# 凸函数

王正强

重庆邮电大学通信学院

# 提纲

1. 凸优化
   - 基础知识
   - 凸函数的定义与性质
   - 保凸的运算
   - 凸优化问题

# 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，

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

# 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.

- 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$\frac{\partial f}{\partial X}=(XY-A)Y^{\mathrm{T}}$.

- *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 ($\alpha$-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$\alpha$-下水平集.

**定义 (上镜图)**

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

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

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$\mathbf{epi}\,f$为闭集，则称$f$为闭函数.

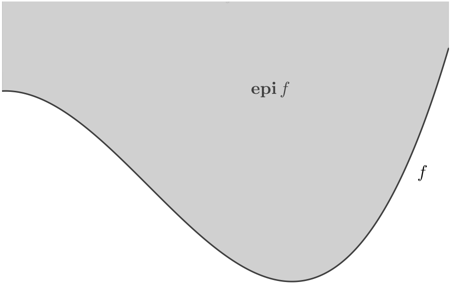



Figure: 函数$f$和其上镜图$\mathbf{epi}\,f$

## $R$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$f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$，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.

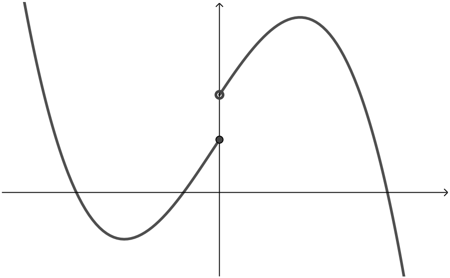


Figure: 下半连续函数$f(x)$

# 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$\alpha$-下水平集都是闭集；
2. $f(x)$是下半连续的；
3. $f(x)$是闭函数.

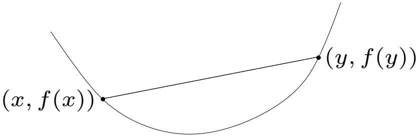
# 凸函数的定义

**定义 (凸函数)**

$f : \mathbb{R}^n \to \mathbb{R}$ 为适当函数，如果$\mathbf{dom}\, f$ 是凸集，且

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

对所有$x, y \in \mathbf{dom}\, f$ , $0 \le \theta \le 1$ 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$x, y \in \mathbf{dom}\, f$ , $x \ne y$, $0 < \theta < 1$，有

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

则称$f$ 是严格凸函数

# 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$x \in \mathrm{dom}\, f, v \in \mathbb{R}^n$，函数$g:\mathbb{R} \to \mathbb{R}$ 是关于$t$的凸函数

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

例：$f(X) = -\log\det X$ 是凸函数，其中$\mathrm{dom}\, f = \mathbb{S}^n_{++}$. 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$\lambda_i$ 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$\mathrm{dom}f$是凸集，取$v=y-x$，以及$t_1=0,t_2=1$，由$\mathrm{dom}g$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$\mathrm{dom}f$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

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

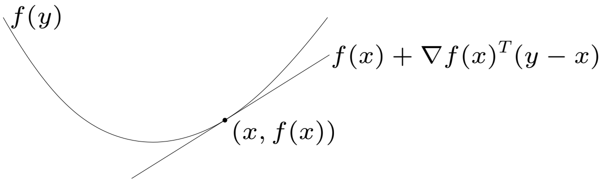
这说明$f(x)$是凸函数． □

# 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$x, y \in \mathrm{dom} f$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. $\square$

# 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$\nabla f$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

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

将两式不等号左右两边相加即可得到结论.

**Proof.**

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

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

□

# 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$f(x) = (1/2)x^T Px + q^T x + r$ (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$P \succeq 0$
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)?

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

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

$$p_j \geq 0, j = 1, \cdots, n. \tag{3}$$

问题(9)-(11) 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。

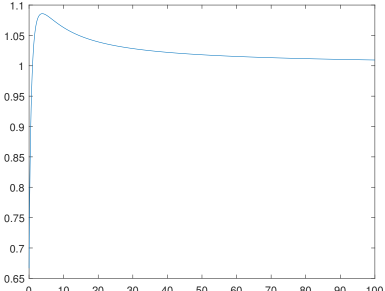
问题来源：

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。

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

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:*

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

Using lemma 2 and $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} h_i p_i / \left(\sum_{i=1}^{n} h_i p_i + \sigma^2\right) < 1$, (9)-(11) 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 (14)-(16) 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, (14)-(16) 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$\alpha \leq 0$，$x^{\alpha}$ 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$0 \leq \alpha \leq 1$，$x^{\alpha}$ 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:

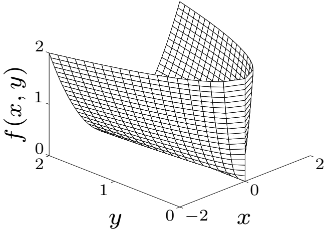
$$f(X) = \mathrm{tr}(A^T X) + b = \sum_{i=1}^m \sum_{j=1}^n A_{ij} X_{ij} + b$$

# 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$\frac{\partial^2 f}{\partial x_i\partial x_j}=\frac{-z_iz_j}{(1^Tz)^2}(i\neq j), \frac{\partial^2 f}{\partial^2 x_i}=\frac{-z_iz_j+z_i1^Tz}{(1^Tz)^2}$

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

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

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

由柯西不等式，得$(\sum_k v_kz_k)^2\leq(\sum_k z_kv_k^2)(\sum_k z_k)$，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$\mathbf{dom}\, f=\mathbf{R}_{++}$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$\alpha$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$\alpha$-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:

$$\left\{x \in \mathbf{R}_+^n \mid G(x) \geq \alpha A(x)\right\}$$

几何平均值大于等于因子$\alpha$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$epif=\{(x,t)|x\in \mathrm{dom}f, f(x)\leq t\}$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

$$hypof=\{(x,t)|x\in \mathrm{dom}f, f(x)\geq t\}$$

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

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

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：

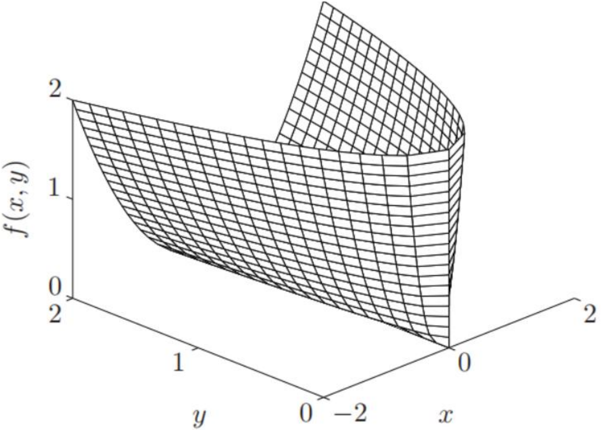
$$\begin{aligned}\mathrm{epi}f&=\left\{(x,Y,t)\mid Y\succ 0,x^TY^{-1}x\leq t\right\}\\&=\left\{(x,Y,t)\mid \begin{bmatrix}Y & x\\ x^T & t\end{bmatrix}\succeq 0,Y\succ 0\right\}\end{aligned}$$

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



# 上镜图

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

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

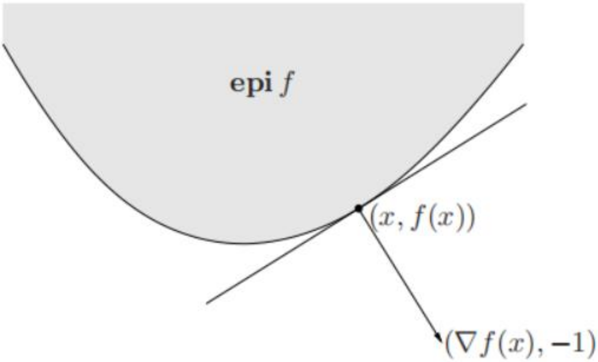
上式可以描述为

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

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$0 \le \theta \le 1$,

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

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

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

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：

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

其中$\int_S p(x)dx = 1, p(x) \geq 0.$

- 算术几何平均不等式:

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

- Hölder不等式：

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

函数$-log\ x$是凸函数：利用Jensen不等式，令$\theta = 1/2$，可得

$$-\log\left(\frac{a+b}{2}\right) \leqslant \frac{-\log a - \log b}{2}.$$

等式两边取指数即可得到式(3.6)。
作为另一个小例子，我们来证明Holder不等式：
对$p > 1, 1/p + 1/q = 1$，以及$x, y \in \mathbf{R}^n$，有

$$\sum_{i=1}^{n} x_i y_i \leqslant \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p} \left(\sum_{i=1}^{n} |y_i|^q\right)^{1/q}$$

由$-log\ x$的凸性以及Jensen不等式，我们可以得到更为一般的算数-几何平均不等式

$$a^\theta b^{1-\theta} \leqslant \theta a + (1-\theta) b,$$

其中$a,b \geqslant 0, 0 \leqslant \theta \leqslant 1$，令

$$a = \frac{|x_i|^p}{\sum\limits_{j=1}^{n} |x_j|^p}, \quad b = \frac{|y_i|^q}{\sum\limits_{j=1}^{n} |y_j|^q}, \quad \theta = 1/p,$$

可以得到如下不等式

$$\left( \frac{|x_i|^p}{\sum\limits_{j=1}^{n} |x_j|^p} \right)^{1/p} \left( \frac{|y_i|^q}{\sum\limits_{j=1}^{n} |y_j|^q} \right)^{1/q} \leqslant \frac{|x_i|^p}{p \sum\limits_{j=1}^{n} |x_j|^p} + \frac{|y_i|^q}{q \sum\limits_{j=1}^{n} |y_j|^q}.$$

对$i$进行求和可以得到Holder不等式。

# 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$\mathrm{epi}\, f$
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

- 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。

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

# 逐点取上确界

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。从上境图的角度理解，一系列函数的逐点上确界函数对应着这些函数上境图的交集：对于函数$f, g$, 和$\mathcal{A}$, 有如下关系成立：

$$\text{epi}\ g = \bigcap_{y \in \mathcal{A}} \text{epi} f(\cdot, y).$$

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:$f(X) = \sup \left\{ \boldsymbol{y}^T \boldsymbol{X} \boldsymbol{y} \mid \|\boldsymbol{y}\|_2 = 1 \right\}$

例3.11 矩阵范数。考虑函数$f(X)=\|X\|_2$，其定义域为$\mathrm{dom}\,f=\mathbf{R}^{p\times q}$，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。

- 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$f=h\circ g:\mathcal{R}^n\to\mathcal{R}$

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

- 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$\mathbf{k}=\mathbf{1}$ 情况, 利用函数$\mathbf{f}$ 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 $g =$ dom g $= R^n$, dom h $= R$，类似地如下结论成立:

- 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，$\mathbf{p} \geq \mathbf{1}, \boldsymbol{g}^p(x)$ 是凸函数。
- 如果$\mathbf{g}$ 是凸函数，则$-\log(-\mathbf{g}(\mathbf{x}))$ 在$\{\mathbf{x}|\mathbf{g}(\mathbf{x}) < 0\}$ 上是凸函数。

# 标量复合

例子，复合定理中h需要满足的条件

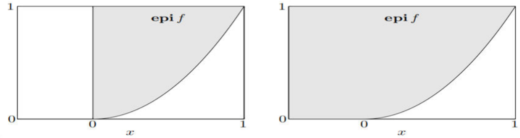



图3.7 左。函数$x^2$, 其定义域为$\mathbf{R}_+$, 在其定义域内是凸且非减的，但是其扩展值延伸不是非减的。右。函数$\max\{x,0\}^2$, 其定义域为$\mathbf{R}$, 函数是凸函数，且其扩展值延伸是非减的。

- 函数$h(x) = x^{1/2}$, 定义域为$\mathrm{dom}\, h = \mathbf{R}_+$, 其为凹函数且$\tilde{h}$ 非减。
- 函数$h(x) = x^{3/2}$, 定义域为$\mathrm{dom}\, h = \mathbf{R}_+$, 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$\mathrm{dom}\, h = \mathbf{R}$, $h$ 是凸函数且满足$\tilde{h}$ 非减的条件。

# 矢量复合

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 $\begin{matrix} g_i \text{ 是凸函数}, h \text{ 是凸函数}, \tilde{h} \text{ 关于每个分量单调不减} \\ g_i \text{ 是凹函数}, h \text{ 是凸函数}, \tilde{h} \text{ 关于每个分量单调不增} \end{matrix}$ ，那么$f$是凸函数。

对$n=1$，$g,h$均可微的情形，我们给出简证

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

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

## 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$\operatorname{dom} h = \mathbf{R}^k$, 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。

- 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。

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

例3.15 Schur补。设二次函数

$f(x,y) = x^T Ax + 2x^T By + y^T Cy$，关于$(x,y)$是凸函数，即：

$$\begin{bmatrix} A & B \\ B^T & C \end{bmatrix} \succeq 0.$$

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：$\mathrm{dist}(x,S) = \inf_{y \in S} \|x - y\|$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: $\mathrm{dom}\, g = \left\{(x,t) \mid \frac{x}{t} \in \mathrm{dom} f, t > 0\right\}$

$f$ 是凸函数，则$g$ 也是凸函数;

$f$ 是凹函数，则$g$ 也是凹函数。

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

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

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

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。

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

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$D_{\mathrm{kl}}(u, v) \geq 0$，当且仅当$u = v$时，$D_{\mathrm{kl}}(u, v) = 0$，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：

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

其中：

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

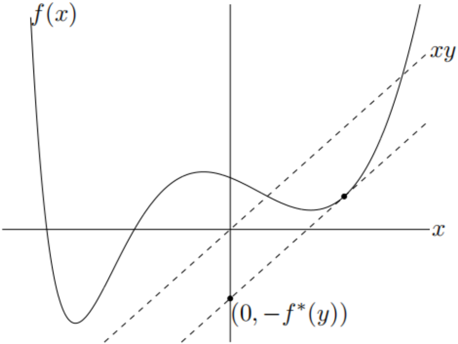
其中$g$是凸的。

# 共轭函数(conjugate function)

**定义**

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

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




函数$f:\mathbf{R}\rightarrow\mathbf{R}$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$x$。

考虑$R$上一些凸函数的共轭函数：

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

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

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

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

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

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。
只有当$Y \prec 0$时，$\operatorname{tr}(YX) + \log\det X$才有上界。如果$Y \nprec 0$，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：

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

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,

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

即: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。

例3.26 范数。令$\|\cdot\|$ 表示为$\mathbf{R}^n$上的范数, 其对偶范数为$\|\cdot\|_*$。我们说明$f(x)=\|x\|$ 的共轭函数为

$$f^*(y)=\begin{cases}0 & \|y\|_*\leq 1\\ \infty & \text{其他}\end{cases}$$

即范数的共轭函数是对偶范数单位球的示性函数。 如果$\|y\|_*>1$, 根据对偶范数的定义，存在$z\in\mathbf{R}^n$，$\|z\|\leq 1$ 使得$y^Tz>1$。取$x=tz$
令$t\to\infty$, 我们有

$$y^Tx-\|x\|=t\left(y^Tz-\|z\|\right)\to\infty,$$

即$f^*(y)=\infty$，没有上界. 反之, 如果$\|y\|_*\leq 1$, 对于任意$x$,我们可得$y^Tx\leq\|x\|\|y\|$。即对于任意$x$, $y^Tx-\|x\|\leq 0$。因此在$x=0$处，$y^Tx-\|x\|$, 达到最大值0 。

例3.27 范数的平方。考虑函数$f(x) = (1/2)\|x\|^2$, 其中$\|\cdot\|$ 是范数, 对偶函数为$\|\cdot\|_*$。我们说明此共轭函数为$f^*(y) = (1/2)\|y\|_*^2$。
由$y^Tx \le \|y\|_*\|x\|$, 可知下式对于任意$x$ 成立。

$$y^Tx - (1/2)\|x\|^2 \le \|y\|_*\|x\| - (1/2)\|x\|^2$$

上式右端是$\|x\|$ 的二次函数, 其最大值为$(1/2)\|y\|_*^2$。因此对于任意$x$, 我们有

$$y^Tx - (1/2)\|x\|^2 \le (1/2)\|y\|_*^2,$$

即$f^*(y) \le (1/2)\|y\|_*^2$。为了说明$f^*(y) \ge (1/2)\|y\|_*^2$, 任取满足$y^Tx = \|y\|_*\|x\|$ 的向量$x$, 对其进行伸缩使得$\|x\| = \|y\|$。对于此$x$有,

$$y^Tx - (1/2)\|x\|^2 = (1/2)\|y\|_*^2,$$

因此说明$f^*(y) \ge (1/2)\|y\|_*^2$。

A.1 范数A.1.1 内积，Euclid范数，夹角定义在$n$维实向量集合$\mathbf{R}^n$上的内积为，对任意for $x, y \in \mathbf{R}^n$。

$$\langle x, y\rangle = x^T y = \sum_{i=1}^{n} x_i y_i,$$

在本书中我们采用符号$x^T y$ 代替$\langle x, y\rangle$。向量$x \in \mathbf{R}^n$ 的Euclid范数，或者$\ell_2$范数，定义为

$$\|x\|_2 = \left(x^T x\right)^{1/2} = \left(x_1^2 + \cdots + x_n^2\right)^{1/2}.$$

对于任意$x, y \in \mathbf{R}^n$.其Cauchy-Schwartz 不等式是$\left|x^T y\right| \leq \|x\|_2 \|y\|_2$。两个非零向量$x, y \in \mathbf{R}^n$ 之间的夹角定义为

$$\angle(x, y) = \cos^{-1}\left(\frac{x^T y}{\|x\|_2 \|y\|_2}\right),$$

这里我们采用$\cos^{-1}(u) \in [0, \pi]$。我们称$x$ 和$y$ 在$x^T y = 0$的情况下正交。定义在$m \times n$ 实矩阵集合$\mathbf{R}^{m\times n}$上的标准内积为

$$\langle X, Y\rangle = \mathrm{tr}\left(X^T Y\right) = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij} Y_{ij},$$

其对于任意$X, Y \in \mathbf{R}^{m\times n}$成立。(此处**tr** 表示矩阵的迹，即其对角元素之和) 我们用符号$\mathrm{tr}\left(X^T Y\right)$ 代替$\langle X, Y\rangle$. 两个矩阵的内积实际上就是将矩阵的元素按一定的顺序（如按行）排序后所生成的$\mathbf{R}^{mn}$ 中相应向量的内积。矩阵$X \in \mathbf{R}^{m\times n}$ 的Frobenius范数矩阵定义为

$$\|X\|_F = \left(\mathrm{tr}\left(X^T X\right)\right)^{1/2} = \left(\sum_{i=1}^{m}\sum_{j=1}^{n} X_{ij}^2\right)^{1/2}.$$

Frobenius范数是实际上就是将矩阵的系数按一定顺序排列后所生成的相应向量的Euclidean范数(一个矩阵的$\ell_2$范数是不同的范数，见§A.1.5.) 定义在$n\times n$ 对称矩阵集合$\mathbf{S}^n$ 上的标准内积为，

$$\langle X, Y\rangle = \mathrm{tr}(XY) = \sum_{i=1}^{n}\sum_{j=1}^{n} X_{ij}Y_{ij} = \sum_{i=1}^{n} X_{ii}Y_{ii} + 2\sum_{i<j} X_{ij}Y_{ij}$$

A.1.2 范数，距离以及单位球

满足以下条件的函数$f : \mathbf{R}^n \to \mathbf{R}, \operatorname{dom} f = \mathbf{R}^n$ 称为范数，
- $f$ 是非负的：对所有的$x \in \mathbf{R}^n$ 成立$f(x) \geqslant 0$,
- $f$ 是正定的：仅对$x = 0$ 成立$f(x) = 0$,
- $f$ 是齐次的：对所有的$x \in \mathbf{R}^n$ 和$t \in \mathbf{R}$ 成立$f(tx) = |t| f(x)$,
- $f$ 满足三角不等式：对所有的$x, y \in \mathbf{R}^n$ 成立$f(x+y) \leqslant f(x) + f(y)$ 。

我们采用符号$f(x) = \|x\|$, 该符号意味着范数是$\mathbf{R}$ 上绝对值函数的推广。我们用$\|x\|_{\text{symb}}$ 表示具体范数，其中下标是区分范数的助记符号。范数是对向量$x$ 的长度的度量；我们可以用两个向量$x$ 和$y$ 的差异的长度度量它们之间的距离，即

$$\operatorname{dist}(x, y) = \|x - y\|.$$

我们用$\operatorname{dist}(x, y)$ 表示$x$ 和$y$ 之间用范数$\|\cdot\|$ 表示的距离。其范数小于或等于1 的所有向量的集合

$$\mathcal{B} = \{x \in \mathbf{R}^n \mid \|x\| \leqslant 1\},$$

# 范数

范数小于或等于1的所有向量的集合$\mathcal{B}=\{x\in R|\ ||x||\leq 1\}$称为范数$\|\cdot\|$的单位球。单位球具有以下性质:

- $\mathcal{B}$ 关于原点对称, 即当且仅当$-x\in\mathcal{B}$ 时成立$x\in\mathcal{B}$,
- $\mathcal{B}$ 是凸集,
- $\mathcal{B}$ 是有界闭集, 内部非空。

反之, 如果$C\subseteq\mathbf{R}^n$ 是满足这三个条件的任何集合, 它就是一种范数的单位球, 该范数由下式给出

$$\|x\|=(\sup\{t\geqslant 0\mid tx\in C\})^{-1}.$$

A.1.3 例子

最简单的范数例子是$\mathbf{R}$ 上的绝对值。另一个简单的例子是上面式(A.1)定义的$\mathbf{R}^n$ 上的Euclid 或$\ell_2$-范数。另外两个经常用到的$\mathbf{R}^n$ 上的范数是定义为

$$\|x\|_1 = |x_1| + \cdots + |x_n|$$

的绝对值之和或$\ell_1$-范数, 以及定义为

$$\|x\|_\infty = \max\left\{|x_1|, \cdots, |x_n|\right\}$$

的Chebyshev 或$\ell_\infty$-范数。这三种范数属于由一个常数决定的参数化的范数类, 习惯上用$p$ 表示这个常数, 要求$p \geqslant 1$, 对应的范数用$\ell_p$-范数表示, 定义为

$$\|x\|_p = \left(|x_1|^p + \cdots + |x_n|^p\right)^{1/p}.$$

取$p = 1$ 就得到$\ell_1$-范数, $p = 2$ 就得到Euclid 范数。不难证明对于任何$x \in \mathbf{R}^n$ 成立

$$\lim_{p\to\infty} \|x\|_p = \max\left\{|x_1|, \cdots, |x_n|\right\},$$

因此$\ell_\infty$-范数作为一种极限情况也属于这类范数。

另一类重要的范数是二次范数。对$P \in \mathbf{S}^n_{++}$，我们定义$P$-二次范数如下

$$\|x\|_P = \left(x^T P x\right)^{1/2} = \left\|P^{1/2}x\right\|_2 .$$

二次范数的单位球是椭圆(反之，如果一个范数的单位球是椭圆，该范数就是二次范数)。

常用的$\mathbf{R}^{m\times n}$ 上的范数有上述式(A.2) 定义的Frobenius 范数，绝对值之和范数

$$\|X\|_{\mathrm{sav}} = \sum_{i=1}^{m}\sum_{j=1}^{n}|X_{ij}|$$

以及最大绝对值范数

$$\|X\|_{\mathsf{mav}} = \max\left\{|X_{ij}| \mid i = 1,\cdots,m, j = 1,\cdots,n\right\}.$$

在$\mathcal{S}A.1.5$ 我们将看到其他一些重要的矩阵范数。

A.1.4 范数的等价性假定$\|\cdot\|_{\mathrm{a}}$ 和$\|\cdot\|_{\mathrm{b}}$ 是$\mathbf{R}^n$ 上的范数。分析中的一个基本结论是，存在正常数$\alpha$ 和$\beta$ 对所有的$x\in\mathbf{R}^n$ 成立

$$\alpha\|x\|_{\mathrm{a}}\leqslant\|x\|_{\mathrm{b}}\leqslant\beta\|x\|_{\mathrm{a}}.$$

该结论意味着所有范数是等价的，即它们定义了相同的开集，相同的收敛序列，等等(见§A.2)。(我们说任何有限维向量空间上的范数都是等价的，但这个结果在无限维向量空间上并不一定成立。) 利用凸分析，我们可以给出一个更明确的结论：如果$\|\cdot\|$ 是$\mathbf{R}^n$ 上的任意范数，那么存在一个二次范数$\|\cdot\|_P$ 对所有的$x$ 成立

$$\|x\|_P\leqslant\|x\|\leqslant\sqrt{n}\|x\|_P.$$

换言之，$\mathbf{R}^n$ 上任何范数可以在$\sqrt{n}$ 倍的范围内被二次范数一致逼近(见$\mathcal{S}8.4.1$ )。

A.1.5 算子范数

假设$\|\cdot\|_{\rm a}$ 和$\|\cdot\|_{\rm b}$ 分别是$\mathbf{R}^m$ 和$\mathbf{R}^n$ 上的范数。对于$X \in \mathbf{R}^{m\times n}$, 我们定义由范数$\|\cdot\|_{\rm a}$ 和$\|\cdot\|_{\rm b}$ 导出的算子范数

$$\|X\|_{\rm a,b} = \sup\left\{\|Xu\|_{\rm a} \mid \|u\|_{\rm b} \leqslant 1\right\}.$$

(可以验证, 该式确实定义了$\mathbf{R}^{m\times n}$ 上的一个范数。) 当$\|\cdot\|_{\rm a}$ 和$\|\cdot\|_{\rm b}$ 都是Euclid 范数时, $X$ 的算子范数是它的最大奇异值, 用$\|X\|_2$ 表示:

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

(该定义当$X \in \mathbf{R}^{m\times 1}$ 时和$\mathbf{R}^m$ 上的Euclid 范数相吻合, 因此在符号上不会产生冲突。) 这个范数也被称为$X$ 的谱范数或$\ell_2$-范数。作为一个例子, 考虑由$\mathbf{R}^m$ 和$\mathbf{R}^n$ 上的$\ell_\infty$-范数导出的范数, 用$\|X\|_\infty$ 表示, 被称为最大行和范数,

$$\|X\|_\infty = \sup\left\{\|Xu\|_\infty \mid \|u\|_\infty \leqslant 1\right\} = \max_{i=1,\cdots,m} \sum_{j=1}^{n} |X_{ij}|.$$

而由$\mathbf{R}^m$ 和$\mathbf{R}^n$ 上的$\ell_1$-范数导出的范数, 用$\|X\|_1$ 表示, 被称为最大列和范数,

$$\|X\|_1 = \max_{j=1,\cdots,n} \sum_{i=1}^{m} |X_{ij}|$$

## A.1.6 对偶范数

令$\|\cdot\|$ 为$\mathbf{R}^n$ 上的范数。对应的对偶范数，用$\|\cdot\|_*$ 表示，定义为

$$\|z\|_* = \sup\left\{z^T x \mid \|x\| \leqslant 1\right\}.$$

(可验证这是一个范数。) 对偶范数可以解释为$z^T$ 的算子范数，由$1\times n$ 矩阵在$\mathbf{R}^n$ 上的范数$\|\cdot\|$ 和$\mathbf{R}$ 上的绝对值导出：

$$\|z\|_* = \sup\left\{\left|z^T x\right| \mid \|x\| \leqslant 1\right\}.$$

从对偶范数的定义我们可以得到对所有的$x$ 和$z$ 都成立的不等式

$$z^T x \leqslant \|x\|\|z\|_*.$$

该不等式在下述意义下是紧致的：对任意$x$，存在$z$ 使不等式成为等式。(类似地，对任意$z$ 存在$x$ 使等式成立。) 对偶范数的对偶就是原范数：对所有的$x$ 我们有$\|x\|_{**} = \|x\|$。( (该性质对无限维向量空间不一定成立。) Euclid 范数的对偶还是Euclid 范数，因为

$$\sup\left\{z^T x \mid \|x\|_2 \leqslant 1\right\} = \|z\|_2.$$

(该结果基于Cauchy-Schwarz 不等式；对非零$z$，使$z^T x$ 在$\|x\|_2 \leqslant 1$ 上达到最大的$x$ 的值为$z/\|z\|_2$ 。)

此外，$\ell_\infty$-范数的对偶是$\ell_1$-范数:

$$\sup\left\{z^T x \mid \|x\|_\infty \leqslant 1\right\} = \sum_{i=1}^{n} |z_i| = \|z\|_1,$$

而$\ell_1$-范数的对偶是$\ell_\infty$-范数。更一般的结论是，$\ell_p$-范数的对偶是$\ell_q$-范数，其中$q$ 满足$1/p+1/q=1$，即$q=p/(p-1)$ 。作为另外一个例子，考虑$\mathbf{R}^{m\times n}$ 上的$\ell_2$-范数或谱范数。对应的对偶范数是

$$\|Z\|_{2*} = \sup\left\{\operatorname{tr}\left(Z^T X\right) \mid \|X\|_2 \leqslant 1\right\},$$

它实际上就是奇异值之和，

$$\|Z\|_{2*} = \sigma_1(Z) + \cdots + \sigma_r(Z) = \operatorname{tr}\left(Z^T Z\right)^{1/2},$$

其中$r=\operatorname{rank} Z$ 。这个范数有时称为核范数。

Fenchel 不等式

$$f(x)+f^*(y)\geq y^Tx.$$

性质：若$f(x)$为凸函数，且$f(x)$ 是闭函数，则有

$$f^{**}=f.\quad \mathrm{f}\,(\mathrm{x})\text{ 为闭函数}\Leftrightarrow\{\mathrm{x}\ ||\mathrm{f}\,(\mathrm{x})|\leq a\}\text{ 为闭集。}$$

性质：设$f(x)$ 为凸函数，且可微，对于$z\in R^n$ ,若

$$y=\nabla f(z)$$

则

$$f^*(y)=z^T\nabla f(z)-f(z)$$

伸缩变换和复合仿射变换

若$a > 0$ 以及$b \in \mathbf{R}$, $g(x) = af(x) + b$ 的共轭函数为$g^*(y) = af^*(y/a) - b$。

设$A \in \mathbf{R}^{n\times n}$ 非奇异, $b \in \mathbf{R}^n$, 则函数$g(x) = f(Ax + b)$ 的共轭函数为

$$g^*(y) = f^*(A^{-T}y) - b^T A^{-T} y,$$

其定义域为$\mathrm{dom} g^* = A^T \mathrm{dom} f^*$。

独立函数的和如果函数$f(u, v) = f_1(u) + f_2(v)$,其中$f_1$ 和$f_2$ 是凸函数,且共轭函数分别为$f_1^*$ 和$f_2^*$,则

$$f^*(w, z) = f_1^*(w) + f_2^*(z).$$

换言之,独立凸函数的和的共轭函数是各个凸函数的共轭函数的和。("独立"的含义是各个函数具有不同的变量。)

- 拟凸函数

$$S_\alpha = \{x \mid f(x) \le \alpha, x \in \mathrm{dom} f\}\text{是凸集合}$$

- 拟凸函数

$$S_\alpha = \{x| -f(x) \le \alpha, x \in dom f\} \text{ 是凸集合}$$

- 拟线性函数：
  即是拟凸函数又是拟凹函数

- 对于拟线性函数，其定义水平集

$$\{x \mid f(x) = \alpha, x \in dom f\}\text{是凸集合}$$

# 拟凸函数例子

对数函数$\log(x)$

向量的长度：$f(x)=\begin{cases}\max\{i|x_i\neq 0\},x\neq 0,\\ 0,x=0\end{cases}$

$f(x)\leq\alpha\Leftrightarrow x_i=0,\forall i=\lfloor\alpha\rfloor+1,\ldots,n.$

线性分式函数：$f(x)=\frac{a^Tx+b}{c^Tx+d}$, $\mathrm{domf}=\{x\mid c^Tx+d>0\}$

上取整函数。函数是$\mathrm{ceil}(x)=\inf\{z\in Z\mid z\geq x\}$拟凸函数（也叫拟凹函数）

例3.31 考虑函数$f : \mathbf{R}^2 \to \mathbf{R}$, 其定义域为$\mathrm{dom}\, f = \mathbf{R}^2_+$,$f(x_1, x_2) = x_1x_2$。此函数既非凸函数, 亦非凹函数, 因为其Hessian 矩阵

$$\nabla^2 f(x) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

是不定的；它的两个特征值一个大于零一个小于零。然而，函数$f$是拟凹函数，因为对任意$\alpha$，函数的上水平集

$$\{x \in \mathbf{R}^2_+ \mid x_1x_2 \geqslant \alpha\}$$

都是凸集。（注意到函数$f$在$\mathbf{R}^2$上不是拟凸函数）

# 拟凸函数的基本性质

- 定理：函数$f(x)$为拟凸函数，当且仅当$\mathrm{dom}f$为凸集，且对$\forall x,y\in\mathrm{dom}f, 0\le\theta\le 1$, 有
$$f(\theta x+(1-\theta)y)\le\max\{f(x),f(y)\}$$
- 定理：若函数$f(x)$一阶可微，则$f(x)$为拟凸函数，$\mathrm{dom}f$为凸集，且对$\forall x,y\in\mathrm{dom}f$, 有
$$f(y)\le f(x)\Rightarrow\nabla^T f(x)(y-x)\le 0$$
- 定理：若函数$f(x)$二阶可微，且满足对$\forall x\in\mathrm{dom}f, y\in\mathcal{R}^n, y\ne 0$, 有
$$y^T\nabla f(x)=0\Rightarrow y^T\nabla^2 f(x)y>0$$
则函数$f(x)$为拟凸函数

# 拟凸函数的基本性质

定理：函数$f(x)$为拟凹函数，当且仅当$\mathrm{dom}f$为凸集，且对$\forall x, y \in \mathrm{dom}f, 0 \leq \theta \leq 1$, 有$f(\theta x + (1-\theta)y) \geq \min\{f(x), f(y)\}$

证明 First, suppose that $f$ is quasi-concave, i.e., that $U_f(a) = \{x | f(x) \geq a\}$ is a convex set for each $a \in \mathbb{R}$. Let $x, y \in \mathcal{D}$ and $\lambda \in (0,1)$. Assume, without loss of generality, that $f(x) \geq f(y)$. Letting $f(y) = a$, we have $x, y \in U_f(a)$. By the convexity of $U_f(a)$, we have $\lambda x + (1-\lambda)y \in U_f(a)$, which means

$$f[\lambda x + (1-\lambda)y] \geq f(y) \geq a = f(y) = \min\{f(x), f(y)\}$$

Now, suppose we have $f[\lambda x + (1-\lambda)y] \geq \min\{f(x), f(y)\}$ for all $x, y \in \mathcal{D}$ and for all $\lambda \in (0,1)$. Let $a \in \mathbb{R}$. If $U_f(a)$ is empty or contains only one point, it is evidently convex, so suppose it contains at least two points $x$ and $y$. Then $f(x) \geq a$ and $f(y) \geq a$, so $\min\{f(x), f(y)\} \geq a$. Now, for any $\lambda \in (0,1)$, we have $f[\lambda x + (1-\lambda)y] \geq \min\{f(x), f(y)\}$ by hypothesis, and so $\lambda x + (1-\lambda)y \in U_f(a)$. Since $a$ was arbitrary, the proof is complete for the case of quasi-concave functions.

拟凹函数的一阶最优条件

Let $f : \mathcal{D} \to \mathbb{R}$ be a $C^1$ function where $\mathcal{D} \subset \mathbb{R}^n$ is convex and open. Then $f$ is a quasi-concave function on $\mathcal{D}$ if and only if it is the case that for any $x, y \in \mathcal{D}$,

$$f(y) \geq f(x) \Rightarrow Df(x)(y-x) \geq 0.$$

**Proof** First, suppose $f$ is quasi-concave on $\mathcal{D}$, and let $x, y \in \mathcal{D}$ be such that $f(y) \geq f(x)$. Let $t \in (0,1)$. Since $f$ is quasi-concave, we have

$$f(x+t(y-x)) = f((1-t)x+ty) \geq \min\{f(x), f(y)\} = f(x).$$

Therefore, it is the case that for all $t \in (0,1)$ :

$$\frac{f(x+t(y-x)) - f(x)}{t} \geq 0.$$

As $t \to 0+$, the LHS of this expression converges to $Df(x)(y-x)$, so $Df(x)(y-x) \geq 0$, establishing one part of the result.

Now suppose that for all $x, y \in \mathcal{D}$ such that $f(y) \geq f(x)$, we have $Df(x)(y-x) \geq 0$. Pick any $x, y \in \mathcal{D}$, and suppose without loss of generality that $f(x) = \min\{f(x), f(y)\}$. We will show that for any $t \in [0,1]$, we must also have $f[(1-t)x + ty] \geq \min\{f(x), f(y)\}$, establishing the quasi-concavity of $f$.
For notational simplicity, let $z(t) = (1-t)x + ty$.
Define $g(t) = f[x + t(y-x)]$. Note that $g(0) = f(x) \leq f(y) = g(1)$; and that $g$ is $C^1$ on $[0,1]$ with $g\prime(t) = Df[x + t(y-x)](y-x)$. We will show that if $t^* \in (0,1)$ is any point such that $f[z(t^*)] \leq f(x)$ (i.e., such that $g(t^*) \leq g(0)$), we must have $g\prime(t^*) = 0$. This evidently precludes the possibility of having any point $\hat{t} \in (0,1)$ such that $g(\hat{t}) < g(0)$, and the desired result is established.
So suppose that $t^* \in (0,1)$ and we have $f(x) \geq f[z(t^*)]$. Then, by hypothesis, we must also have
$Df[z(t^*)](x - z(t^*)) = -t^* Df[z(t^*)](y-x) \geq 0$. Since $t > 0$, this implies $g\prime(t^*) \leq 0$. On the other hand, since it is also true that $f(y) \geq f(x)$, we have $f(y) \geq f[z(t^*)]$, so we must also have $Df[z(t^*)](y - z(t^*)) = (1-t^*) Df[z(t^*)](y-x) \geq 0$. Since $t^* < 1$, this implies in turn that $g\prime(t^*) \geq 0$.
It follows that $g\prime(t^*) = 0$.

# 拟凸函数的基本性质

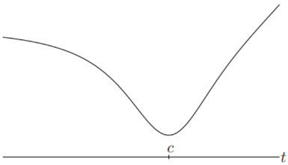



图3.11 R 上的拟凸函数。函数在$t \leqslant c$ 时非增，在$t \geqslant c$ 时非减。

**R 上的拟凸函数**

对**R** 上的拟凸函数，我们给出一个简单的刻画。由于考虑一般的函数较为繁琐，所以我们考虑连续函数。连续函数$f : \mathbf{R} \to \mathbf{R}$ 是拟凸的，当且仅当下述条件至少有一个成立。

- 函数$f$ 是非减的；
- 函数$f$ 是非增的；
- 存在一点$c \in \operatorname{dom} f$，使得对于$t \leqslant c$ (且$t \in \operatorname{dom} f$),$f$ 非增，对于$t \geqslant c$(且$t \in \operatorname{dom} f$),$f$ 非减。

点$c$ 可以在$f$ 的全局最小点中任选一个。图3.11描述了这样的情形。

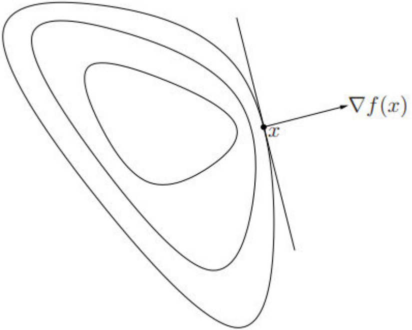



图3.12 拟凸函数$f$ 的三条等值线如图所示。梯度向量$\nabla f(x)$ 在点$x$ 处定义了水平集$\{z|f(z) \leqslant f(x)\}$ 的一个支撑超平面。

# 保持拟凸性的算子

- 非负权值函数的最大值函数

  $f = \max\{w_1 f_1, \cdots, w_n f_n\}, \quad w_i \geq 0, i = 1, \cdots, n$

- 复合函数

  如果$g : R^n \to R$ 是拟凸函数，且函数$h : R \to R$ 是非减的，则$f = h \circ g$ 是拟凸的。

- 最小化

  如果函数$f(x, y)$ 是$x$ 和$y$ 的联合拟凸函数，且C 是凸集，则函数$g(x) = \inf_{y \in C} f(x, y)$ 是拟凸的。

# 拟凸函数的凸函数族表示

- 若$f(x)$ 为拟凸函数, 根据$f(x)$ 的任意$t$ 下水平集, 我们可以构造一个凸函数族$\phi_t(x)$ ，使得
$$f(x) \leq t \Leftrightarrow \phi_t(x) \leq 0$$
- 性质：若$\phi_t(x)$ 为拟凸函数$f(x)$ 的凸函数族表示, 对每一个$x \in \mathrm{dom} f$, 若$s \geq t$, 则有
$$\phi_s(x) \leq \phi_t(x).$$

可以选取: $\phi_t(x) = \begin{cases} 0, f(x) \leq t, \\ \infty, \text{ else} \end{cases}$

$\phi_t(x) = \mathrm{dist}(x, \{z | f(z) \leq t\})$, 如果$f$ 的下水平集是闭集。

- 例凸凹函数之比。设$\mathbf{p}(\mathbf{x})$ 是凸函数, $\mathbf{q}(\mathbf{x})$ 是凹函数, 在凸集$\mathbf{C}$ 上, $\mathbf{p}(\mathbf{x})$ 非负, $\mathbf{q}(\mathbf{x})$ 大于零, 定义在集合$\mathbf{C}$ 上的函数$f(x) = \mathbf{p}(x)/\mathbf{q}(x)$, 则$f$ 是拟凸函数。

  可以选取:

$$\phi_t(x) = p(x) - tq(x), t \geq 0$$

# 对数凸函数

## 定义

函数$f(x)$称为对数凸（凹）函数，若函数$f(x)$满足：

1.dom$f$为凸集

2.$f(x) > 0$

3.log$f(x)$为凸（凹）函数

## 命题

$f(x)$为对数凸函数，当且仅当对$1/f(x)$对数凹函数。

## 定理

函数$f(x)$的定义域为凸集，且$f(x) > 0$，则$f(x)$为对数凸函数，当且仅当对$\forall x, y \in \mathrm{dom} f, 0 \leq \theta \leq 1$
有
$$f(\theta x + (1-\theta)y) \leq f(x)^{\theta} f(y)^{1-\theta}$$

# 对数凸函数

**命题**

- 若$f$是对数凸函数，则$f$是凸函数。
  利用函数复合性质。
- 如$f$是凹函数，$f > 0$,则$f$为对数凹函数。
  利用函数复合性质。
- 如$f$为对数凸函数，则$f$为拟凸函数。
  利用对数函数的单调性。
- 如$f$为对数凹函数，则$f$为拟凹函数。
  利用对数函数的单调性。

例3.39 一些简单的对数-凹函数和对数-凸函数

- 仿射函数。函数$f(x)=a^Tx+b$在$\{x\mid a^Tx+b>0\}$上是对数-凹函数。
- 幂函数。函数$f(x)=x^a$在$\mathbf{R}_{++}$上当$a\le 0$时是对数-凸函数，当$a\geqslant 0$时是对数-凹函数。
- 指数函数。函数$f(x)=e^{ax}$既是对数-凸函数也是对数-凹函数。
- Gauss概率密度函数的累积分布函数
$$\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-u^2/2}du,$$
是对数-凹函数（参见习题3.54）。
- Gamma函数。Gamma函数
$$\Gamma(x)=\int_0^{\infty}u^{x-1}e^{-u}du,$$
当$x\geqslant 1$时是对数-凸函数（参见习题3.52）。
- 行列式。det X在$\mathbf{S}^n_{++}$上是对数-凹函数。
- 行列式与迹之比。$\det X/\operatorname{tr}X$在$\mathbf{S}^n_{++}$上是对数-凹函数（参见习题3.49）。

# 对数凸函数和凹函数的性质

**定理**

函数$f(x)$二阶可微，则$f(x)$为对数凸函数当且仅当
$$f(x)\nabla^2 f(x) \succ= \nabla f(x)\nabla f(x)^T$$

- 性质：对数凸性与凹性对函数乘积和正数数乘运算均保持封闭。
- 性质：对数凸性对函数加运算保持封闭。但对数凹性对函数加运算不封闭。
- 性质：对数凹性对卷积运算保持封闭。

**定理**

- 函数$f(x,y)$对每一个$y \in C$在$x$上对数凸，则函数$g(x) = \int_C f(x,y)dy$也是对数凸函数。

**定理**

函数$f(x,y): \mathcal{R}^n \times \mathcal{R}^m \to \mathcal{R}$为对数凹函数，则函数$g(x) = \int f(x,y)dy$是对数凹函数。

**定义**

广义单调性 设$K \subseteq \mathcal{R}^n$为真锥，函数$f:\mathcal{R}^n \to \mathcal{R}$称为$K$-单调增，若函数$f(x)$满足：

$$x \preceq_K y \Rightarrow f(x) \leq f(y)$$

**定义**

广义凸函数 设$K \subseteq \mathcal{R}^n$为真锥，函数$f:\mathcal{R}^n \to \mathcal{R}^m$称为$K$-凸，若函数$f(x)$满足对$\forall x, y \in \mathrm{dom} f, 0 \leq \theta \leq 1$均有$f(\theta x + (1-\theta)y) \preceq_K \theta f(x) + (1-\theta) f(y)$.

**定理**

定理（对偶等价）：函数$f(x)$为$K$-凸函数，当且仅当对所有$w \succeq_{K^*} 0$, $w^T f(x)$为凸函数。

# 作业（1）

- 3.16
- 3.18
- 3.21
- 3.45
- 3.49(a)(b)(c)

# 第四章
# 凸优化问题

# 优化问题的基本形式

- 优化问题的标准描述：

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_j(x) = 0, \quad j = 1, \ldots, p
\end{aligned}
$$

优化变量 $x \in \mathcal{R}^n$
不等式约束 $f_i(x) \le 0$
等式约束 $h_j(x) = 0$
无约束优化 $m = p = 0$

# 优化问题的基本形式

- 优化问题的域：$D = \bigcap_{i=0}^{m} \mathrm{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathrm{dom}\, h_i$
- 可行点(解) (feasible) $x \in D$ 满足约束条件
- 可行域(可解集) 所有点的集合
- 最优值

$$p^* = \inf \{f_0(x) \mid f_i(x) \leq 0, i = 1, \ldots, m, h_i(x) = 0, i = 1, \ldots, p\}$$

- 最优化解:若$x^*$可行，且$p^* = f_0(x^*)$，最优解集为：$X_{opt} = \{x | x \in X_{feasible}, f_0(x) = p^*\}$

例4.2 框约束，考虑优化问题：

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & l_i \leqslant x_i \leqslant u_i, \quad i = 1, \cdots, n, \end{array}$$

其中$x \in \mathbf{R}^n$为优化变量。这些约束称为变量的界或框约束。
将该问题表示为标准形式：

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & l_i \leqslant x_i \leqslant u_i, \quad i = 1, \cdots, n, \\ & x_i - u_i \leqslant 0, \quad i = 1, \cdots, n. \end{array}$$

这里有$2n$个不等式约束函数：

$$f_i(x) = l_i - x_i, \quad i = 1, \cdots, n,$$

及

$$f_i(x) = x_{i-n} - u_{i-n}, \quad i = n+1, \cdots, 2n$$

# 局部最优问题

局部最优问题（存在$R>0$）,可行解$x$是问题局部最优：

$$
\begin{array}{lll}
\text{minimize} & f_0(\mathrm{z}), & x \in \mathcal{R}^n \\
\text{subject to} & f_i(z) \le 0, & i = 1, \ldots, m \\
& h_i(z) = 0, & j = 1, \ldots, p \\
& \|z - x\|_2 \le R, R > 0 &
\end{array}
$$

例子：

- $f_0(x) = 1/x, \operatorname{dom} f_0 = \mathbf{R}_{++} : p^\star = 0$，没有最优点
- $f_0(x) = -\log x, \operatorname{dom} f_0 = \mathbf{R}_{++} : p^\star = -\infty$
- $f_0(x) = x \log x, \operatorname{dom} f_0 = \mathbf{R}_{++} : p^\star = -1/e, x = 1/e$是最优
- $f_0(x) = x^3 - 3x, p^\star = -\infty$，局部最优点$x = 1$。

# 等价问题（缩放（Scaling））

**定理**

定理：若$\alpha_i > 0, i = 0, \ldots, m, \beta_i \neq 0, i = 1, \ldots, p$，则原优化问题与以下优化问题等价：

$$
\begin{aligned}
\text{minimize} \quad & \tilde{f}_0(x) = \alpha_0 f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} \quad & \tilde{f}_i(x) = \alpha_i f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& \tilde{h}_i(x) = \beta_i h_i(x) = 0, \quad i = 1, \ldots, p
\end{aligned}
$$

# 等价问题（变量替换）

**定理**

定理：设$\phi : \mathcal{R}^n \rightarrow \mathcal{R}^n$为一一对应，且$D \subseteq \phi(\mathrm{dom}\,\phi)$，则原优化问题与以下优化问题等价：

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(z) = f_0(\phi(z)), \quad z \in \mathcal{R}^n \\ \text{subject to} & \tilde{f}_i(z) = f_i(\phi(z)) \leq 0, \quad i = 1, \ldots, m \\ & \tilde{h}_i(z) = h_i(\phi(z)) = 0, \quad i = 1, \ldots, p \end{array}$$

则$z$是变量替换后优化问题的最优解，$x = \phi(z)$是原优化问题最优解。

# 目标函数和约束函数的变换

**定理**

定理：设$\psi_0 : \mathcal{R} \to \mathcal{R}$为严格单调增函数；$\psi_1, \dots, \psi_m$满足$\psi_i(u) \le 0$当且仅当$u \le 0$；$\omega_1, \dots, \omega_p$满足$\omega_i(u) = 0$当且仅当$u = 0$。则原优化问题与以下优化问题等价：

$$\begin{array}{lll} \text{minimize} & \tilde{f}_0(z) = \psi_0\left(f_0(z)\right), & z \in \mathcal{R}^n \\ \text{subject to} & \tilde{f}_i(z) = \psi_i\left(f_i(z)\right) \le 0, & i = 1, \dots, m \\ & \tilde{h}_i(z) = \omega_i\left(h_i(z)\right) = 0, & \mathrm{i} = 1, \dots, p \end{array}$$

例：最小化范数和范数平方：

$$\min \|Ax - b\|_2, x \in R^n$$

$$\Leftrightarrow \min \|Ax - b\|_2^2, \quad x \in R^2$$

# 松弛变量

**定理**

标准形式的优化问题与以下优化问题等价：

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} \quad & f_i(x) + s_i = 0, \quad i = 1, \dots, m \\
& s_i \geq 0, \quad i = 1, \dots, m \\
& h_i(x) = 0, \quad j = 1, \dots, p
\end{aligned}
$$

$s_i$称为不等式约束$f_i(x) \leq 0$的松弛变量。

# 松弛变量

## 定理

设凸优化问题：

$$
\begin{aligned}
&\text{minimize} \quad f_0(x), x \in \mathcal{R}^n \\
&\text{subject to} \quad a_i^T x \le b_i, \quad i = 1, \ldots, m
\end{aligned}
$$

等价于：

$$
\begin{aligned}
&\text{minimize} \quad f_0(x), \quad x \in \mathcal{R}^n \\
&\text{subject to} \quad a_i^T x + s_i = b_i, \quad i = 1, \ldots, m \\
&\qquad\qquad\qquad s_i \ge 0, \quad i = 1, \ldots, m
\end{aligned}
$$

## 定理

定理：设$\phi : \mathcal{R}^k \to \mathcal{R}^n$是这样的方式：$x$满足等式$h_i(x) = 0, j = 1, \ldots, p$成立等价于存在一些$z \in \mathcal{R}^n$使得$x = \phi(z)$。则优化问题

$$\begin{array}{ll} \text{minimize} & f_0(x), \quad x \in \mathcal{R}^n \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \ldots, m \\ & h_j(x) = 0, \quad j = 1, \ldots, p \end{array}$$

与以下优化问题等价：

$$\begin{array}{ll} \text{minimize} & f_0(\phi(z)), \quad x \in \mathcal{R}^n \\ \text{subject to} & f_i(\phi(z)) \leq 0, \quad i = 1, \ldots, m \end{array}$$

非凸功率控制问题：

$$\text{maximize} \sum_{i=1}^{n} \frac{Lw_i h_i p_i}{\sum_{k \neq i} h_k p_k + L h_i p_i + \sigma^2} \tag{9}$$

$$\text{subject to} \sum_{j=1}^{n} g_j p_j \leq T, \tag{10}$$

$$p_j \geq 0, j = 1, \cdots, n. \tag{11}$$

问题(9)-(11) 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.

# 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:*

$$\sum_{i=1}^{n} \frac{L w_i h_i p_i}{\sum_{k \neq i} h_k p_k + \sigma^2 + L h_i p_i} = \sum_{i=1}^{n} \frac{L w_i a_i}{((L-1)a_i + 1)} \quad (12)$$

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

$$p_i = \frac{\sigma^2 a_i}{\left(1 - \sum\limits_{i=1}^{n} a_i\right) h_i}, \forall i \in \{1, \ldots, n\}. \quad (13)$$

Using lemma 2 and $\sum_{i=1}^{n} a_i = \sum_{i=1}^{n} h_i p_i / \left( \sum_{i=1}^{n} h_i p_i + \sigma^2 \right) < 1$, (9)-(11) are equivalent to the following problem:

$$\text{maximize } \sum_{i=1}^{n} \frac{L w_i a_i}{((L-1)a_i + 1)} \tag{14}$$

$$\text{subject to } \sigma^2 \sum_{i=1}^{n} \frac{g_i a_i}{h_i} + T \sum_{i=1}^{n} a_i \leq T, \tag{15}$$

$$a_i \geq 0, i = 1, \cdots, n. \tag{16}$$

- 定理：设凸优化问题

$$
\begin{array}{rl}
\text{minimize} & f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} & f_i(x) \leq 0, i = 1, \ldots, m, Ax = b
\end{array}
$$

等价于

$$
\begin{array}{rl}
\text{minimize} & f_0\left(Fz + x_0\right), x \in \mathcal{R}^n \\
\text{subject to} & f_i\left(Fz + x_0\right) \leq 0, \quad i = 1, \ldots, m
\end{array}
$$

其中
$Ax = b \Leftrightarrow x = Fz + x_0$, $F \in R^{n*k}$为满足$R(F) = N(A)$的矩阵，$x_0$为等式约束任意可行解。$z \in R^k$, 可以选择$F$为满秩矩阵，从而有$k = n - rank(A)$, 邓加厚优化问题不含有等式约束并且减少了$rank(A)$个变量。

# 引入等式约束

$$\begin{array}{ll}\text{minimize} & f_0\left(A_0x+b_0\right)\\ \text{subject to} & f_i\left(A_ix+b_i\right)\leqslant 0, i=1,\cdots,m, h_i(x)=0, i=1,\cdots,p\end{array}$$

我们引入新的变量$y_i \in \mathbf{R}^{k_i}$ 和新的等式约束$y_i = A_ix+b_i, i=0,\cdots,m$,
从而构造等价问题

$$\begin{array}{ll}\text{minimize} & f_0\left(y_0\right)\\ \text{subject to} & f_i\left(y_i\right)\leqslant 0, \quad i=1,\cdots,m\\ & y_i=A_ix+b_i, \quad i=0,\cdots,m\\ & h_i(x)=0, \quad i=1,\cdots,p\end{array}$$

该问题含有$k_0+\cdots+k_m$ 个新变量:

$$y_0\in\mathbf{R}^{k_0}, \quad \cdots, \quad y_m\in\mathbf{R}^{k_m}$$

及$k_0+\cdots+k_m$ 个新的等式约束:

$$y_0=A_0x+b_0, \quad \cdots, \quad y_m=A_mx+b_m$$

# 优化部分变量

- 性质：

$$\inf_{x,y} f(x,y) = \inf_x \tilde{f}(x), \text{其中} \tilde{f}(x) = \inf_y f(x,y)$$

- 定理：优化问题

$$\begin{array}{ll} \text{minimize} & f_0(x_1,x_2), \quad x_1 \in R^{n_1}, x_2 \in R^{n_2}, x = (x_1,x_2) \in R^{n_1+n_2} \\ \text{subject to} & f_i(x_1) \le 0, \quad i = 1,\ldots,m_1 \\ & \tilde{f}_i(x_2) \le 0, \quad i = 1,\ldots,m_2 \end{array}$$

可以分离变量$x_1, x_2$

$$\text{令} \tilde{f}_0(x) = \inf \left\{ f_0(x_1,z) \mid \tilde{f}_i(\mathrm{z}) \le 0, \mathrm{i} = 1,\cdots,\mathrm{m}_2 \right\}$$

等价问题：

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(x) \\ \text{subject to} & f_i(y_i) \leqslant 0, \quad i = 1,\cdots,m_1. \end{array}$$

例4.4
在约束下优化二次函数的部分变量。考虑具有严格凸的二次目标的问题，其中某些变量不受约束：

$$\begin{array}{ll} \text{minimize} & x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2 \\ \text{subject to} & f_i(x_1) \leqslant 0, \quad i = 1, \cdots, m \end{array}$$

这里，我们可以解析地优化$x_2$：

$$\inf_{x_2}\left(x_1^T P_{11} x_1 + 2x_1^T P_{12} x_2 + x_2^T P_{22} x_2\right) = x_1^T\left(P_{11} - P_{12} P_{22}^{-1} P_{12}^T\right) x_1$$

(参见§A.5.5)。因此，原问题等价于

$$\begin{array}{ll} \text{minimize} & x_1^T\left(P_{11} - P_{12} P_{22}^{-1} P_{12}^T\right) x_1 \\ \text{subject to} & f_i(x_1) \leqslant 0, \quad i = 1, \cdots, m. \end{array}$$

# 上境图问题形式

$$
\begin{array}{ll}
\text{minimize} & t \\
\text{subject to} & f_0(x) - t \le 0, \\
& f_i(x) \le 0, \quad i = 1, \dots, m \\
& h_i(x) = 0, j = 1, \dots, p
\end{array}
$$

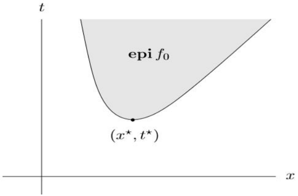



图4.1 无约束问题的上境图形式问题的几何解释。其目标是寻找上境图中(阴影所示) 极小化$t$ 的一点，即上境图中最“低” 的一点。最优点是$(x^\star, t^\star)$ 。

# 隐式与显式约束

标准形式问题可以表示为如下一个无约束问题

$$x \in \mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i$$

- 我们把$\mathcal{D}$认为问题的可行域。
- 约束$f_i(x) \leq 0, h_i(x) = 0$是显式约束。
- 如果问题没有显式约束，则它是不受约束的$(m = p = 0)$

例:

$$\text{minimize} f_0(x) = -\sum_{i=1}^{k} \log\left(b_i - a_i^T x\right)$$

是一个具有隐式约束的无约束问题$a_i^T x < b_i$

# 可行解问题

可行解问题

$$\begin{array}{rll} \text{find} & x & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & i = 1, \ldots, p \end{array}$$

可以被认为是一般问题的一个特例$f_0(x) = 0$:

$$\begin{array}{rll} \text{minimize} & 0 & \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & i = 1, \ldots, p \end{array}$$

- - $p^\star = 0$ 如果约束是可行的；任何可行的$x$ 是最优的
- - $p^\star = \infty$ 如果约束不可行

# 标准形式凸优化问题

- 凸优化问题的基本描述：

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), \quad x \in \mathcal{R}^n \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p
\end{aligned}
$$

$f_i(x)$ 为凸函数
$h_i(x)$ 为仿射函数
若$f_0(x)$ 为拟凸函数, 则优化问题称为拟凸优化问题。
性质：凸优化问题的可行域是凸集。

- 例

$$\begin{aligned} \text{minimize} \quad & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} \quad & f_1(x) = x_1 / \left(1 + x_2^2\right) \leq 0 \\ & h_1(x) = (x_1 + x_2)^2 = 0 \end{aligned}$$

$f_0$ 凸集, 可行集合$\{(x_1, x_2) \,|x_1 = -x_2, x_1 \leq 0\}$ 是凸集，非凸优化问题的标准形式, $\mathrm{h}_1$ 不是仿射, $\mathrm{f}_1$ 不是凸函数
等价于凸优化问题

$$\begin{aligned} \text{minimize} \quad & f_0(x) = x_1^2 + x_2^2 \\ \text{subject to} \quad & \tilde{f}_1(x) = x_1 \leq 0 \\ & h_1\tilde{(}x) = (x_1 + x_2)^2 = 0 \end{aligned}$$

凸优化问题的任意局部极小点都是全局最优

证明：假设$x$是局部极小，$y$全局最优且$f_0(y) < f_0(x)$.
$x$ 局部最优意味着存在$R > 0$使得

$$z\text{可行}, \quad \|z - x\|_2 \le R \quad \Longrightarrow \quad f_0(z) \ge f_0(x).$$

考虑$z = \theta y + (1-\theta)x$ 且$\theta = R/(2\|y - x\|_2)$

- $\|y - x\|_2 > R$, 因此$0 < \theta < 1/2$
- $z$ 是两个可行点的凸组合，因此也可行。
- $\|z - x\|_2 = R/2$，并且

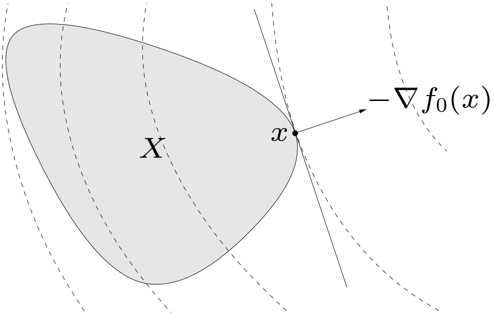
$$f_0(z) \le \theta f_0(x) + (1-\theta) f_0(y) < f_0(x),$$

这与$x$是局部极小的假设矛盾.

# 可微凸优化问题的最优性条件

$x$ 是凸优化问题$\min_{x\in X} f_0(x)$最优解当且仅当$x$可行且满足：

$$\nabla f_0(x)^T(y-x) \geq 0, \quad \forall y \in X.$$

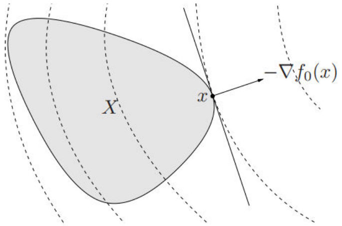



如果$\nabla f_0(x)$非零，它定义了可行集$X$在$x$处的支撑超平面.

# 可微凸优化问题的最优性条件

- 定理: 设$X$ 为凸优化问题的可行域, $f_0(x)$ 可微。则$x$为最优解当且仅当$x \in X, \nabla f_0(x)^T(y-x) \geq 0, y \in X$成立, 其中$X = \{x | f_i(x) \leq 0, i = 1, \cdots, m, h_i(x) = 0, i = 1, \cdots, p\}$

- 定理: 无约束凸优化问题中, 若$f_0(x)$ 可微。则$x$ 为最优解当且仅当$\nabla f_0(x) = 0$ 成立。



例4.5 无约束二次规划。考虑极小化二次函数

$$f_0(x) = (1/2)x^T Px + q^T x + r$$

其中, $P \in \mathbf{S}_+^n$(保证了$f_0$为凸函数)。x 为$f_0$的最小解的充要条件是:

$$\nabla f_0(x) = Px + q = 0$$

根据这个(线性) 方程无解、有唯一解或有多解的不同, 有几种可能的情况

- 如果$q \notin \mathcal{R}(P)$ , 则无解。此类情况$f_0$ 无下界。
- 如果$P \succ 0$ (意味着$f_0$ 严格凸), 则存在唯一的最小解$x^\star = -P^{-1}q$。
- 如果P奇异但$q \in \mathcal{R}(P)$ , 则最优解集合为(仿射) 集合$X_{\text{opt}} = -P^\dagger q + \mathcal{N}(P)$,其中$P^\dagger$ 表示P 的伪逆(参见$SA.5.4$ )。

例4.6 解析中心。考虑(无约束的)极小化(凸)函数$f_0 : \mathbf{R}^n \to \mathbf{R}$ 的问题,

$$f_0(x) = -\sum_{i=1}^m \log(b_i - a_i^T x), \quad \mathbf{dom}\, f_0 = \{x \mid Ax \prec b\}$$

其中$a_1^T, \cdots, a_m^T$ 表示$A$ 的行向量。函数$f_0$ 可微,因此$x$ 是最优解的充要条件为:

$$Ax \prec b, \quad \nabla f_0(x) = \sum_{i=1}^m \frac{1}{b_i - a_i^T x} a_i = 0. \quad (4.23)$$

(条件$Ax \prec b$ 即是$x \in \mathbf{dom} f_0$)如果$Ax \prec b$ 不可行,则$f_0$ 的定义域为空。假设$Ax \prec b$可行、存在几种可能的情况(参见习题4.2):

- 式(4.23) 无解,则问题无最优解。这种情况当且仅当$f_0$ 无下界时发生。
- 式(4.23) 有多解,在这种情况下。可以证明其解构成了一个仿射集合。
- 式(4.23) 有唯一解,即$f_0$ 有唯一的极小值。这种情况当且仅当开多面体$\{x \mid Ax \prec b\}$有界非空时发生。

# 等式凸优化问题

**定理**

设凸优化问题仅有等式约束

$$\begin{aligned} \text{minimize} \quad & f_0(x), x \in \mathcal{R}^n \\ \text{subject to} \quad & Ax = b \end{aligned}$$

则$x$为最优解当且仅当$x \in X$, 且存在向量$\nu$ 满足

$$\nabla f_0(x) + A^T \nu = 0$$

**定理**

设凸优化问题

$$\begin{aligned} \text{minimize} \quad & f_0(x), x \in \mathcal{R}^n \\ \text{subject to} \quad & x \succ= 0 \end{aligned}$$

则$x$为最优解当且仅当$x \succ= 0$，且

$$\nabla f_0(x) \succ= 0, \nabla \Big(f_0(x)\Big)_i x_i = 0, i = 1, \cdots, n.$$

# 拟凸优化问题

- 拟凸优化问题的基本描述

$$
\begin{aligned}
\text{minimize} \quad & f_0(x), x \in \mathcal{R}^n \\
\text{subject to} \quad & f_i(x) \le 0,\ i = 1, ..., m \\
& h_i(x) = 0,\ j = 1, ..., p
\end{aligned}
$$

$f_0(x)$为拟凸函数,$f_1(x), ..., f_m(x)$为凸函数。
注:拟凸优化问题的局部最优解不一定是全局最优解。

# 通过凸可行问题求解拟凸函数

- 设$\phi_t(x)$为拟凸函数$f_0(x)$ 的凸函数族表示，即

$$f_0(x) \le t \Leftrightarrow \phi_t(x) \le 0$$

则拟凸优化问题的可行解问题为

$$\begin{array}{rl} \text{find} & x \\ \text{subject to} & \phi_t(x) \le 0, \\ & f_i(x) \le 0,\ i = 1, ..., m \\ & Ax = b \end{array}$$

设$p^*$为拟凸优化问题的最优解，若上述问题可解，则$p^* \le t$；否则$p^* \ge t$ 。

# 求解拟凸优化问题的二分法

给定一个足够小的$l$和足够大的$u$，使得区间$[l,u]$能包含最优解$p^*$, $l \leq p^* \leq u^*$, 给定容忍度$\varepsilon > 0$
重复:

- 令$t = (l+u)/2$
- 求解可行解问题;
- 若可解, 则令$u = t$，否则令$l = t$

- LP问题的一般描述

$$
\begin{array}{ll}
\text{minimize} & c^T x + d \\
\text{subjectto} & Gx \prec= h \\
 & Ax = b \\
G \in \mathcal{R}^{m\times n}, & A \in \mathcal{R}^{p\times n}
\end{array}
$$

# 线性规划（linear program，LP）

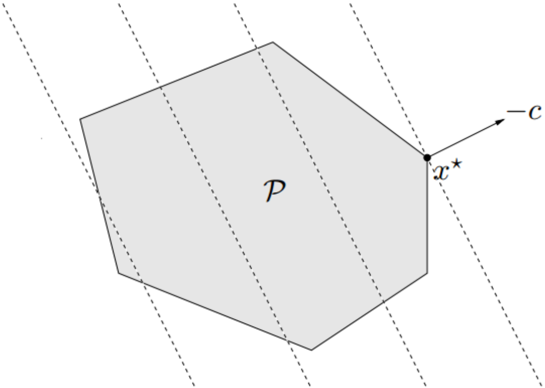



图4.4 线性规划的几何解释。可行集$\mathcal{P}$ 是多面体，如阴影所示。目标$c^Tx$ 是线性的，所以其等位曲线是与$c$ 正交的超平面(如虚线所示)。点$x^\star$ 是最优的，它是$\mathcal{P}$ 中在方向$-c$ 上最远的点。

# 线性规划的标准形式和不等式形式

- 标准形式LP问题

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subjectto} \quad & Ax = b \\ & x \succ= 0 \end{aligned}$$

- 不等式形式LP问题

$$\begin{aligned} \text{minimize} \quad & c^T x + d \\ \text{subjectto} \quad & Ax \prec= b \end{aligned}$$

# 将线性规划转化为标准形式

$$\begin{aligned} \text{minimize} \quad & c^T x + d \\ \text{subject to} \quad & Gx \prec= h \\ & Ax = b \end{aligned}$$

通过引入松弛变量$s$和$x = x^+ - x^-, x^+, x^- \geq 0.$

$$\begin{aligned} \text{minimize} \quad & c^T x^+ - c^T x^- + d \\ \text{subject to} \quad & Gx^+ - Gx^- + s = h \\ & Ax^+ - Ax^- = b \\ & x^+ \succ= 0, x^- \succ= 0, s \succ= 0 \end{aligned}$$

# LP问题的例子

- 多面体的切比雪夫中心
- 分片线性函数最小化
- 线性分式规划

# 多面体的切比雪夫中心

- 多面体　$P = \{x \in R^n \mid a_i^T x \leq b_i, i = 1, \cdots, m\}$

- Chebyshev center:到多面体边界距离最远的内点（最深的点）

- 问题描述

$$
\begin{aligned}
&\text{maximize} \quad r \\
&\text{subject to} \quad B(x_c, r) \subseteq P
\end{aligned}
$$

$$B(x_c, r) = \{x_c + u \mid \|u\|_2 \leq r\}$$

- LP形式

$$
\begin{aligned}
&\text{minimize} \quad -r \\
&\text{subject to} \quad a_i^T x_c + r\|a_i\|_2 \leq b_i, i = 1, \cdots, m
\end{aligned}
$$

# 分片线性函数最小化

- 问题描述

$$\text{minimize} \quad f(x) = \max_{i=1,\ldots,m} \left(a_i^T x + b_i\right)$$

- 上半图形式

$$\begin{aligned} &\text{maximize} && t \\ &\text{subject to} && \max_{i=1,\ldots,m} \left(a_i^T x + b_i\right) \le t \end{aligned}$$

- LP形式

$$\begin{aligned} &\text{minimize} && t \\ &\text{subject to} && a_i^T x + b_i \le t, i = 1, \cdots, m \end{aligned}$$

# 线性分式规划

- 问题描述

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) = \frac{c^T x + d}{e^T x + f}, \mathrm{dom} f_0 = \{x | e^T x + f > 0\} \\
\text{subject to} \quad & Gx \prec= h \\
& Ax = b
\end{aligned}
$$

- LP形式

$$
\begin{aligned}
\text{minimize} \quad & c^T y + dz \\
\text{subject to} \quad & Gy - hz \prec= 0 \\
& Ay - bz = 0 \\
& e^T y + fz = 1 \\
& z \geq 0
\end{aligned}
$$

$$
y = \frac{x}{e^T x + f} \qquad z = \frac{1}{e^T x + f}
$$

- 食谱问题：选择$n$种食品的数量：$x_1, \ldots, x_n$
  - 每种食品含有热量$c_j$,并且$a_{ij}$表示第$j$种食物含有的第$i$种营养.
  - 每天摄入的营养含量不低于$b \in \mathbb{R}^m$.

  故良好的饮食条件需满足：

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Ax \geq b, x \geq 0 \end{aligned} \tag{17}$$

- 分段线性最小化问题

$$\min \quad \max_{i=1,\ldots,m} (a_i^T x + b_i) \tag{18}$$

  也即求一个线性规划问题:

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & a_i^T x + b_i \leq t, \quad i = 1, \ldots, m \end{aligned} \tag{19}$$

基追踪问题是压缩感知中的一个基本问题，可以写为

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \|x\|_1, \\ \text{s.t.} \quad & Ax = b. \end{aligned} \tag{20}$$

对每个$|x_i|$ 引入一个新的变量$z_i$，可以将问题(20) 转化为

$$\begin{aligned} \min_{z\in\mathbb{R}^n} \quad & \sum_{i=1}^n z_i, \\ \text{s.t.} \quad & Ax = b, \\ & -z_i \le x_i \le z_i, \quad i = 1, 2, \cdots, n, \end{aligned} \tag{21}$$

这是一个线性规划问题．