathbf{R}^n$ 上的凸函数, 则复合函数$f = h \circ g$, 即最大r 个g_i 函数的逐点和, 是凸函数。
- 函数$h(z) = \log\left(\sum_{i=1}^k e^{z_i}\right)$ 是凸函数且在每一维分量上非减, 因此只要g_i 是凸函数, $\log\left(\sum_{i=1}^k e^{g_i}\right)$ 就是凸函数。

矢量复合的例子

- 对$0 < p \leqslant 1$，定义在$\mathbf{R}_+^k$ 上的函数$h(z) = \left(\sum_{i=1}^k z_i^p\right)^{1/p}$ 是凹的，且其扩展值延伸(当$z \not\succeq 0$ 时为$-\infty$) 在每维分量上非减，则若g_i 是凹函数且非负，$f(x) = \left(\sum_{i=1}^k g_i(x)^p\right)^{1/p}$ 是凹函数。
- 设$p \geqslant 1, g_1, \cdots, g_k$ 是凸函数且非负。则函数$\left(\sum_{i=1}^k g_i(x)^p\right)^{1/p}$ 是凸函数。为了说明这一点，考虑函数$h : \mathbf{R}^k \to \mathbf{R}$

$$h(z) = \left(\sum_{i=1}^k \max\{z_i, 0\}^p\right)^{1/p}$$

其中$\mathrm{dom}\, h = \mathbf{R}^k$，因此$h = \tilde{h}$。由函数$h$ 是凸函数且非减可知$h(g(x))$ 关于x 是凸函数。对$z \succeq 0$，我们有$h(z) = \left(\sum_{i=1}^k z_i^p\right)^{1/p}$，所以$\left(\sum_{i=1}^k g_i(x)^p\right)^{1/p}$ 是凸函数。

矢量复合的例子

- 几何平均函数$h(z)=\left(\prod_{i=1}^k z_i\right)^{1/k}$，定义域为$\mathbf{R}_+^k$，它是凹函数，且其扩展值延伸在每维分量上非减。因此若$g_1,\cdots,g_k$是非负凹函数，它们的几何平均$\left(\prod_{i=1}^k g_i\right)^{1/k}$也是非负凹函数。

- 设$p\geqslant 1, g_1,\cdots,g_k$是凸函数且非负。则函数$\left(\sum_{i=1}^k g_i(x)^p\right)^{1/p}$是凸函数。为了说明这一点，考虑函数$h:\mathbf{R}^k\to\mathbf{R}$

$$h(z)=\left(\sum_{i=1}^k \max\{z_i,0\}^p\right)^{1/p}$$

其中$\operatorname{dom} h=\mathbf{R}^k$，因此$h=\tilde{h}$。由函数$h$是凸函数且非减可知$h(g(x))$关于$x$是凸函数。对$z\succeq 0$，我们有$h(z)=\left(\sum_{i=1}^k z_i^p\right)^{1/p}$，所以$\left(\sum_{i=1}^k g_i(x)^p\right)^{1/p}$是凸函数。

矢量复合的例子

- 几何平均函数$h(z)=\left(\prod_{i=1}^{k} z_i\right)^{1/k}$，定义域为$\mathbf{R}_{+}^{k}$，它是凹函数，且其扩展值延伸在每维分量上非减。因此若$g_1,\cdots,g_k$是非负凹函数，它们的几何平均$\left(\prod_{i=1}^{k} g_i\right)^{1/k}$也是非负凹函数。

最小化

如果函数$f(x,y)$ 是关于(x,y) 的凸函数，集合C 是一个非空凸集,定义函数：

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

是凸函数

example:

$f(x,y) = x^TAx + 2x^TBy + y^TCy$，其中：

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

最小化y 得到$g(x) = \inf_y f(x,y) = x^T\left(A - BC^{-1}B^T\right)x$，$g$ 是凸函数, 因此Schur补：$A - BC^{-1}B^T \succeq 0$

若S是凸集 则到某一集合的距离: $\operatorname{dist}(x,S) = \inf_{y\in S}\|x-y\|$是凸函数

凸函数的透视运算保凸：

$$g(x,t) = \mathrm{tf}(x/t)$$

其定义域: $\mathrm{dom}\, g = \{(x,t) \mid \frac{x}{t} \in \mathrm{dom} f, t > 0\}$

f 是凸函数，则g 也是凸函数;

f 是凹函数，则g 也是凹函数。

$$\begin{aligned}(x,t,s) \in \ \mathsf{epi}\, g &\Longleftrightarrow tf(x/t) \le s \\ &\Longleftrightarrow f(x/t) \le s/t \\ &\Longleftrightarrow (x/t, s/t) \in \ \mathsf{epi}\, f.\end{aligned}$$

例3.18 Euclid范数的平方.$\mathbf{R}^n$上的凸函数$f(x)=x^Tx$的透视函数由下式给出：

$$g(x,t)=t(x/t)^T(x/t)=\frac{x^Tx}{t}$$

当$t>0$时它关于(x,t)是凸函数。

我们可以利用其他方法导出g的凸性，首先，将g表示为一系列二次-线性分式x_i^2/t的和,在§3.1.5中，我们可以将g表述为一种特殊的矩阵分式函数$x^T(tI)^{-1}x$，由此导出凸性(见例3.4)。

例3.19 负对数。考虑函数$\mathbf{R}_{++}$上的凸函数$f(x)=-\log x$。其透视函数为：

$$g(x,t)=-t\log(x/t)=t\log(t/x)=t\log t-t\log x,$$

在$\mathbf{R}^2_{++}$上它是凸函数。函数g称为关于t和x的相对熵。

当$x=1$时，g即为负熵函数。

基于g的凸性。我们可以得到一些有趣的相关函数的凸性或凹性。首先，定义两个向量$u,v\in\mathbf{R}^n_{++}$的相对熵。

$$\sum_{i=1}^n u_i\log\left(u_i/v_i\right)$$

我们可以利用其他方法导出g的凸性，首先，将g表示为一系列二次-线性分式x_i^2/t的和,在§3.1.5中，我们可以将g表述为一种特殊的矩阵分式函数$x^T(tI)^{-1}x$，由此导出凸性(见例3.4)。

由于它是一系列u_i,v_i的相对熵和，因此(u,v)是凸函数。

另一个密切相关的函数是向量$u, v \in \mathbf{R}^n_{++}$之间的Kullback-Leibler散度，其形式为：

$$D_{kl}(u, v) = \sum \left(u_i \log \left(u_i / v_i\right) - u_i + v_i\right)$$

因为它是(u, v)的相对熵和线性函数的和，所以它也是凸函数。Kullback-Leibler散度总是满足$D_{kl}(u, v) \geq 0$，当且仅当$u = v$时，$D_{kl}(u, v) = 0$，因此Kullback-Leibler散度可以用来衡量两个正向量之间的偏差；见习题3.13。(注意到当u和v都是概率向量，即$\mathbf{1}^T u = \mathbf{1}^T v = 1$时，相对熵和Kullback-Leibler散度是等价的。)

例3.20 设$f : \mathbf{R}^m \to \mathbf{R}$是凸函数，$A \in \mathbf{R}^{m \times n}, b \in \mathbf{R}^m, c \in \mathbf{R}^n$, and $d \in \mathbf{R}$。我们定义：

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

其中：

$$\operatorname{dom} g = \left\{x \mid c^T x + d > 0, (Ax + b) / \left(c^T x + d\right) \in \operatorname{dom} f\right\}$$

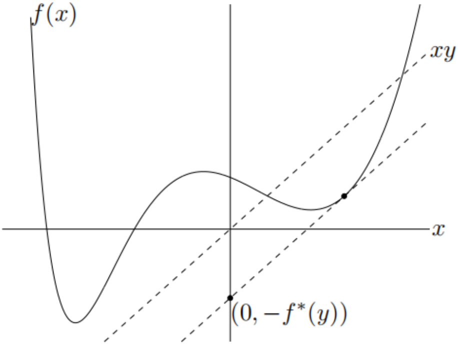
其中g是凸的。

共轭函数(conjugate function)

定义

设函数$f:\mathcal{R}^n\to\mathcal{R}$，其共轭函数$f^*:\mathcal{R}^n\to\mathcal{R}$，定义为：

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



函数$f:\mathbf{R}\to\mathbf{R}$以及某一$y\in\mathbf{R}$。共轭函数$f^*(y)$是线性函数$yx$和$f(x)$之间的最大差值，如果$f$可微，在满足$f'(x)=y$的点处的差值最大$x$。

共轭函数(conjugate function)

共轭函数：

$$f(x) = ax + b, f^*(a) = -b, \quad \mathrm{dom} f^* = \{a\}$$

$$f(x) = e^x, f^*(y) = y\log(y) - y(y \geq 0)$$

$$f(x) = x\log x, f^*(y) = \mathrm{e}^{y-1}, \mathrm{dom} f^* = R$$

$$f(x) = \frac{1}{x}, \quad x \in R_{++}, f^*(y) = -2(-\mathrm{y})^{1/2}, \mathrm{dom} f^* = -R_+$$

$$f(x) = -\log(x), \quad x \in R_{++}, f^*(y) = -\log(-y) - 1, y < 0.$$

例3.22 严格凸的二次函数。考虑函数$f(x)=\frac{1}{2}x^TQx$，$Q\in\mathbf{S}^n_{++}$。对所有x，y的函数$y^Tx-\frac{1}{2}x^TQx$都有上界并在$x=Q^{-1}y$处达到上确界，因此：

$$f^*(y)=\frac{1}{2}y^TQ^{-1}y$$

例3.24 示性函数。设I_S是某个集合$S\subseteq\mathbf{R}^n$(不一定是凸集)的示性函数，即当x在$\operatorname{dom} I_S=S$内时，$I_S(x)=0$示性函数的共轭函数为

$$I_S^*(y)=\sup_{x\in S} y^Tx,$$

它是集合S的支撑函数。

例3.23 对数-行列式。我们考虑$\mathbf{S}_{++}^n$上定义的函数$f(X)=\log\det X^{-1}$。其共轭函数定义为：

$$f^*(Y)=\sup_{X\succ 0}(\operatorname{tr}(YX)+\log\det X),$$

其中$\operatorname{tr}(YX)$是$\mathbf{S}^n$上的标准内积。首先我们说明只有当$Y\prec 0$时，$\operatorname{tr}(YX)+\log\det X$才有上界。如果$Y\not\prec 0$，则$Y$有特征向量$v$，$\|v\|_2=1$且对应的特征值$\lambda\geq 0$。令$X=I+tvv^T$我们有：

$$\operatorname{tr}(YX)+\log\det X=\operatorname{tr}Y+t\lambda+\log\det\left(I+tvv^T\right)=\operatorname{tr}Y+t\lambda+\log(1+t),$$

当$t\to\infty$时，上式无界。接下来考虑$Y\prec 0$的情形。为了追求最大值，令对X的偏导为零，则:

$$\nabla_X(\operatorname{tr}(YX)+\log\det X)=Y+X^{-1}=0$$

(见$\mathcal{S}$ A.4.1), 得到$X=-Y^{-1}$ (X是正定的)。因此

$$f^*(Y)=\log\det(-Y)^{-1}-n$$

其定义域为$\operatorname{dom}f^*=-\mathbf{S}_{++}^n$。

例3.25 指数和对数函数。为了得到指数和的对数函数$f(x)=\log\left(\sum_{i=1}^n e^{x_i}\right)$的对数函数，
首先考察y取何值时$y^Tx-f(x)$的最大值可以得到。对x求导，令其为零，我们可以得到如下条件：

$$y_i=\frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}},\quad i=1,\ldots,n.$$

当且仅当$y\succ 0$以及$1^Ty=1$时上述方程有解。将y_i的表达式代入$y^Tx-f(x)$我们可以得到$f^*(y)=\sum_{i=1}^n y_i\log y_i$。
根据前面的约定，$0\log 0$等于0，因此只要满足$y\succeq 0$以及$1^Ty=1$，即使当y的某些分量为0时，f^*的表达式仍然正确。
事实上f^*的定义域即为$1^Ty=1, y\succeq 0$。为了说明这一点，假设y的某个分量是负的，比如说$y_k<0$，令$x_k=-t, x_i=0, i\neq k$, 令t趋向于无穷，$y^Tx-f(x)$无上界。

共轭函数(conjugate function)

如果$y \succeq 0$但是$1^T y \neq 1$,我们令$x = t1$,可以得到：

$$y^T x - f(x) = t1^T y - t - \log n.$$

若$1^T y > 1$,当$t \to \infty$时上述表达式无界;当$1^T y < 1$,若$t \to -\infty$时其无界。总之,

$$f^*(y) = \begin{cases} \sum_{i=1}^n y_i \log y_i & \text{如果} y \succeq 0 \text{ 且} 1^T y = 1 \\ \infty & \text{其他情况} \end{cases}$$

换言之，指数和的对数函数的共轭函数是概率单纯形内的负熵函数

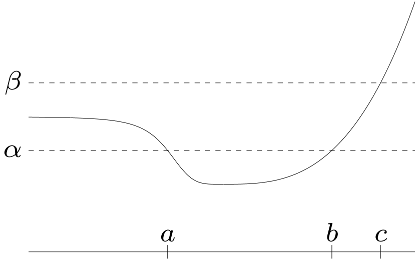
提纲

拟凸函数

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

$$S_\alpha = \{x \in \operatorname{dom} f | f(x) \le \alpha\}$$

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

拟凸、凹函数的例子

- $\sqrt{|x|}$ 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

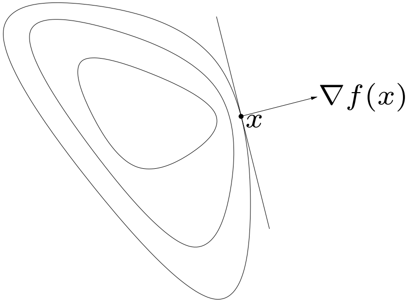
拟凸函数的性质

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

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

如果$\log f$ 是凹函数，则f 称为对数凹函数，

- 幂函数：当$a\leq 0$时，x^a 是$\mathbb{R}_{++}$ 上的对数凸函数；当$a\geq 0$，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