凸优化

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
   - 半定规划

2. 凸优化理论与应用

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.

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$\operatorname{epi} f$为闭集，则称$f$为闭函数.

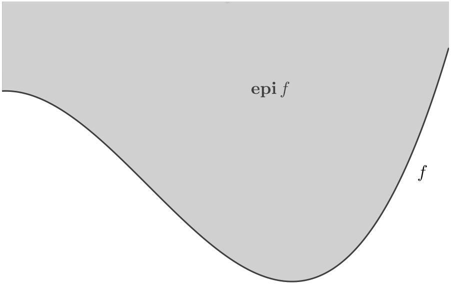


Figure: 函数f和其上镜图$\operatorname{epi} f$

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。

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.

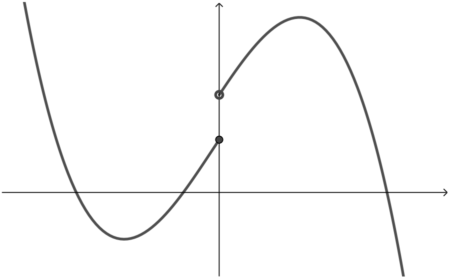


Figure: 下半连续函数$f(x)$

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.

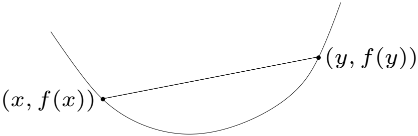
凸函数的定义

定义 (凸函数)

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

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

对所有$x, y \in \mathrm{dom}\, f$, $0 \le \theta \le 1$ 都成立，则称f 是凸函数



- 若f 是凸函数，则$-f$ 是凹函数
- 若对所有$x, y \in \mathrm{dom}\, f$, $x \ne y, 0 < \theta < 1$，有

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

则称f 是严格凸函数

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

$f:\mathbb{R}^n\rightarrow\mathbb{R}$ 是凸函数，当且仅当对每个$x\in\mathbf{dom}\,f, v\in\mathbb{R}^n$，函数$g:\mathbb{R}\rightarrow\mathbb{R}$ 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．

凸函数判定定理

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

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

而等式左边有

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

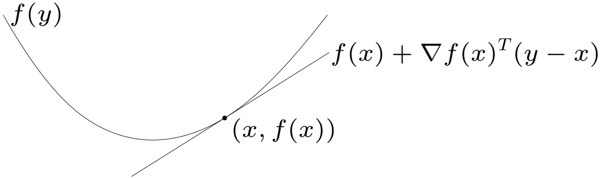
这说明$f(x)$是凸函数． □

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

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

将上述第一个不等式两边同时乘t，第二个不等式两边同时乘$1-t$，相加得

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

这正是凸函数的定义，因此充分性成立．□

梯度单调性

定理

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

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

Proof.

必要性：若f可微且为凸函数，根据一阶条件，我们有

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

将两式不等号左右两边相加即可得到结论.

Proof.

充分性：若∇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$\text{epi}f$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$\text{epi}f$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$f(x) = (1/2)x^T Px + q^T x + r$ (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)?

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

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

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

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：

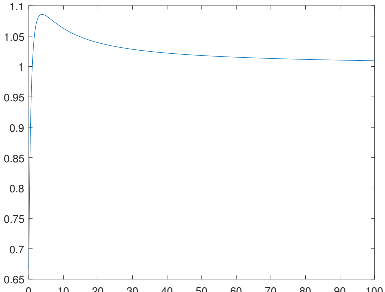
Zhengqiang Wang, Lingge Jiang, Chen He. Optimal Price-Based Power Control Algorithm in Cognitive Radio Networks. IEEE Transactions on Wireless Communications, 2014, 13(11):5909–5920, Nov. 2014.

功率控制问题

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:*

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

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^{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:

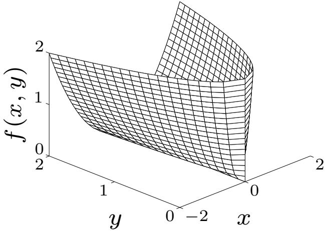
$$f(X) = \mathrm{tr}(A^T X) + b = \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}X_{ij} + b$$

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

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$\mathbf{dom}\, f = \mathbf{R}_{++}$是凹的。
其Hessian矩阵$\nabla^2 f(x)$可以通过下面两个式子给出：

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

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$\left(a^T a\right)\left(b^T b\right) \geqslant \left(a^T b\right)^2$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:

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

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$epif = \{(x,t)|x \in \mathrm{dom} f, f(x) \leq t\}$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$f:\mathbf{R}^n\times\mathbf{S}^n\to\mathbf{R}$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：

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

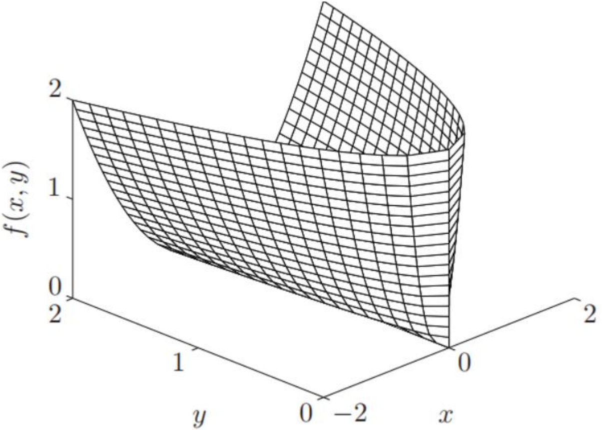
通过舒尔补条件判断分块矩阵的半正定性（见附录A.5.5）。最后一个条件关于(x,Y,t)中的线性矩阵不等式，因此epi$f$是凸的。

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

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

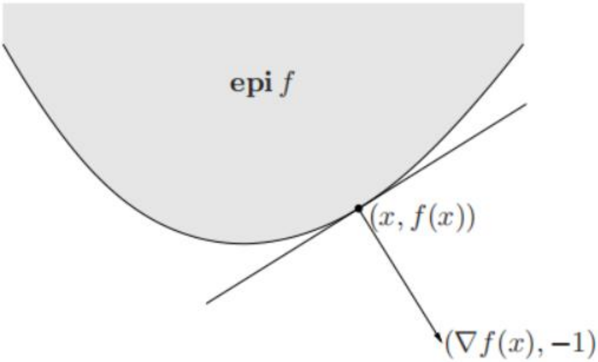
上式可以描述为

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

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$0 \le \theta \le 1$,

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

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

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

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:

$$\sqrt{ab} \leq \frac{a+b}{2}, a, b \geq 0.$$

- Hölder不等式：

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

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

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

$$\mathrm{domf} = \{x \mid Ax + b \in \mathrm{domf}\}$$

- 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。

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

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：

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

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$\mathbf{dom} f=\mathbf{R}^{p\times q}$，其中$\|\cdot\|_2$表示谱范数或者最大奇值。函数$f$可以表述为

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

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

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

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

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

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

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

- 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$k=1$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，$\mathbf{p} \geq \mathbf{1}, g^p(x)$ 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。

例子，复合定理中h需要满足的条件

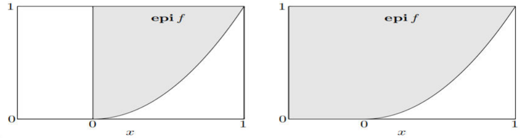


图3.7 左。函数x^2, 其定义域为$\mathbf{R}_+$, 在其定义域内是凸且非减的, 但是其扩展值延伸不是非减的。右。函数$\max\{x,0\}^2$, 其定义域为$\mathbf{R}$, 函数是凸函数, 且其扩展值延伸是非减的。

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

对$n=1$，g,h均可微的情形，我们给出简证

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

推论

- 如果g_i 是正值凹函数，则$\sum_{i=1}^m \log g_i(x)$ 是凹函数
- 如果g_i 是凸函数，则$\log\sum_{i=1}^m \exp g_i(x)$是凸函数

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$h(z)=\log\left(\sum_{i=1}^k \mathrm{e}^{z_i}\right)$ 是凸函数且在每一维分量上非减，因此只要g_i 是凸函数，$\log\left(\sum_{i=1}^k \mathrm{e}^{g_i}\right)$ 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

$f(x,y)=x^TAx+2x^TBy+y^TCy$，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

透视函数

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。

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

例3.18 Euclid范数的平方.$\mathbf{R}^n$上的凸函数$f(x) = x^Tx$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。

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

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。

透视函数

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：

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

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：

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

其中：

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

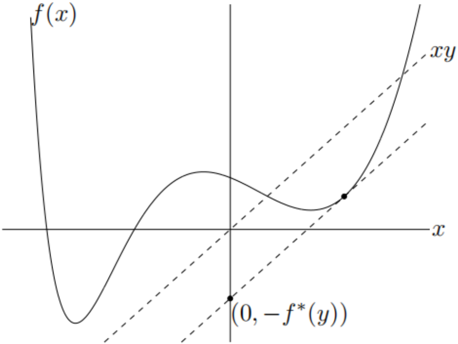
其中g是凸的。

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：

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



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$x$。

考虑R上一些凸函数的共轭函数：

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

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

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

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

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

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$Y \not\prec 0$，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$y^T x - f(x)$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$1^T y = 1$时上述方程有解。将y_i的表达式代入$y^T x - f(x)$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$1^T y = 1$，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$1^T y = 1, y \succeq 0$。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，$y^T x - f(x)$无上界。

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

$$y^T x - f(x) = t\mathbf{1}^T y - t - \log n.$$

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

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

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。

共轭函数(conjugate function)

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

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

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$z\in\mathbf{R}^n$，$\|z\|\le 1$ 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$\|y\|_*\le 1$, 对于任意x,我们可得$y^Tx\le\|x\|\|y\|$。即对于任意x, $y^Tx-\|x\|\le 0$。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ℓ_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

$$\langle X, Y\rangle = \operatorname{tr}\left(X^T Y\right) = \sum^{m}\sum^{n} X_{ij} Y_{ij},$$

其对于任意$X, Y \in \mathbf{R}^{m\times n}$成立。(此处$\mathbf{tr}$ 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ℓ_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$f:\mathbf{R}^n\to\mathbf{R}, \mathrm{dom} f=\mathbf{R}^n$ 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$\|x\|_{\mathrm{symb}}$ 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

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

我们用$\mathrm{dist}(x,y)$ 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ℓ_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ℓ_1-范数，以及定义为

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

的Chebyshev 或ℓ_∞-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ℓ_p-范数表示，定义为

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

取$p = 1$ 就得到ℓ_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

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

因此ℓ_∞-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

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

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α 和β 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，相同的收玫序列，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$\|\cdot\|_{\mathrm{a}}$和$\|\cdot\|_{\mathrm{b}}$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$\|\cdot\|_{\mathrm{a}}$和$\|\cdot\|_{\mathrm{b}}$导出的算子范数

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

(可以验证, 该式确实定义了$\mathbf{R}^{m\times n}$上的一个范数。) 当$\|\cdot\|_{\mathrm{a}}$和$\|\cdot\|_{\mathrm{b}}$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:

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

(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ℓ_2-范数。作为一个例子, 考虑由$\mathbf{R}^m$和$\mathbf{R}^n$上的ℓ_∞-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ℓ_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，ℓ_∞-范数的对偶是ℓ_1-范数：

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

而ℓ_1-范数的对偶是ℓ_∞-范数。更一般的结论是，ℓ_p-范数的对偶是ℓ_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ℓ_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：$f(x) = \frac{a^T x + b}{c^T x + d}, \mathrm{domf} = \{x \mid c^T x + d > 0\}$

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）

拟凹函数例子

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

$$\nabla^2 f(x)=\left[\begin{array}{cc}0 & 1\\ 1 & 0\end{array}\right]$$

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α，函数的上水平集

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

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$\forall x, y \in \mathrm{dom} f, 0 \leq \theta \leq 1$, 有
$$f(\theta x + (1-\theta)y) \leq \max\{f(x), f(y)\}$$
- 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
$$f(y) \leq f(x) \Rightarrow \nabla^T f(x)(y-x) \leq 0$$
- 定理：若函数$f(x)$二阶可微，且满足对$\forall x \in \mathrm{dom} f, y \in \mathcal{R}^n, y \neq 0$, 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

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

证明 First, suppose that f is quasi-concave, i.e., that $U_f(a) = \{x | f(x) \ge a\}$ 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 $f(x) \ge f(y)$. 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

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

Now, suppose we have $f[\lambda x + (1-\lambda)y] \ge \min\{f(x), f(y)\}$ 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 $f(x) \ge a$ and $f(y) \ge a$, so $\min\{f(x), f(y)\} \ge a$. Now, for any $\lambda \in (0, 1)$, we have $f[\lambda x + (1-\lambda)y] \ge \min\{f(x), f(y)\}$ 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 $g\prime(t^*) = 0$.

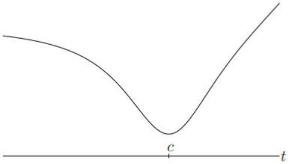


图3.11 R 上的拟凸函数。函数在$t \leqslant c$ 时非增，在$t \geqslant c$ 时非减。

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$c \in \mathrm{dom} f$，使得对于$t \leqslant c$ (且$t \in \mathrm{dom} f$),f 非增，对于$t \geqslant c$(且$t \in \mathrm{dom} f$),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。

拟凸函数的基本性质

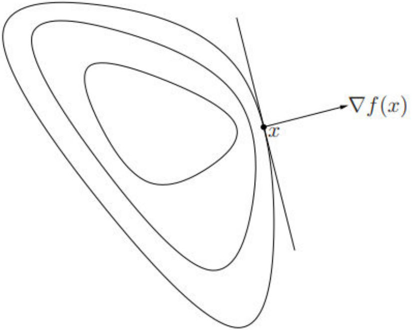


图3.12 拟凸函数f 的三条等值线如图所示。梯度向量$\nabla f(x)$ 在点x 处定义了水平集$\{z|f(z) \leqslant f(x)\}$ 的一个支撑超平面。

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。

- 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

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

- 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$s \ge t$, 则有

$$\phi_s(x) \le \phi_t(x).$$

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

$\phi_t(x) = \mathrm{dist}(x, \{z | f(z) \le t\})$, 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$\forall x, y \in \mathrm{dom} f, 0 \le \theta \le 1$

有
$$f(\theta x + (1-\theta)y) \le f(x)^{\theta} f(y)^{1-\theta}$$

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。

对数凹和对数凸函数例子

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$a\leq 0$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：

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

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：$D=\bigcap_{i=0}^{m}\operatorname{dom}f_i\cap\bigcap_{i=1}^{p}\operatorname{dom}h_i$
- 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

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

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

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：

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

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$\phi : \mathcal{R}^n \to \mathcal{R}^n$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：

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

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；$\psi_1, \ldots, \psi_m$满足$\psi_i(u) \leq 0$当且仅当$u \leq 0$；$\omega_1, \ldots, \omega_p$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：

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

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：

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

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：

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

等价于：

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

消除等式约束

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

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

与以下优化问题等价：

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

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.

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

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

等价于

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

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

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

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：

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

- 定理：优化问题

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

可以分离变量x_1, x_2

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

等价问题：

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

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

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

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

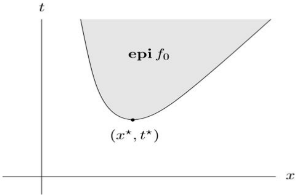


图4.1 无约束问题的上境图形式问题的几何解释。其目标是寻找上境图中(阴影所示) 极小化t 的一点，即上境图中最“低”的一点。最优点是$(x^\star, t^\star)$ 。

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：

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

$f_i(x)$ 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。

凸优化问题的例

- 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, h_1 不是仿射, f_1 不是凸函数
等价于凸优化问题

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

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

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

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

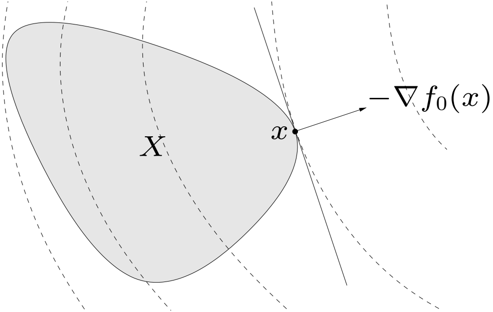
$$f_0(z) \leq \theta f_0(x) + (1 - \theta)f_0(y) < f_0(x),$$

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：

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

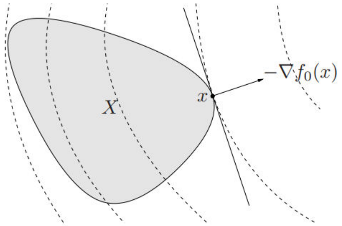


如果$\nabla f_0(x)$非零，它定义了可行集X在x处的支撑超平面.

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$\mathcal{S}A.5.4$)。

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,

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

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:

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

(条件$Ax\prec b$ 即是$x\in\textbf{dom}f_0$)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ν 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

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

$f_0(x)$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^*, $l \le p^* \le u^*$, 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）

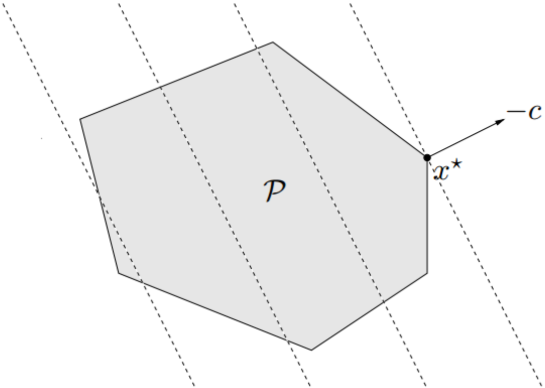


图4.4 线性规划的几何解释。可行集$\mathcal{P}$ 是多面体，如阴影所示。目标$c^T x$ 是线性的，所以其等位曲线是与c 正交的超平面(如虚线所示)。点$x^\star$ 是最优的，它是$\mathcal{P}$ 中在方向$-c$ 上最远的点。

线性规划的标准形式和不等式形式

- 标准形式LP问题

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

- 不等式形式LP问题

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

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

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

- Chebyshev center of a polyhedron
- Piecewise-linear minimization
- Linear-fractional programming

Chebyshev center of a polyhedron

- 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

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

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

- LP形式

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

Piecewise-linear minimization

- 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

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

- LP形式

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

Linear-fractional programming

- 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：$x_1,...,x_n$
  - 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

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

也即求一个线性规划问题:

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

基追踪问题

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.

- 除此之外，也可以引入x_i的正部和负部,其中$x_i^+ = \max\{x_i, 0\}$, $x_i^- = \max\{-x_i, 0\}$..
- 利用$x_i = x_i^+ - x_i^-$, $|x_i| = x_i^+ + x_i^-$，则问题(20) 转化为的另外一种等价的线性规划形式可以写成

$$\begin{aligned} \min_{x^+, x^- \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} (x_i^+ + x_i^-), \\ \text{s.t.} \quad & Ax^+ - Ax^- = b, \\ & x^+, x^- \geq 0. \end{aligned}$$

可以看出这也是一个线性规划问题，且与原问题等价.

数据拟合

在数据拟合中，除了常用的最小二乘模型外，还有最小ℓ_1 范数模型

$$\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_1, \tag{22}$$

和最小ℓ_∞ 范数模型

$$\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_\infty. \tag{23}$$

这两个问题都可以转化成线性规划的形式.

- 对于问题(22)，通过引入变量$y=Ax-b$，可以得到如下等价问题：

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

- 利用基追踪问题中类似的技巧，可以将上述绝对值优化问题转化成线性规划问题.

数据拟合

- 对于问题(23)，令$t = \|Ax - b\|_\infty$，则得到等价问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n,\, t\in\mathbb{R}} \quad & t, \\ \text{s.t.} \quad & \|Ax - b\|_\infty \le t. \end{aligned}$$

- 利用ℓ_∞ 范数的定义，可以进一步写为

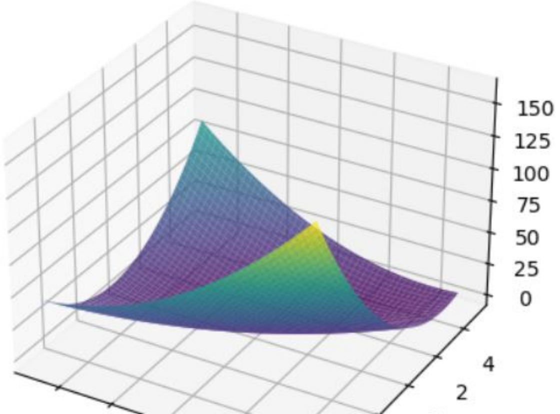
$$\begin{aligned} \min_{x\in\mathbb{R}^n,\, t\in\mathbb{R}} \quad & t, \\ \text{s.t.} \quad & -t\mathbf{1} \le Ax - b \le t\mathbf{1}, \end{aligned}$$

这是一个线性规划问题.

二次规划问题（Quadratic Programming, QP）

$$\min \quad f(x_1, x_2) = x_1^2 + 2x_2^2 - 2x_1x_2 + 2x_1 - 6x_2$$

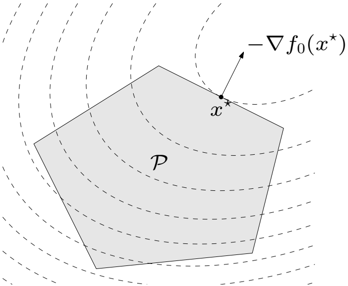
$$\text{s.t.} \quad \begin{aligned} & x_1 + x_2 \geq 2 \\ & x_1 - 2x_2 \leq 4 \\ & 2x_1 + x_2 = 6 \end{aligned}$$



二次规划问题（Quadratic Programming, QP）

$$
\begin{aligned}
\min \quad & \frac{1}{2}x^T P x + q^T x + r \\
\text{s.t.} \quad & Gx \le h \\
& Ax = b
\end{aligned}
\tag{24}
$$

- $P \in \mathcal{S}_+^n$,故目标函数是二次的
- 在一个多面体内最小化一个二次凸问题



二次规划的应用

- 最小二乘问题：
$$\min \|Ax-b\|_2^2 \tag{25}$$
  - 该问题的解析解为$x^* = A^\dagger b$ (其中$A^\dagger$为广义逆)
- 随机线性规划

$$\begin{aligned} \min \quad & \bar{c}^T x + \gamma x^T \Sigma x = \mathbf{E} c^T x + \gamma \mathrm{var}(c^T x) \\ \text{s.t.} \quad & Gx \le h, \quad Ax = b \end{aligned}$$

- c 是随机向量并且均值为$\bar{c}$，方差为Σ
- $c^T x$ 均值为$\bar{c}^T x$,方差为$x^T \Sigma x$
- $\gamma > 0$ 为风险参数,控制预期成本与风险.

带有二次约束的二次规划问题(QCQP)

考虑带有二次约束的二次规划问题：

$$\begin{array}{ll} \min & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{s.t.} & (1/2)x^T P_i x + q_i^T x + r_i \le 0, \quad i = 1, ..., m \\ & Ax = b \end{array}$$

- $P_i \in \mathbb{S}_+^n$;即目标函数与约束均为二次凸的
- 如果$P_1, ..., P_m \in \mathbb{S}_{++}^n$, 则可行域为$m$个椭球与一个仿射集的交集.

广义不等式约束

- 凸问题的广义不等式约束

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \preceq_{K_i} 0, \quad i = 1, \dots, m \\ & Ax = b \end{aligned}$$

  - $f_0 : \mathbb{R}^n \to \mathbb{R}$ 为凸函数; $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ 关于适当锥K_i为K_i-凸的.
  - 与标准凸问题有相同的性质

- 锥形式问题(具有仿射目标函数与约束的特殊情况)

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Fx + g \preceq_K 0 \\ & Ax = b \end{aligned}$$

将线性规划问题延伸到了非多面体锥上($K = \mathbb{R}^m_+$)

QP问题的例子

- Least-squares and regression
- Distance between polyhedra

Least-squares and regression

- 问题描述

$$
\begin{aligned}
&\text{minimize} \quad \|Ax - b\|_2^2 \\
&\|Ax - b\|_2^2 = x^T A^T A x - 2b^T A x + b^T b
\end{aligned}
$$

具有解析解$x = A^\dagger b$，$A^\dagger$是A的伪逆

多面体间距离

- 问题描述：R^n上多面体
  $P_1=\{x\mid A_1x\preceq= b_1\}$和$P_2=\{x\mid A_2x\preceq= b_2\}$的（Euclid）距离定义为
  $$\operatorname{dist}(P_1,P_2)=\inf\{\|x_1-x_2\|\mid x_1\in P_1,x_2\in P_2\}$$
  $$P_1=\{x\mid A_1x\preceq= b_1\}\quad P_2=\{x\mid A_2x\preceq= b_2\}$$

- QP形式
  $$\begin{array}{ll}\text{minimize} & \|x_1-x_2\|_2^2\\ \text{subject to} & A_1x\preceq= b_1, A_2x\preceq= b_2\end{array}$$

这一问题无可行解的充要条件是，其中一个多面体是空的。其最优解为零的充要条件是，多面体相交，这种情况下，最优的x_1和x_2是相等的（并且是交集$\mathcal{P}_1\cap\mathcal{P}_2$中的点）。否则最优的$x_1$和$x_2$分别在$\mathcal{P}_1$和$\mathcal{P}_2$中，并且是最接近的。（我们将在第8章中更加详细地讨论涉及距离的几何问题。）

- SOCP问题的基本描述：

$$
\begin{aligned}
\text{minimize} \quad & f^T x \\
\text{subject to} \quad & \|A_i x + b\|_2 \leq c_i^T x + d_i, i = 1, \cdots, m \\
& Fx = g,
\end{aligned}
$$

- 二次锥约束条件：

$$
\|Ax + b\|_2 \leq c^T x + d
$$

当$c_i = 0, i = 1, \cdots, m$, SOCP等同于QCQP;
$A_i = 0, i = 1, \cdots, m$, SOCP退化为线性规划。

SOCP问题的例 – Robust linear programming

- 问题描述：

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \le b_i \end{array}$$

其中参数c,a_i和b_i含有一些不确定性或变化。

- 例c,b_i为确定的常数，a_i为变量，其范围满足：

$$a_i \in \varepsilon_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \le 1\}$$

- SOCP形式

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i^T x + \left\|P_i^T x\right\|_2 \le b_i \end{array}$$

几何规划(Geometric Programming，GP)

- 几何规划的基本描述：

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & f_i(x) \leq 1, i = 1, \ldots, m \\ & h_i(x) = 1, i = 1, \ldots, p \end{array}$$

其中f_i为正项式，h_i为单项式，$D = R^n_{++}$。

- 单项式与正项式$(c > 0, c_k > 0,)$：

$$f(x) = c x_1^{a_1} \cdots x_n^{a_n}$$

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} \cdots x_n^{a_{nk}}$$

- 正项式对于加法，数乘和非负的伸缩变化是封闭的。
- 单项式对于数乘和除是封闭的。
- 一个正项式和一个单项式相乘，其结果为一个正项式。
- 一个正项式除以一个单项式，其结果仍为正项式。

几何规划（一般）不是凸优化问题，但是通过变量代换以及目标、约束函数的转化，GP问题可以转化为凸优化问题。

- 令：

$$y_i = \log x_i$$

- 几何规划的凸形式：

$$\begin{array}{ll} \text{minimize} & \tilde{f}_0(y) = \log\left(\sum_{k=1}^{K_0} e^{a_{0k}^T y + b_{0k}}\right) \\ \text{subject to} & \tilde{f}_i(y) = \log\left(\sum_{k=1}^{K_i} e^{a_{ik}^T y + b_{ik}}\right) \leq 0, i = 1, \ldots, m \\ & \tilde{h}_i(y) = g_i^T y + h_i = 0, i = 1, \ldots, p \end{array}$$

广义不等式约束

- 广义不等式约束的优化问题：

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \preceq_{K_i} 0, i = 1, \dots, m \\ & Ax = b \end{aligned}$$

- SOCP的描述：

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & -\left(A_i x + b_i, c_i^T x + d_i\right) \preceq_{K_i} 0, i = 1, \dots, m \\ & Fx = g \\ & K_i = \left\{(y, t) \in R^{k_i+1} \mid \|y\|_2 \leq t\right\} \end{aligned}$$

半定规划问题的一般形式如下：

$$\begin{aligned} \min \quad & c^{\mathrm{T}}x, \\ \text{s.t.} \quad & x_1A_1 + x_2A_2 + \cdots + x_nA_n + B \preceq 0, \\ & Gx = h, \end{aligned} \tag{26}$$

其中$c \in \mathbb{R}^n, A_i \in \mathcal{S}^m, i = 1, 2, \cdots, m, B \in \mathcal{S}^m, G \in \mathbb{R}^{p\times n}, h \in \mathbb{R}^p$ 为已知的向量和矩阵，$x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$是自变量．

- 半定规划（SDP）是线性规划在矩阵空间中的一种推广．它的目标函数和等式约束均为关于矩阵的线性函数，而它与线性规划不同的地方是其自变量取值于半正定矩阵空间．
- 作为一种特殊的矩阵优化问题，半定规划在某些结构上和线性规划非常相似，很多研究线性规划的方法都可以作为研究半定规划的基础．

类似于线性规划问题，我们考虑半定规划的标准形式

$$\begin{aligned} \min \quad & \langle C, X\rangle, \\ \text{s.t.} \quad & \langle A_1, X\rangle = b_1, \\ & \cdots \\ & \langle A_m, X\rangle = b_m, \\ & X \succeq 0, \end{aligned} \tag{27}$$

和对偶形式：

$$\begin{aligned} \min \quad & -b^{\mathrm{T}}y, \\ \text{s.t.} \quad & y_1A_1 + y_2A_2 + \cdots + y_nA_n \preceq C. \end{aligned} \tag{28}$$

形如(26) 式的优化问题都可以转化成(27) 式或者(28) 式的形式.

LP与SDP

$$\text{LP:} \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax \leq b \end{array} \qquad \text{SDP:} \quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \mathrm{diag}(Ax - b) \preceq 0 \end{array}$$

SOCP 与SDP

$$\text{SOCP:} \quad \begin{array}{ll} \min & f^T x \\ \text{s.t.} & \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m \end{array}$$

$$\text{SDP:} \quad \begin{array}{ll} \min & f^T x \\ \text{s.t.} & \begin{bmatrix} (c_i^T x + d_i)I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m \end{array}$$

二次约束二次规划问题的半定规划松弛

- 考虑二次约束二次规划问题

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & x^{\mathrm{T}}A_0x + 2b_0^{\mathrm{T}}x + c_0, \\
\text{s.t.} \quad & x^{\mathrm{T}}A_ix + 2b_i^{\mathrm{T}}x + c_i \le 0, \quad i = 1, 2, \cdots, m,
\end{aligned} \tag{29}
$$

其中A_i 为$n \times n$ 对称矩阵．当部分A_i为对称不定矩阵时，问题(29)是NP 难的非凸优化问题．

- 写出问题(29) 的半定规划松弛问题．对任意$x \in \mathbb{R}^n$以及$A \in \mathcal{S}^n$，有恒等式

$$
x^{\mathrm{T}}Ax = \mathrm{Tr}(x^{\mathrm{T}}Ax) = \mathrm{Tr}(Axx^{\mathrm{T}}) = \left\langle A, xx^{\mathrm{T}} \right\rangle,
$$

因此问题(29) 中所有的二次项均可用下面的方式进行等价刻画：

$$
x^{\mathrm{T}}A_ix + 2b_i^{\mathrm{T}}x + c_i = \left\langle A_i, xx^{\mathrm{T}} \right\rangle + 2b_i^{\mathrm{T}}x + c_i.
$$

由上述分析，原始问题等价于

$$\begin{aligned}\min_{x\in\mathbb{R}^n} \quad & \langle A_0, X\rangle + 2b_0^{\mathrm{T}}x + c_0 \\ \text{s.t.} \quad & \langle A_i, X\rangle + 2b_i^{\mathrm{T}}x + c_i \leq 0, \quad i = 1, 2, \cdots, m, \\ & X = xx^{\mathrm{T}}.\end{aligned} \tag{30}$$

进一步地，

$$\begin{aligned} x^{\mathrm{T}}A_i x + 2b_i^{\mathrm{T}}x + c_i &= \left\langle \begin{pmatrix} A_i & b_i \\ b_i^{\mathrm{T}} & c_i \end{pmatrix}, \begin{pmatrix} X & x \\ x^{\mathrm{T}} & 1 \end{pmatrix} \right\rangle, \\ &\stackrel{\text{def}}{=\!=} \langle \overline{A}_i, \overline{X}\rangle, \quad i = 0, 1, \cdots, m. \end{aligned}$$

接下来将等价问题(30) 松弛为半定规划问题.

- 在问题(30) 中，唯一的非线性部分是约束$X = xx^{\mathrm{T}}$，我们将其松弛成半正定约束$X \succeq xx^{\mathrm{T}}$. 可以证明，$\overline{X} \succeq 0$ 与$X \succeq xx^{\mathrm{T}}$ 是等价的.
- 因此这个问题的半定规划松弛可以写成

$$\begin{aligned} \min \quad & \langle \overline{A_0}, \overline{X} \rangle \\ \text{s.t.} \quad & \langle \overline{A_i}, \overline{X} \rangle \leq 0, \quad i = 1, 2, \cdots, m, \\ & \overline{X} \succeq 0, \\ & \overline{X}_{n+1,n+1} = 1. \end{aligned}$$

其中“松弛”来源于我们将$X = xx^{\mathrm{T}}$ 替换成了$X \succeq xx^{\mathrm{T}}$.

最大割问题的半定规划松弛

- 令G为一个无向图，其节点集合为$V=\{1,2,\cdots,n\}$和边的集合为E. 令$w_{ij}=w_{ji}$为边$(i,j)\in E$上的权重，并假设$w_{ij}\geq 0,\ (i,j)\in E$. 最大割问题是找到节点集合$V$的一个子集$S$使得$S$与它的补集$\overline{S}$之间相连边的权重之和最大化.
- 可以将最大割问题写成如下整数规划的形式：令$x_j=1,\ j\in S$和$x_j=-1,\ j\in\overline{S}$，则

$$\begin{aligned} \max\quad & \frac{1}{2}\sum_{i<j}(1-x_ix_j)w_{ij} \\ \text{s.t.}\quad & x_j\in\{-1,1\},\ j=1,2,\cdots,n. \end{aligned} \tag{31}$$

- 在问题(31)中，只有当x_i与x_j不同时，目标函数中w_{ij}的系数非零. 最大割问题是一个离散优化问题，很难在多项式时间内找到它的最优解.

接下来介绍如何将问题(31) 松弛成一个半定规划问题.

- 令$W=(w_{ij})\in\mathcal{S}^n$，并定义$C=-\frac{1}{4}(\mathrm{Diag}(W\mathbf{1})-W)$为图$G$的拉普拉斯矩阵的$-\frac{1}{4}$倍，则问题(31) 可以等价地写为

$$\begin{aligned}\min\quad & x^\mathrm{T}Cx,\\ \mathrm{s.t.}\quad & x_i^2=1,\ i=1,2,\cdots,n.\end{aligned}$$

  由于目标函数是关于x的二次函数，可将其等价替换为$\langle C,xx^\mathrm{T}\rangle$.
- 接下来令$X=xx^\mathrm{T}$，注意到约束$x_i^2=1$，这意味着矩阵X对角线元素$X_{ii}=1$. 因此利用矩阵形式我们将最大割问题转化为

$$\begin{aligned}\min\quad & \langle C,X\rangle,\\ \mathrm{s.t.}\quad & X_{ii}=1,\ i=1,2,\cdots,n,\\ & X\succeq 0,\ \mathrm{rank}(X)=1.\end{aligned}\tag{32}$$

- 问题(32) 和(31) 是等价的，这是因为$X=xx^\mathrm{T}$ 可以用约束$X\succeq 0$和$\mathrm{rank}(X)=1$等价刻画.

- SDP问题描述：

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n + G \preceq 0 \\ & Ax = b \end{array}$$

其中$F_i, G \in \mathbf{S}^k$。

- 不等式约束包含多个线性矩阵不等式（LMI）：

$$x_1 \hat{F}_1 + \cdots + x_n \hat{F}_n + \hat{G} \preceq 0, \quad x_1 \tilde{F}_1 + \cdots + x_n \tilde{F}_n + \tilde{G} \preceq 0$$

等价于单个LMI：

$$x_1 \begin{bmatrix} \hat{F}_1 & 0 \\ 0 & \tilde{F}_1 \end{bmatrix} + x_2 \begin{bmatrix} \hat{F}_2 & 0 \\ 0 & \tilde{F}_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \hat{F}_n & 0 \\ 0 & \tilde{F}_n \end{bmatrix} + \begin{bmatrix} \hat{G} & 0 \\ 0 & \tilde{G} \end{bmatrix} \preceq 0$$

- LP与等价的SDP：

$$\text{LP}: \text{minimize} \quad c^T x$$
$$\text{subject to} \quad Ax \preceq b$$

$$\text{SDP}: \text{minimize} \quad c^T x$$
$$\text{subject to} \quad diag(Ax - b) \preceq 0$$

- SOCP与等价的SDP：

$$\text{SOCP}: \text{minimize} \quad f^T x$$
$$\text{subject to} \quad \|A_i x + b_i\|_2 \leq c_i^T x + d_i, \quad i = 1, \ldots, m$$

$$\text{SDP}: \text{minimize} \quad f^T x$$
$$\text{subject to} \quad \begin{bmatrix} \left(c_i^T x + d_i\right) I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m$$

特征值最小化

- 特征值最小化：

$$\text{minimize} \quad \lambda_{\max}(A(x))$$

给定$A_i \in \mathbf{S}^k$，$A(x) = A_0 + x_1A_1 + \cdots + x_nA_n$。

- 等价SDP：

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & A(x) \preceq tI \end{array}$$

其中$x \in \mathbf{R}^n, t \in \mathbf{R}$。

$$\lambda_{\max}(A) \leq t \quad \Longleftrightarrow \quad A \preceq tI$$

第5章
对偶问题

优化问题的拉格朗日函数

- 设优化问题：

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

- 拉格朗日函数$L : D \times R^n \times R^p \to R$：

$$
L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x)
$$

- 对固定的x,拉格朗日函数$L(x, \lambda, v)$为关于λ和v的仿射函数。

拉格朗日对偶函数

- 拉格朗日对偶函数(lagrange dual function)：

$g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v) = \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x)\right)$

若拉格朗日函数没有下界，则令

$$g(\lambda, v) = -\infty$$

- 拉格朗日对偶函数为凹函数。
- 对$\forall\lambda \succ= 0$和$\forall v$，若原最优化问题有最优值p^*，则

$$g(\lambda, v) \leq p^*$$

对偶函数的例子

- 线性方程的最小二乘解
- 标准形式的线性规划问题
- 双向划分问题

- 原问题：

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

- 拉格朗日函数：
$$L(x, v) = x^T x + v^T (Ax - b)$$
- 拉格朗日对偶函数：
$$g(v) = -\frac{1}{4} v^T AA^T v - b^T v$$

$g(v)$是一个二次凹函数，定义域为R^p。利用对偶函数给出原问题下界的性质，从而对于任意的$v \in R^p$,有$-\frac{1}{4}v^T AA^T v - b^T v \leq inf\{x^T x | Ax = b\}$

标准形式的线性规划问题

- 原问题：

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \succeq 0 \end{aligned}$$

- 拉格朗日函数：

$$\begin{aligned} L(x, \lambda, v) &= c^T x - \lambda^T x + v^T (Ax - b) \\ &= -v^T b + \left(c - \lambda + A^T v\right)^T x \end{aligned}$$

- 拉格朗日对偶函数：

$$g(\lambda, v) = \begin{cases} -b^T v & A^T v - \lambda + c = 0 \\ -\infty & \text{其他情况} \end{cases}$$

注：对偶函数g只有在$R^m \times R^p$上的一个正常方式仿射子集上取值有限。

双向划分问题

- 原问题：

$$\begin{array}{ll} \text{minimize} & x^T W x, W \in S^n \\ \text{subject to} & x_i^2 = 1, i = 1, \ldots, n \end{array}$$

- 拉格朗日函数：

$$\begin{aligned} L(x, v) &= x^T W x + \sum_i^n v_i \left(x_i^2 - 1\right) \\ &= x^T (W + \mathrm{diag}(v)) x - 1^T v \end{aligned}$$

- 拉格朗日对偶函数：

$$g(v) = \begin{cases} -1^T v & W + \mathrm{diag}(v) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

- 弱对偶性: $p^* \geq -\mathbf{1}^T \nu$ 若$W + \mathrm{diag}(\nu) \succeq 0$
- 取$\nu = -\lambda_{\min}(W)\mathbf{1}$ 得最优解p^* 的一个下界的估计, $p^* \geq n\lambda_{\min}(W)$

- 共轭函数
$$f^*(y) = \sup_{x\in \mathrm{dom} f}\left(y^T x - f(x)\right)$$
- 共轭函数与对偶函数存在密切联系
- 具有线性不等式约束和线性等式约束的优化问题：

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & Ax \prec= b, Cx = d \end{aligned}$$

对偶函数：
$$\begin{aligned} g(\lambda, v) &= -b^T\lambda - d^T v - f_0^*\left(-A^T\lambda - C^T v\right) \\ \mathrm{dom} g &= \left\{(\lambda, v) \mid -A^T\lambda - C^T v \in \mathrm{dom} f_0^*\right\} \end{aligned}$$

实例:等式约束下的范数最小化

$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b \end{array}$$

对偶函数

$$g(\nu) = \inf_x(\|x\| - \nu^T Ax + b^T\nu) = \begin{cases} b^T\nu & \|A^T\nu\|_* \le 1 \\ -\infty & \text{其他} \end{cases}$$

其中 $\|v\|_* = \sup_{\|u\|\le 1} u^T v$ 是 $\|\cdot\|$ 的对偶范数

证明: 利用 $\inf_x(\|x\| - y^T x)$ 在 $\|y\|_* \le 1$ 时等于0 否则等于$-\infty$

- 若 $\|y\|_* \le 1$, 则 $\|x\| - y^T x \ge 0$ 对任意x 都成立, 当 $x = 0$ 时取等
- 若 $\|y\|_* > 1$, 取 $x = tu$, 其中 $\|u\| \le 1$, $u^T y = \|y\|_* > 1$:

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \text{ 当 } t \to \infty$$

弱对偶性: $p^* \ge b^T\nu$ 若 $\|A^T\nu\|_* \le 1$

熵的最大化

- 原始问题：

$$\begin{array}{ll} \text{minimize} & \sum\limits_{i=1}^{\text{n}} x_i \log x_i, \quad D = R_+^n \\ \text{subject to} & Ax \prec= b, 1^T x = 1 \end{array}$$

- 共轭函数：

$$f_0^*(y) = \sum_{i=1}^{n} e^{y_i - 1}$$

- 对偶函数：

$$\begin{aligned} g(\lambda, v) &= -b^T\lambda - v - f_0^*\left(-A^T\lambda - 1v\right) \\ &= -b^T\lambda - v - \sum_{i=1}^{m} e^{-a_i^T\lambda - v - 1} \\ &= -b^T\lambda - v - e^{-v-1} \sum_{i=1}^{m} e^{-a_i^T\lambda} \end{aligned}$$

- 原始问题：

$$\begin{array}{ll} \text{minimize} & \log\det X^{-1}, \quad D = S_{++}^n \\ \text{subject to} & a_i^T X a_i \leq 1, i = 1, \ldots, m \end{array}$$

- 共轭函数：

$$f_0^*(Y) = \log\det(-Y)^{-1} - n$$

- 对偶函数：

$$g(\lambda) = \begin{cases} \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - 1^T\lambda + n & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty & \text{otherwise} \end{cases}$$

- 拉格朗日对偶问题的描述：

$$\begin{aligned} \text{maximize} \quad & g(\lambda, v) \\ \text{subject to} \quad & \lambda \succ= 0 \end{aligned}$$

- 对偶可行域

$$\lambda \succ= 0$$
$$g(\lambda, v) > -\infty$$

- 最优值

$$d^*$$

- 最优解

$$(\lambda^*, v^*)$$

LP问题的对偶问题

- 标准LP问题

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & Ax = b, x \succ= 0 \end{array}$$

- 对偶函数

$$g(\lambda, v) = \begin{cases} -b^T v & A^T v - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}$$

- 对偶问题

$$\begin{array}{rl} \text{maximize} & g(\lambda, v) \\ \text{subject to} & \lambda \succ= 0 \end{array}$$

- 等价描述

$$\begin{array}{rl} \text{maximize} & g(\lambda, v) \\ \text{subject to} & A^T \nu - \lambda + c = 0, \lambda \succ= 0 \end{array}$$

LP问题的对偶问题

- 标准LP问题

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

- 对偶函数

$$g(\lambda, \nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \textit{otherwise} \end{cases}$$

- 对偶问题

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & \lambda \succ= 0 \end{aligned}$$

- 等价描述

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & A^T\nu - \lambda + c = 0 \\ & \lambda \succ= 0 \end{aligned}$$

弱对偶性

定理

弱对偶性：设原始问题的最优值为p^*,对偶问题的最优值为d^*,则$d^* \leq p^*$成立。

- 最优对偶间隙：optimal duality gap
$$p^* - d^*$$
- 可以利用对偶问题找到原始问题最优解的下界。

强对偶性

定义

强对偶性　设原始问题的最优值为p^*,对偶问题的最优值为d^*。若$d^* = p^*$成立，则称原始问题和对偶问题之间具有强对偶性。

- 强对偶性并不是总是成立的。
- 凸优化问题通常（但并不总是）具有强对偶性。

定理

*Slater*定理：若凸优化问题存在严格可行解，即存在$x \in \text{relint } D$，满足

$$f_i(x) < 0, i = 1, \cdots, m,$$
$$Ax = b$$

则优化问题具有强对偶性。该条件称为*Slater*条件。

Slater约束品性与强对偶原理

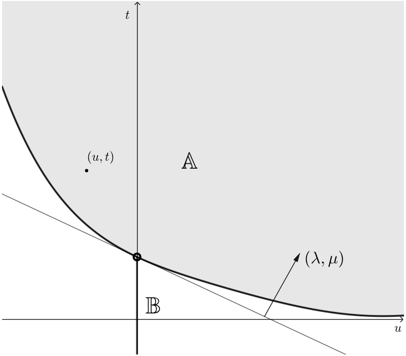
定理

若凸优化问题满足*Slater*条件, 则强对偶原理成立.

- 当$d^* > -\infty$ 时, 对偶问题的最优解可以取到, 即存在对偶可行解(λ^*, ν^*), 满足$g(\lambda^*, \nu^*) = d^* = p^*$.
- 假设集合$\mathcal{D}$ 内部非空(即relint$\mathcal{D} = \mathcal{D}$), A 行满秩(否则可以去掉多余的线性等式约束)以及原始问题最优函数值p^* 有限.
- 定义集合

$$
\begin{aligned}
\mathbb{A} = \{(u, v, t) \mid \exists x \in \mathcal{D},\ c_i(x) \le u_i,\ i = 1, 2, \cdots, m, \\
Ax - b = v, f(x) \le t\}. \\
\mathbb{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid s < p^*\}.
\end{aligned}
$$

- 可以证明集合$\mathbb{A}$ 和$\mathbb{B}$ 是不相交的.
- 假设$(u, v, t) \in \mathbb{A} \cap \mathbb{B}$.根据$(u, v, t) \in \mathbb{B}$, 有$u = 0, v = 0$ 和$t < p^*$.
- 由$(u, v, t) \in \mathbb{A}$, 可知$f(x) \le t < p^*$, 这与p^* 是原始问题最优值矛盾.



justification=centering

Figure: 集合$\mathbb{A}$ 和$\mathbb{B}$ 在u - t 方向投影的示意图
($\mathbb{A}$ 一般为有内点的凸集, $\mathbb{B}$ 是一条射线且不含点$(0,0,p^*)$)

因为$\mathbb{A}$ 和$\mathbb{B}$ 均为凸集, 由超平面分离定理, 存在$(\lambda,\nu,\mu)\neq 0$ 和α, 使得

$$\lambda^{\mathrm{T}}u+\nu^{\mathrm{T}}v+\mu t\geq\alpha,\quad\forall\,(u,v,t)\in\mathbb{A},$$
$$\lambda^{\mathrm{T}}u+\nu^{\mathrm{T}}v+\mu t\leq\alpha,\quad\forall\,(u,v,t)\in\mathbb{B}.$$

Slater约束品性与强对偶原理

- 我们断言$\lambda \geq 0$和$\mu \geq 0$ (否则可以取u_i和t为任意大的正实数以及$\nu = 0$, 这会导致$\lambda^{\mathrm{T}}u + \mu t$在集合$\mathbb{A}$上无下界).
- 同时, 由于$\mu t \leq \alpha$对于所有$t < p^*$成立, 可得$\mu p^* \leq \alpha$.
- 对任意$x \in \mathcal{D}$, 取$(u, v, t) = (c_i(x), Ax - b, f(x)) \in \mathbb{A}$, 可知

$$\sum_{i=1}^{m} \lambda_i c_i(x) + \nu^{\mathrm{T}}(Ax - b) + \mu f(x) \geq \alpha \geq \mu p^*.$$

- 假设$\mu > 0$, 则

$$L(x, \frac{\lambda}{\mu}, \frac{\nu}{\mu}) \geq p^*.$$

进一步地, 我们有$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) \geq p^*$, 根据弱对偶性$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) \leq p^*$自然成立.因此, 必有$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) = p^*$成立.说明在此情况下强对偶性满足, 且对偶最优解可以达到.

- 考虑$\mu=0$ 的情况, 可以从上面得到对于所有的$x\in\mathcal{D}$,

$$\sum_{i=1}^{m}\lambda_i c_i(x)+\nu^{\mathrm{T}}(Ax-b)\geq 0.$$

- 取满足Slater条件的点x_S, 有$\sum_{i=1}^{m}\lambda_i c_i(x_S)\geq 0$.
- 又$c_i(x_S)<0$ 和$\lambda_i\geq 0$, 我们得到$\lambda=0$, 上式化为

$$\nu^{\mathrm{T}}(Ax-b)\geq 0,\quad \forall\, x\in\mathcal{D}.$$

根据$(\lambda,\nu,\mu)\neq 0$ 可知$\nu\neq 0$, 结合A 行满秩可以得到$A^{\mathrm{T}}\nu\neq 0$.由于x_S 是可行解, 我们有$\nu^{\mathrm{T}}(Ax_S-b)=0$.

- 因为$x_S\in\mathcal{D}$, 则存在点e使得$\tilde{x}=x_S+e\in\mathcal{D}$,
且$\nu^{\mathrm{T}}Ae=\nu^{\mathrm{T}}(A\tilde{x}-b)<0$.这与$\nu^{\mathrm{T}}(Ax-b)\geq 0$, $\forall\, x\in\mathcal{D}$ 矛盾.
- 综上所述, Slater条件能保证强对偶性.
- 在定理的证明中, Slater条件保证了$\mu\neq 0$.

一阶充要条件

- 对于一般的约束优化问题，当问题满足特定约束品性时，我们知道KKT条件是局部最优解处的必要条件.
- 而对于凸优化问题，当Slater条件满足时，KKT条件则变为局部最优解的充要条件(根据凸性，局部最优解也是全局最优解).

定理 (凸优化问题的一阶充要条件)

对于凸优化问题，用a_i 表示矩阵A^{T} 的第i 列，$\partial f, \partial c_i$ 表示次梯度，如果*Slater*条件成立，那么x^*, λ^* 分别是原始，对偶全局最优解当且仅当

$$
\begin{aligned}
\text{稳定性条件} \quad & 0 \in \partial f(x^*) + \sum_{i\in\mathcal{I}} \lambda_i^* \partial c_i(x^*) + \sum_{i\in\mathcal{E}} \lambda_i^* a_i, \\
\text{原始可行性条件} \quad & Ax^* = b,\ \forall i \in \mathcal{E}, \\
\text{原始可行性条件} \quad & c_i(x^*) \le 0,\ \forall i \in \mathcal{I}, \\
\text{对偶可行性条件} \quad & \lambda_i^* \ge 0,\ \forall i \in \mathcal{I}, \\
\text{互补松弛条件} \quad & \lambda_i^* c_i(x^*) = 0,\ \forall i \in \mathcal{I}.
\end{aligned}
$$

一阶充要条件:充分性

- 设存在$(\bar{x}, \bar{\lambda})$ 满足KKT条件, 我们考虑凸优化问题的拉格朗日函数
$$L(x, \lambda) = f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} \lambda_i (a_i^{\mathrm{T}} x - b_i).$$
- 当固定$\lambda = \bar{\lambda}$ 时, 注意到$\bar{\lambda}_i \geq 0, i \in \mathcal{I}$ 以及$\bar{\lambda}_i (a_i^{\mathrm{T}} x), i \in \mathcal{E}$ 是线性函数可知$L(x, \bar{\lambda})$ 是关于x 的凸函数.
- 由凸函数全局最优点的一阶充要性可知, 此时$\bar{x}$ 就是$L(x, \bar{\lambda})$ 的全局极小点.根据拉格朗日对偶函数的定义,
$$L(\bar{x}, \bar{\lambda}) = \inf_{x \in \mathcal{D}} L(x, \bar{\lambda}) = g(\bar{\lambda}).$$
- 根据原始可行性条件$A\bar{x} = b$ 以及互补松弛条件$\bar{\lambda}_i c_i(\bar{x}) = 0, i \in \mathcal{I}$,
$$L(\bar{x}, \bar{\lambda}) = f(\bar{x}) + 0 + 0 = f(\bar{x}).$$
- 根据弱对偶原理,
$$L(\bar{x}, \bar{\lambda}) = f(\bar{x}) \geq p^* \geq d^* \geq g(\bar{\lambda}) = L(\bar{x}, \bar{\lambda}) \Rightarrow p^* = d^*,$$
故$\bar{x}, \bar{\lambda}$ 分别是原始问题和对偶问题的最优解.

关于充分性的评述

- 定理的充分性说明，若能直接求解出凸优化问题的KKT对，则其就是对应问题的最优解.
- 在充分性部分的证明中，我们没有使用Slater条件，这是因为在证明的一开始假设了KKT点是存在的.
- Slater条件的意义在于当问题最优解存在时，其相应KKT条件也会得到满足.
- 当Slater条件不满足时，即使原始问题存在全局极小值点，也可能不存在(x^*, λ^*) 满足KKT条件.
- 一阶充要条件：必要性的证明见P188-193.

- 弱化的Slater条件：若不等式约束条件的前k个约束函数是仿射，则Slater条件可以弱化为：
  存在$x \in \mathrm{relint}\ D$，满足

$$
\begin{aligned}
&f_i(x) \leq 0, i = 1, \cdots, k, \\
&f_i(x) < 0, i = k+1, \cdots, m, \\
&Ax = b
\end{aligned}
$$

- 原问题：

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

- 对偶问题：

$$\text{maximize} \quad g(\nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

- 具有强对偶性

Lagrange dual of QCQP

$$
\begin{array}{ll}
\text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\
\text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \le 0, i = 1, \cdots, m
\end{array}
$$

- 拉格朗日函数：

$$L(x, \lambda) = \frac{1}{2}x^T P(\lambda)x + q(\lambda)^T x + r(\lambda)$$

$$P(\lambda) = P_0 + \sum_{i=1}^{m} \lambda_i P_i \quad q(\lambda) = q_0 + \sum_{i=1}^{m} \lambda_i q_i \quad r(\lambda) = r_0 + \sum_{i=1}^{m} \lambda_i r_i$$

- 对偶函数：

$$g(\lambda) = \inf_x L(x, \lambda) = -\frac{1}{2}q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda)$$

Lagrange dual of QCQP

- 对偶问题：

$$
\begin{array}{ll}
\text{maximize} & -\dfrac{1}{2}q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda) \\
\text{subject to} & \lambda \succ= 0
\end{array}
$$

- Slater条件：存在x，满足

$$
(1/2)x^T P_i x + q_i^T x + r_i < 0, i = 1, \cdots, m
$$

Entropy maximization

- 原始问题：

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{n} x_i \log x_i, \quad D = R_+^n \\ \text{subject to} \quad & Ax \prec= b \\ & 1^T x = 1 \end{aligned}$$

- 对偶函数：

$$g(\lambda, \nu) = -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^T\lambda}$$

- 对偶问题：

$$\begin{aligned} \text{maximize} \quad & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-\alpha_i^T\lambda} \\ \text{subject to} \quad & \lambda \succ= 0 \end{aligned}$$

Entropy maximization 简化对偶问题

关于对偶变量ν解析求最大可以简化对偶问题(5.30)。对于任意固定λ，当目标函数对ν的导数为零时，即

$$\nu = \log \sum_{i=1}^{n} e^{-a_i^T \lambda} - 1.$$

目标函数取最大值。将ν的最优值代入对偶问题可以得到

$$\begin{aligned} \text{maximize} \quad & -b^T\lambda - \log\left(\sum_{i=1}^{n} e^{-a_i^T\lambda}\right) \\ \text{subject to} \quad & \lambda \succ= 0 \end{aligned}$$

这是一个非负约束的几何规划问题（凸优化问题）。弱化的Slater条件：存在$x \succ 0$，满足

$$Ax \prec= b$$

$$1^T x = 1$$

Minimum volume covering ellipsoid

- 原始问题：

$$\begin{array}{ll} \text{minimize} & \log\det X^{-1}, D = S^n_{++} \\ \text{subject to} & a_i^T X a_i \leq 1, i = 1, \ldots, m \end{array}$$

- 对偶函数：

$$g(\lambda) = \begin{cases} \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - 1^T\lambda + n & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty & otherwise \end{cases}$$

- 对偶问题：

$$\begin{array}{ll} \text{minimize} & \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - 1^T\lambda + n \\ \text{subject to} & \lambda \succ= 0 \end{array}$$

Minimum volume covering ellipsoid

- 弱化的**Slater**条件：存在$X \in S^n_{++}$，满足

$$a_i^T X a_i \leq 1, i = 1, \ldots, m$$

- 弱化的**Slater**条件总成立，因此该优化问题具有强对偶性。

次优解决认证和终止条件

- 对于一优化问题，若x为可行解，(λ, v)为对偶问题可行解，则有如下不等式：

$$f_0(x) - p^* \le f_0(x) - g(\lambda, v)$$

x为ε次优解，其中

$$\varepsilon = f_0(x) - g(\lambda, v)$$

- 不等式可以用于对次优解的精度估计。

互补松弛条件

设x^*为原始优化问题的最优解，(λ^*, v^*) 为对偶问题的最优解，若两者具有强对偶性，则

$$
\begin{aligned}
f_0(x^*) &= g(\lambda^*, v^*) \\
&= \inf_x \left(f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i v_i^* h_i(x) \right) \\
&\leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i v_i^* h_i(x^*) \\
&\leq f_0(x^*)
\end{aligned}
$$

所以 $\sum_i \lambda_i^* f_i(x^*) = 0$.

即 $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$.

KKT优化条件

设优化问题中，函数$f_0(x),\ldots,f_m(x), h_0(x),\ldots,h_p(x)$可微。设$x^*$为原始优化问题的最优解，$(\lambda^*,\nu^*)$为对偶问题的最优解，且两者具有强对偶性，则$(x^*,\lambda^*,\nu^*)$满足如下Karush-Kuhn-Tucker(KKT)条件：

- $f_i\left(x^*\right)\leq 0, i=1,\ldots,m.$
- $h_i\left(x^*\right)=0, i=1,\ldots,p.$
- $\lambda_i^*\geq 0, i=1,\ldots,m.$
- $\lambda_i^* f_i\left(x^*\right)=0, i=1,\ldots,m.$
- $\nabla f_0\left(x^*\right)+\sum_{i=1}^m \lambda_i^*\nabla f_i\left(x^*\right)+\sum_{i=1}^p \nu_i^*\nabla h_i\left(x^*\right)=0.$

凸优化问题的KKT条件

设原始问题为凸优化问题中，函数$f_0(x),\ldots,f_m(x), h_0(x),\ldots,h_p(x)$可微，设$(\tilde{x},\tilde{\lambda},\tilde{\nu})$满足**KKT**条件，则$\tilde{x}$为原始问题的最优解，而$(\tilde{\lambda},\tilde{\nu})$ 为对偶问题的最优解，且两者具有强对偶性。

- $f_i(\tilde{x}) \leq 0, i = 1,\ldots,m.$
- $h_i(\tilde{x}) = 0, i = 1,\ldots,p.$
- $\tilde{\lambda}_i \geq 0, i = 1,\ldots,m$
- $\tilde{\lambda}_i f_i(\tilde{x}) = 0, i = 1,\ldots,m$
- $\nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) = 0.$

例5.1 等式约束凸二次最小化。我们考虑这个问题：

$$\begin{array}{ll}\text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b,\end{array}$$

其中，$P \in \mathbf{S}_+^n$。这个问题的KKT条件是

$$Ax^\star = b, \quad Px^\star + q + A^T \nu^\star = 0,$$

我们可以写成

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \tag{33}$$

解决了$m+n$个变量的$m+n$个方程。优化问题等价为方程组求解问题。

注水问题（water-filling）

原始凸优化问题：

$$\begin{aligned} \text{minimize} \quad & -\sum_i \log(\alpha_i + x_i) \\ \text{subject to} \quad & x \succ= 0 \\ & 1^T x = 1 \end{aligned}$$

KKT条件：

- $x^* \succ= 0, 1^T x^* = 1, \lambda^* \succ= 0,$
- $\lambda_i^* x_i^* = 0, i = 1, \ldots, n,$
- $-1/(\alpha_i + x_i^*) - \lambda_i^* + v^* = 0, i = 1, \ldots, n$

解得：

$$x_i^* = \begin{cases} 1/v^* - \alpha_i & v^* < 1/\alpha_i \\ 0 & v^* \geq 1/\alpha_i \end{cases}$$

其中：

n

凸优化问题的对偶求解

设原始优化问题与对偶问题具有强对偶性，且(λ^*, ν^*) 为对偶问题的最优解，假设$L(x, \lambda^*, \nu^*)$存在唯一的最小解，即

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x)$$

存在唯一解x^*。若x^*为原问题的可行解，则x^*即为原始问题的最优解；若x^*不是原始问题的可行解，则原始问题不存在最优点，即原问题最优解无法达到。当对偶问题比原问题更容易求解时候，例如对偶问题可以解析求解或者有某些结构更加容易分析，上述方法有意义。

例 5.3 熵的最大化。考虑熵的最大化问题

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array}$$

其中定义域为 $\mathbf{R}^n_{++}$，其对偶问题为

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^n e^{-a_i^T\lambda} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

（参见第 213 页以及第 220 页。）假设 Slater 条件的弱化形式成立，即存在 $x \succ 0$ 使得 $Ax \preceq b$ 以及 $\mathbf{1}^T x = 1$，因此强对偶性成立，存在一个对偶最优解 $(\lambda^\star, \nu^\star)$。

设对偶问题已经解出。$(\lambda^\star, \nu^\star)$ 处的 Lagrange 函数为

$$L(x, \lambda^\star, \nu^\star) = \sum_{i=1}^n x_i \log x_i + \lambda^{\star T}(Ax - b) + \nu^\star(\mathbf{1}^T x - 1)$$

它在 $\mathcal{D}$ 上严格凸且有下界，因此有一个唯一解 $x^\star$，

$$x_i^\star = 1/\exp(a_i^T\lambda^\star + \nu^\star + 1), \quad i = 1, \cdots, n,$$

其中 a_i 是矩阵 A 的列向量。如果 $x^\star$ 是原问题可行解，则其必是原问题 (5.13) 的最优解。如果 $x^\star$ 不是原问题可行解，那么我们可以说原问题的最优解不能达到。

Example 5.4 *Minimizing a separable function subject to an equality constraint.* We consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n f_i(x_i) \\ \text{subject to} & a^T x = b, \end{array}$$

where $a \in \mathbf{R}^n$, $b \in \mathbf{R}$, and $f_i : \mathbf{R} \to \mathbf{R}$ are differentiable and strictly convex. The objective function is called *separable* since it is a sum of functions of the individual variables $x_1, \ldots, x_n$. We assume that the domain of f_0 intersects the constraint set, *i.e.*, there exists a point $x_0 \in \mathbf{dom}\, f_0$ with $a^T x_0 = b$. This implies the problem has a unique optimal point $x^\star$.

The Lagrangian is

$$L(x,\nu) = \sum_{i=1}^n f_i(x_i) + \nu(a^T x - b) = -b\nu + \sum_{i=1}^n (f_i(x_i) + \nu a_i x_i),$$

which is also separable, so the dual function is

$$\begin{aligned} g(\nu) &= -b\nu + \inf_x \left(\sum_{i=1}^n (f_i(x_i) + \nu a_i x_i) \right) \\ &= -b\nu + \sum_{i=1}^n \inf_{x_i} (f_i(x_i) + \nu a_i x_i) \\ &= -b\nu - \sum_{i=1}^n f_i^*(-\nu a_i). \end{aligned}$$

The dual problem is thus

$$\text{maximize} \quad -b\nu - \sum_{i=1}^n f_i^*(-\nu a_i),$$

with (scalar) variable $\nu \in \mathbf{R}$.

Now suppose we have found an optimal dual variable $\nu^\star$. (There are several simple methods for solving a convex problem with one scalar variable, such as the bisection method.) Since each f_i is strictly convex, the function $L(x, \nu^\star)$ is strictly convex in x, and so has a unique minimizer $\tilde{x}$. But we also know that $x^\star$ minimizes $L(x, \nu^\star)$, so we must have $\tilde{x} = x^\star$. We can recover $x^\star$ from $\nabla_x L(x, \nu^\star) = 0$, *i.e.*, by solving the equations $f_i'(x_i^\star) = -\nu^\star a_i$.

扰动及灵敏度分析

- 扰动问题:

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \le u_i, & i = 1, \ldots, m \\ & h_j(x) = v_j, & j = 1, \ldots, p \end{array}$$

- 当$u = 0, v = 0$ 时即为原始问题。
- 若u_i 为正, 则第i 个不等式约束被放宽; 若u_i 为负, 则第i 个不等式约束被收紧。
- 记$p^*(u, v)$ 为扰动问题的最优解。若扰动问题无最优解, 则记

$$p^*(u, v) = \infty$$

全局不等式和局部灵敏度分析

- 设对偶问题存在最优解，且与原始问题具有强对偶性，若非扰动问题的最优对偶解为(λ^*, ν^*)，则有：

$$p^*(u,v) \geq p^*(0,0) - {\lambda^*}^T u - {\nu^*}^T v$$

- 当若$p^*(u,v)$ 在$u=0, v=0$ 处可微，则

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i} \quad \nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

全局不等式证明

定理

对于所有的u 和v，我们有

$$p^\star(u,v) \geq p^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v.$$

证明：假设x 扰动问题的任一可行解, 即, $f_i(x) \leq u_i$ 对$i = 1,\ldots,m$成立, 同时$h_i(x) = v_i$ 对$i = 1,\ldots,p$成立. 利用强对偶性有,

$$\begin{aligned} p^\star(0,0) = g\left(\lambda^\star, \nu^\star\right) &\leq f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x) \\ &\leq f_0(x) + \lambda^{\star T} u + \nu^{\star T} v. \end{aligned}$$

(第一个不等式利用$g\left(\lambda^\star, \nu^\star\right)$的定义得到; 第二个不等式由$\lambda^\star \succeq 0$得到。)因此，我们有对于扰动问题的任意可行解$x$, 有下式成立

$$f_0(x) \geq p^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v,$$

从而全局不等式$p^\star(u,v) \geq p^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v$成立.

灵敏度解释：$p^*(u,v) \geq p^*(0,0) - \lambda^{*T}u - v^{*T}v$

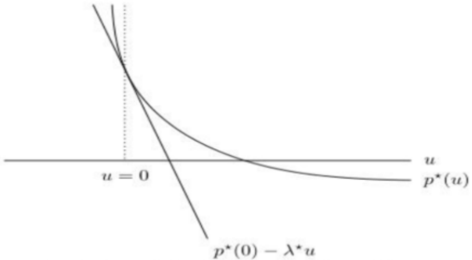


图5.10 具有一个约束$f_1(x) \leq u$ 的凸问题的最优值$p^\star(u)$ 的图像，它是u 的函数。当$u=0$时，对应原始未被扰动的问题；当$u<0$ 时，约束加强了，当$u>0$ 时，约束放松了。仿射函数$p^\star(0)-\lambda^\star u$ 是$p^\star$ 的一个下界。

- 如果$\lambda_i^\star$ 比较大，我们加强第i 个约束(即选择$u_i<0$)，则最优值$p^\star(u,v)$ 必会大幅增加。
- 如果$\nu_i^\star$ 较大且大于零，我们选择$v_i<0$，或者如果$\nu_i^\star$ 较大且小于零，我们选择$v_i>0$，在这两种情况下最优值$p^\star(u,v)$ 必会大幅增加。
- 如果$\lambda_i^\star$ 较小，我们放松第i 个约束($u_i>0$)，那么最优值$p^\star(u,v)$ 不会减小太多。
- 如果$\nu_i^\star$ 较小且大于零，$v_i>0$ 或者如果$\nu_i^\star$ 较小且小于零，$v_i<0$，那么最优值$p^\star(u,v)$ 不会减小太多。

局部灵敏度分析

局部灵敏度: 假设$p^\star(u,v)$ 在$(0,0)$ 处是可微的, 然后

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \quad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

利用全局不等式结果证明(对于$\lambda_i^\star$)

$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t\searrow 0} \frac{p^\star(te_i,0) - p^\star(0,0)}{t} \geq -\lambda_i^\star$$
$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t>0} \frac{p^\star(te_i,0) - p^\star(0,0)}{t} \leq -\lambda_i^\star$$

因此$p^\star(u)$对于一个不等式的问题也有同样约束约束

广义不等式

本节将Lagrange 对偶理论扩展到具有广义不等式约束的问题，即广义不等式

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subjectto} & f_i(x) \preceq_{K_i} 0, \quad i = 1, \cdots, m \\ & h_i(x) = 0, \quad i = 1, \cdots, p, \end{array}$$

$\preceq_{K_i}$ 是$\mathbf{R}^{k_i}$上的广义不等式对于问题(5.91) 中的每个广义不等式$f_i(x) \preceq_{K_i} 0$, 引入Lagrange 乘子向量$\lambda_i \in \mathbf{R}^{k_i}$ 并定义相关的Lagrange 函数如下

$$L(x, \lambda, \nu) = f_0(x) + \lambda_1^T f_1(x) + \cdots + \lambda_m^T f_m(x) + \nu_1 h_1(x) + \cdots + \nu_p h_p(x),$$

其中$\lambda = (\lambda_1, \cdots, \lambda_m)$, $\nu = (\nu_1, \cdots, \nu_p)$ 。对偶函数的定义和原问题只有数值不等式的情形一样,对偶函数$g : \mathbf{R}^{k_1} \times \cdots \times \mathbf{R}^{k_m} \times \mathbf{R}^p \to \mathbf{R}$, 被定义为

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i^T f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$$

下界性质：如果$\lambda_i \succeq_{K_i^*} 0$，同时$g(\lambda_1, \ldots, \lambda_m, \nu) \leq p^\star$ 证明：如果$\tilde{x}$是可行点，同时$\lambda \succeq_{K_i^*} 0$，有

$$\begin{aligned} f_0(\tilde{x}) &\geq f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \\ &\geq \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu) \\ &= g(\lambda_1, \ldots, \lambda_m, \nu) \end{aligned}$$

最小化所有可行点$\tilde{x}$可以得到$p^\star \geq g(\lambda_1, \ldots, \lambda_m, \nu)$

Lagrange对偶优化问题为：

$$\begin{array}{ll} \text{maximize} & g(\lambda_1, \ldots, \lambda_m, \nu) \\ \text{subjectto} & \lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m \end{array}$$

弱对偶性总是成立，即：$p^\star \geq d^\star$ 强对偶性在广义不等式的情况下同样成立：$p^\star = d^\star$，前提是原问题是凸的且满足合适的约束准则

广义不等式

半定规划

primal SDP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subjectto} & x_1 F_1 + \cdots + x_n F_n + G \preceq 0 \end{array}$$

- 拉格朗日乘数是矩阵$Z \in \mathbf{S}^k$
- 拉格朗日$L(x, Z) = c^T x + \operatorname{tr}\left(Z\left(x_1 F_1 + \cdots + x_n F_n - G\right)\right)$
- 对偶函数

$$g(Z) = \inf_x L(x, Z) = \begin{cases} -\operatorname{tr}(GZ) & \operatorname{tr}\left(F_i Z\right) + c_i = 0, \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

对偶SDP(半定锥自对偶)

$$\begin{array}{ll} \text{maximize} & -\operatorname{tr}(GZ) \\ \text{subjectto} & Z \succeq 0, \quad \operatorname{tr}\left(F_i Z\right) + c_i = 0, \quad i = 1, \ldots, n \end{array}$$

如果primal SDP 是严格可行的($\exists x$ with $x_1 F_1 + \cdots + x_n F_n \prec G$)，$p^\star = d^\star$

提纲

七章
无约束优化

无约束优化问题

- 问题描述: minimize $f(x)$
- $f(x)$ 为凸函数, 且二次可微。
- 无约束问题求解的两种方法:
- 求解梯度方程:

$$\nabla f\left(x^{*}\right)=0$$

- 迭代逼近:

$$f\left(x^{(k)}\right) \rightarrow p^{*}$$

例

- 二次优化:

$$\text{minimize } \frac{1}{2}x^T Px + q^T x + r, P \in S_+^n$$

梯度方程

$$Px^* + q = 0$$

迭代起始点

- 起始点$x^{(0)}$满足：
  1.$x^{(0)} \in \mathrm{dom}\ f$;
  2.$S = \{x \in \mathrm{dom}\ f \mid f(x) \leq f(x^{(0)})\}$　为闭集。
- 满足条件2的几种函数：
  - 函数$f(x)$任意下水平集都是闭集；
  - 函数的定义域为R^n
  - 当$x \to$ bd dom f时，$f(x) \to \infty$

定义

定义 函数$f(x)$在S上具有强凸性，若$f(x)$满足

$$\nabla^2 f(x) \succ= mI, m > 0$$

- 若函数$f(x)$具有强凸性，则有

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{m}{2} \|y - x\|_2^2$$
$$\geq f(x) - \frac{1}{2m} \|\nabla f(x)\|_2^2$$

- p^*为最优值，则

$$f(x) - p^* \leq \frac{1}{2m} \|\nabla f(x)\|_2^2$$

强凸性

- 若函数$f(x)$在S上具有强凸性，则可以证明存在$M>0$，满足
$$\nabla^2 f(x) \prec= MI$$
则有
$$f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{M}{2} \|y-x\|_2^2$$

- p^*为最优值，则
$$p^* \leq f(x) - \frac{1}{2M} \|\nabla f(x)\|_2^2$$

强凸性

- 对于$x \in S$，矩阵$\nabla^2 f(x) \in S^n_{++}$的特征值从大到小依次为$\{\lambda_1, ..., \lambda_n\}$。则有：

$$mI \prec= \lambda_n I \prec= \nabla^2 f(x) \prec= \lambda_1 I \prec= MI$$

定义

定义 矩阵$\nabla^2 f(x) \in S^n_{++}$的条件数为最大特征值与最小特征值之比，即$r = \lambda_1 / \lambda_n$。

- 条件数的上界：

$$r \le M/m$$

下降法

- 下降法的基本原理：
  迭代$x^{(\dot{\kappa}+1)} = x^{(\dot{\kappa})} + t\Delta x$，满足$f(x^{(\dot{\kappa}+1)}) < f(x^{(\dot{\kappa})})$
  Δx为下降方向，t为步长因子。

- 对于凸函数$f(x)$，当Δx满足$\nabla f(x)^T \Delta x < 0$时，存在某个$t$，使得$f(x^{(\dot{\kappa}+1)}) < f(x^{(\dot{\kappa})})$。

下降方法

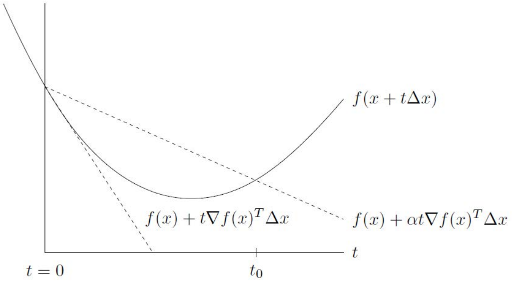
通用下降方法：

- 给出初始点$x \in \mathbf{dom}\, f$；
- 循环迭代
  - 计算下降方向Δx；
  - 搜索步长因子t；
  - 迭代：$x = x + t\Delta x$

- 精确直线搜索：$t = \arg\min_{t>0} f(x + t\Delta x)$

- 回溯直线搜索：
  给定参数$\alpha \in (0, 0.5), \beta \in (0, 1)$
  - 初始化：令$t = 1$；
  - 循环迭代
    - 若$f(x + t\Delta x) \le f(x) + \alpha t \nabla f(x)^T \Delta x$，则终止退出；
    - 否则令$t = \beta t$

步长因子搜索


$f(x+t\Delta x)$
$f(x)+t\nabla f(x)^T\Delta x$
$f(x)+\alpha t\nabla f(x)^T\Delta x$
t
$t=0$
t_0

- 下降方向：$\Delta x = -\nabla f(x)$
- 终止条件：$\|\nabla f(x)\|_2 \le \varepsilon$
- 收敛性：
$$f(x^{(k)}) - p^* \le c^{\dot{k}}(f(x^{(0)}) - p^*)$$
其中$c \in (0, 1)$。=,其中常数c依赖于采用的搜索准则。采用精确直线搜索时，$c = 1 - m/M < 1$；采用回溯直线搜索时，$c = 1 - min\{2m\alpha, 2\beta * \alpha * m/M1$
- 算法简单，当m/M收敛速度较慢。

收敛性分析

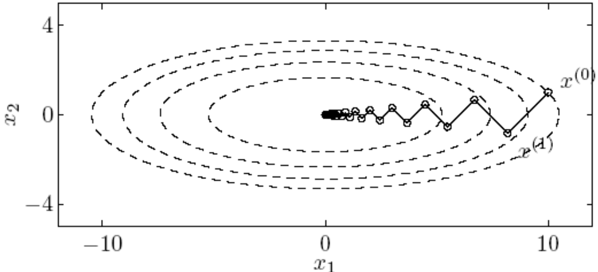
- 设函数$f(x)$具有强凸性，则存在$m>0$和$M>0$，满足：
$$mI \preceq= \nabla^2 f(x) \preceq= MI$$
则有：
$$f(x-t\Delta x) \le f(x) - t\left\|\nabla f(x)\right\|_2^2 + \frac{Mt^2}{2}\left\|\nabla f(x)\right\|_2^2，其中\Delta x = -\nabla f(x)$$
- 若t采用精确一维搜索，则$t=1/M$，收敛速度因子：
$$c = 1 - m/M$$
- 若t采用回溯一维搜索，收敛速度因子：
$$c = 1 - \min\{2m\alpha, 2\beta\alpha m/M\}$$
- 条件数越大，收敛速度越小。

例子

R^2 空间的二次问题

$$\text{minimize}\ \frac{1}{2}\left(x_1^2+\gamma x_2^2\right),\gamma>0$$

- 初始解为$(\gamma,1)$, 采用精确一维搜索;
- 迭代:
$$x_1^{(k)}=\gamma\left(\frac{\gamma-1}{\gamma+1}\right)^k \quad x_2^{(k)}=\left(-\frac{\gamma-1}{\gamma+1}\right)^k$$
- 当$\gamma \ll 1$ or $\gamma \gg 1$ 时,算法收敛速度很慢。

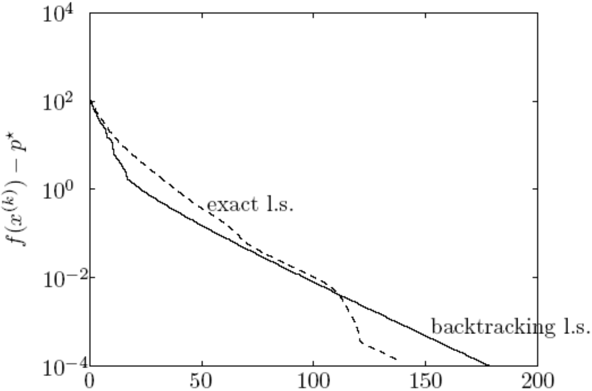


例子

R^100空间的一个问题

$$\text{minimize} \quad c^T x - \sum_{i=1}^{m} \log\left(b_i - a_i^T x\right), ypm = 500, pn = 100$$

- 采用回溯直线搜索和精确直线搜索的梯度优化方法产生误差和迭代次数的关系



梯度下降法

结论

- 梯度方法通常呈现近似线性收敛性质，即误差$f(x^(k)) - p^*$以类似几何数列方式收敛到零。
- 回溯参数α和的取值对收敛性有明显影响，但不会产生戏剧性的效果。精确直线搜索有时可以改善梯度方法的收敛性，但是效果不是很大（或许不能抵消进行精确直线搜索增加的计算量）
- 收敛速度强烈依赖于Hessian矩阵的条件数，问题条件数不能太大（比如等于100），收敛速度也可能很慢。
- 算法优点是方法简单，确定是收敛速度强烈依赖于Hessian矩阵的条件数。

最速下降法

- 规范化最速下降方向:

$$\Delta x_{nsd} = \arg\min_{v} \left\{ \nabla f(x)^T v \,\middle|\, \|v\| = 1 \right\}$$

- 非规范化最速下降方向

$$\Delta x_{sd} = \|\nabla f(x)\|_* \Delta x_{nsd}$$

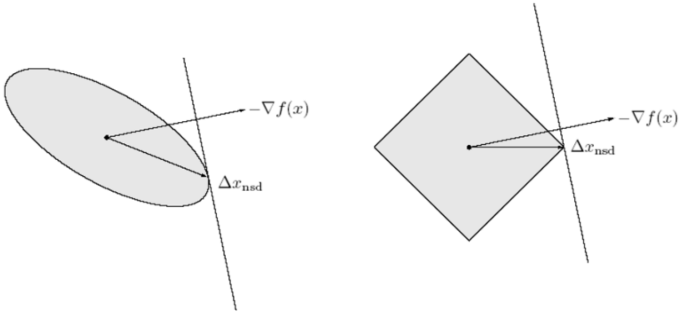
- 欧式范数: $\Delta x_{sd} = -\nabla f(x)$
- 二次范数$\|x\|_P = \left(x^T P x\right)^{1/2}$:

$$\Delta x_{sd} = -P^{-1} \nabla f(x)$$

- l_1-范数:

$$\Delta x_{sd} = -\frac{\partial f(x)}{\partial x_i} e_i$$

牛顿法



最速下降法

- 设函数$f(x)$ 二阶可微，则在x 附近，$f(x)$ 的泰勒展式为：

$$\widehat{f}(x+\Delta x)=f(x)+\nabla f(x)^T\Delta x+\frac{1}{2}\Delta x^T\nabla^2 f(x)\Delta x$$

- 泰勒展开可作为$f(x)$ 在x 附近的近似；
- 下降方向（Newton步径）：

$$\Delta x_{nt}=-\nabla^2 f(x)^{-1}\nabla f(x)$$

- 为二次范数$\|x\|_{\nabla^2 f(x)}=\left(x^T\nabla^2 f(x)x\right)^{1/2}$ 上的最速下降方向。

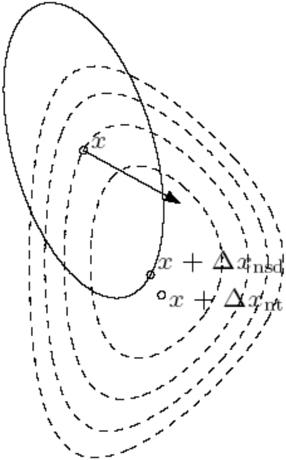


图9.17 虚线是一个凸函数的等值线。椭圆(实线) 表示$\{x+v \mid v^T\nabla^2 f(x)v \leqslant 1\}$ 。箭头表示梯度下降方向$-\nabla f(x)$ 。Newton 步径Δx_{nt} 是范数$\|\cdot\|_{\nabla^2 f(x)}$ 导出的最速下降方向。图中也给出了采用相同范数的规范化的最速下降方向Δx_{nsd} 。

牛顿减量法

- 令：$\lambda(x) = \left(\nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)\right)^{1/2}$
- $\lambda(x)$ 为$f(x)$ 在x 处的牛顿减量。
- 牛顿减量的性质**1**：

$$f(x) - \inf_y \hat{f}(y) = f(x) - \widehat{f}\,(x + \Delta x_{nt}) = \frac{1}{2}\lambda(x)^2$$

  牛顿减量可作为迭代求解的误差估计。
- 性质2:

$$\nabla f(x)^T \Delta x_{nt} = -\lambda(x)^2$$

- 性质3：牛顿减量具有仿射不变性。

牛顿方法

- 初始化: 给定初始解$x \in \mathrm{dom} f$ 以及$\varepsilon > 0$
- LOOP:
- 计算Newton步径和减量: $\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$

$$\lambda^2 = \nabla f(x)^T \nabla^2 f(x)^{-1} \nabla f(x)$$

- 若$\lambda^2/2 < \varepsilon$ 则终止退出;
- 直线搜索: 通过回溯搜索确定步长t;
- 改进: $x = x + t\Delta x_{nt}$

定理

1. 定理：假设$f(x)$二阶连续可微，具有强凸性，即存在$M > m > 0$，满足：

$$mI \prec= \nabla^2 f(x) \prec= MI$$

且*Hessian*矩阵满足*Lipschitz*条件，即存在$L > 0$,满足：
$\left\|\nabla^2 f(x) - \nabla^2 f(y)\right\|_2 \le L\|x-y\|_2$
则存在$0 < \eta < m^2/L, \gamma > 0$，

$$\text{若} \left\|\nabla f\left(x^{(k)}\right)\right\|_2 \ge \eta,\ \text{则} f\left(x^{(k+1)}\right) - f\left(x^{(k)}\right) < -\gamma;$$

$$\text{若} \left\|\nabla f\left(x^{(k)}\right)\right\|_2 < \eta,\ \text{则} t^{(k)} = 1 \text{，且}$$

$$\frac{L^2}{2m^2}\left\|\nabla f\left(x^{(k+1)}\right)\right\|_2 \le \left(\frac{L}{2m^2}\left\|\nabla f\left(x^{(k)}\right)\right\|_2\right)^2$$

定理

1. 定理: 假设$f(x)$ 二阶连续可微, 具有强凸性, 即存在$M > m > 0$, 满足:

$$mI \preceq= \nabla^2 f(x) \preceq= MI$$

且海森矩阵满足Lipschitz条件, 即存在$L > 0$, 满足:

$$\left\| \nabla^2 f(x) - \nabla^2 f(y) \right\|_2 \le L\|x - y\|_2$$

则存在$k > 0$, 对于$l > k$, 有

$$f\left(x^{(l)}\right) - p^* \le \frac{1}{2m} \left\| \nabla f\left(x^{(l)}\right) \right\|_2^2 \le \frac{2m^3}{L^2} \left(\frac{1}{2}\right)^{2^{l-k+1}}$$

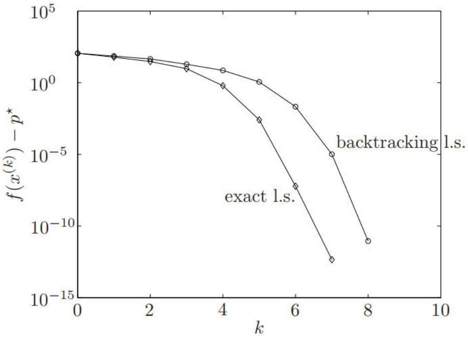


图9.21 用Newton 方法求解$\mathbf{R}^{100}$ 中一个问题的误差和迭代次数k 的关系。回溯直线搜索参数取为$\alpha = 0.01, \beta = 0.5$ 。收敛速度同样极其迅速：经过7 次或8 次迭代就获得了很高的精度。采用精确直线搜索的Newton 方法的迭代次数只比回溯直线搜索的少一次。

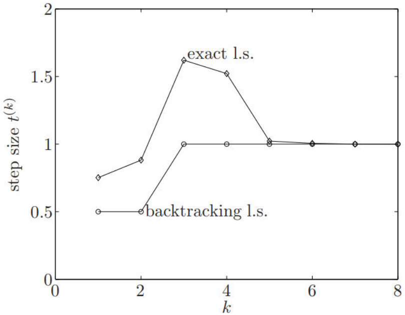


图9.22 对$\mathbf{R}^{100}$ 中的一个问题，采用回溯直线搜索和精确直线搜索的Newton 方法的步长t和迭代次数的关系。回溯直线搜索在最初两次迭代中采用回溯后的步长，以后总是采用$t = 1$ 。

凸优化理论与应用

第10章等式约束优化

等式约束优化问题

- 问题描述：

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

- $f(x)$ 为凸函数，且二次连续可微，且$A \in R^{p\times n}, p < n, \operatorname{rank} A = p$
- 假设最优值p^* 存在，则x^* 为最优解当且仅当存在v^*,满足（KKT条件）：

$$\nabla f(x^*) + A^T v^* = 0, Ax^* = b$$

例

- 二次优化：

$$\begin{array}{ll}\text{minimize} & \frac{1}{2}x^TPx+q^Tx+r, P\in S_+^n \\ \text{subject to} & Ax=b\end{array}$$

KKT系统：

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} x^* \\ v^* \end{bmatrix}=\begin{bmatrix} -q \\ b \end{bmatrix}$$

- KKT系统可解，则二次优化问题存在最优解。
- 系数矩阵称为KKT矩阵。KKT矩阵非奇异当且仅当：

$$Ax=0, x\neq 0 \Rightarrow x^TPx>0$$

消去等式约束

- 方程组$Ax=b$ 的解集:
$$\{x \mid Ax=b\}=\{Fz+\tilde{x} \mid z\in R^{n-p}\}$$
- $\tilde{x}$ 为方程组的一个特解, F 为A 的零空间的任何矩阵。
- 无约束优化形式：
$$\text{minimize} \quad f(Fz+\tilde{x}), z\in R^{n-p}$$
- 若z^* 为最优解, 则有
$$x^*=Fz^*+\tilde{x}$$
最优对偶变量:
$$v^*=-\left(A^TA\right)^{-1}A\nabla f\left(x^*\right)$$

对偶方法求等式约束问题

对偶函数

$$\begin{aligned} g(\nu) \quad &= \quad -b^T\nu + \inf_x (f(x) + \nu^T Ax) \\ &= -b^T\nu - \sup_x \left((-A^T\nu)^T x - f(x)\right) \\ &= -b^T\nu - f^*(-A^T\nu) \end{aligned}$$

- 对偶形式

$$\text{maximize} \quad -b^T v - f^* \left(A^T v\right)$$

牛顿法

- x 为等式约束优化的可行解，则在x附近原问题的二次近似为：

$$\begin{array}{ll} \text{minimize} & \widehat{f}(x+v) = f(x) + \nabla f(x)^T v + \frac{1}{2} v^T \nabla^2 f(x) v \\ \text{subject to} & A(x+v) = b \end{array}$$

- 设Δx_{nt} 为Newton方法和ω 分别为该问题和对偶问题的最优解, 则满足:

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ \omega \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}$$

- 牛顿减量

$$\lambda(x) = \left(\Delta x_{nt}^T \nabla^2 f(x)^{-1} \Delta x_{nt}\right)^{1/2}$$

牛顿减量

- 牛顿减量

$$\lambda(x) = \left(\Delta x_{nt}^T \nabla^2 f(x)^{-1} \Delta x_{nt}\right)^{1/2}$$

- 牛顿减量的性质：

$$f(x) - \inf\{\widehat{f}(x+v) \mid A(x+v) = b\} = \frac{1}{2}\lambda(x)^2$$

- 性质2：牛顿减量具有仿射不变性。

- 可行下降方向: 设x 满足方程组$Ax=b$ 。若v 满足方程组$Av=0$, 则$A(x+tv)=b$ 称为可行方向。若对于较小的$t>0$, 有$f(x+tv)<f(x)$, 则v 为可行下降方向。

1. 初始化：给定初始解$x \in \mathrm{dom}\ f$满足$Ax = b$，以及$\varepsilon > 0$
2. **LOOP**:
   - 计算Δx_{nt}及λ^2；
   - 若$\lambda^2/2 < \varepsilon$，则终止退出；
   - 一维线性搜索：计算步长因子t。
   - 迭代：$x = x + t\Delta x_{nt}$

消去等式约束的牛顿方法

问题：

$$\text{minimize} \quad \tilde{f}(z) = f(Fz + \tilde{x}), z \in R^{n-p}$$

1. 初始化：$z^{(0)}$，第k次迭代值$z^{(k)}$；
2. 转换为等式约束下的牛顿方法：
   - 初始值：$x^{(0)} = Fz^{(0)} + \tilde{x}$
   - 迭代值：$x^{(k)} = Fz^{(k)} + \tilde{x}$

非可行解为初始点的牛顿法

- x为等式约束优化的非可行解，则增量Δx应尽可能满足**KKT**条件，即: $x + \Delta x \approx x^*$，则最优性条件为

$$A(x + \Delta x) = b \quad \nabla f(x + \Delta x) + A^T \omega = 0$$

- 函数$f(x)$二阶连续可微，因此有

$$\nabla f(x + \Delta x) \approx \nabla f(x) + \nabla^2 f(x) \Delta x$$

- 设Δx_{nt}和ω为**KKT**条件的解，即有：

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{nt} \\ \omega \end{bmatrix} = - \begin{bmatrix} \nabla f(x) \\ Ax - b \end{bmatrix}$$

非可行解为初始点的牛顿法

- 令$y=(x,v)$，$r(y)=\big(\nabla f(x)+A^Tv, Ax-b\big)$
- 则**KKT**可表示为：$r(y)=0$
- 设y为不满足**KKT**条件，则迭代量需满足：

$$r(y+\Delta y)\approx r(y)+\mathrm{Dr}(y)\Delta y=0$$

即

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix}\begin{bmatrix} \Delta x_{nt} \\ \Delta v_{nt} \end{bmatrix}=-\begin{bmatrix} \nabla f(x)+A^Tv \\ Ax-b \end{bmatrix}$$

非可行解为初始点的牛顿法

1. 初始化：给定初始解$x \in \mathrm{dom}\ f$，v，以及$\varepsilon > 0$
2. **LOOP**:
   - 计算Δx_{nt}及Δv_{nt}；
   - 回溯一维线性搜索：
     - 令$t = 1$；
     - **While** $\|r\left(x + t\Delta x_{nt}, v + t\Delta v_{nt}\right)\|_2 \geq (1 - \alpha t)\|r(x, v)\|_2$

$$t = \beta t$$

   - 迭代：$x = x + t\Delta x_{nt} \quad v = v + t\Delta v_{nt}$
   - 当$Ax = b$且$\|r(y)\|_2 < \varepsilon$时，终止迭代。

KKT系统的求解

- **KKT**系统：

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g \\ h \end{bmatrix}$$

- LDL^T分解；
- 若H非奇异，则可消元：

$$AH^{-1}A^Tw = h - AH^{-1}g, Hv = -\left(g + A^Tw\right)$$

- 若H奇异，则**KKT**系统可改写为：

$$\begin{bmatrix} H + A^TQA & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} v \\ w \end{bmatrix} = - \begin{bmatrix} g + A^TQh \\ h \end{bmatrix}$$

其中$Q \succ= 0$，且满足$H + A^TQA \succ 0$

第9章　内点法

不等式约束优化问题

- 问题描述：

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \le 0, i = 1, \dots, m \\ & Ax = b \end{aligned}$$

- $f_i(x)$为凸函数，且二次连续可微，且：
- 若H非奇异，则可消元：

$$A \in R^{p \times n}, p < n, \operatorname{rank} A = p$$

- 假设最优值p^*存在；
- 假设存在$\tilde{x} \in \operatorname{dom} f$，满足严格不等式条件$f_i(x) < 0$
- 则优化问题具有强对偶性，其对偶问题亦可解。

- 示性函数消去不等式约束：：

$$\begin{array}{ll} \text{minimize} & f_0(x) + \sum_{i=1}^{m} I_-\left(f_i(x)\right) \\ \text{subject to} & Ax = b \end{array}$$

- $I_-(u)$不具备良好的连续可微性，考虑用对数障碍函数来近似替代。

$$I_-(u) = \begin{cases} 0 & u \leq 0 \\ \infty & u > 0 \end{cases}$$

对数障碍函数

- 对于$t > 0$, $-1/t\log(-u)$是$I_-(u)$的光滑逼近。且当$t \to \infty$时，有：

$$-1/t\log(-u) \to I_-(u)$$

- 令

$$\phi(x) = -\sum_{i=1}^{m} \log\left(-f_i(x)\right)$$

- 带示性函数的优化问题可近似为：

$$\begin{array}{ll} \text{minimize} & f_0(x) + \frac{1}{t}\phi(x), t > 0 \\ \text{subject to} & Ax = b \end{array}$$

对数障碍函数

- 对数障碍函数$\phi(x)$是凸函数
- 对数障碍函数二阶连续可微，其导数为

$$\nabla\phi(x) = \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x)$$

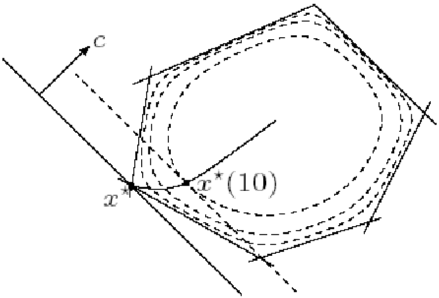
$$\nabla^2\phi(x) = \sum_{i=1}^{m} \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla^2 f_i(x)$$

中心路径

- 对数障碍近似问题的等价问题：

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x), t > 0 \\ \text{subject to} & Ax = b \end{array}$$

- 最优解为$x^*(t)$，则最优解集称为优化问题的中心路径。



中心路径的对偶点

- 设$x = x^*(t)$，则存在w满足**KKT**条件：

$$t\nabla f_0(x) + \sum_{i=1}^{m} \frac{1}{-f_i(x)} \nabla f_i(x) + A^T w = 0, Ax = b$$

- 令$\lambda_i^*(t) = -\frac{1}{tf_i(x^*(t))}, v^*(t) = w/t$
  则$x = x^*(t)$是拉格朗日函数$L(x, \lambda^*(t), v^*(t))$的最小值解。

$$L(x, \lambda^*(t), v^*(t)) = f_0(x) + \sum_{i=1} \lambda_i^*(t) f_i(x) + v^*(t)(Ax - b)$$

- $(\lambda^*(t), v^*(t))$为对偶问题的可行解。

中心线的对偶点

- 设p^*为原始问题的最优解，则有：

$$
\begin{aligned}
p^* &\geq g\left(\lambda^*(t), v^*(t)\right) \\
&= L\left(x^*(t), \lambda^*(t), v^*(t)\right) \\
&= f_0\left(x^*(t)\right) - m/t
\end{aligned}
$$

- 因此，当$t \to \infty$时，有$f_0\left(x^*(t)\right) \to p^*$。$x^*(t)$为原始问题的$m/t-$次优解。

障碍方法

- 初始化: 给定严格可行解$x, t > 0, \mu > 1$, 及$\varepsilon > 0$
- LOOP:
- 中心步骤: 以x 为初始点求解优化问题$x^*(t)$,

$$\begin{array}{ll} \text{minimize} & tf_0(x) + \phi(x) \\ \text{subjectto} & Ax = b \end{array}$$

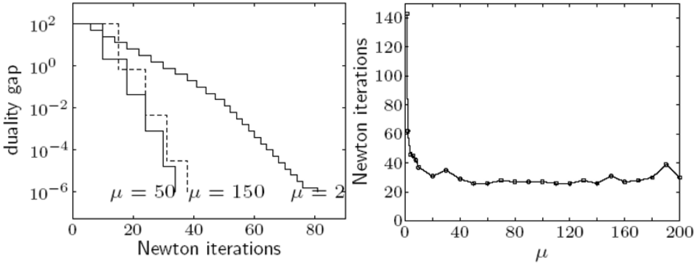
- 迭代: $x = x^*(t)$
- 终止条件: 若$m/t < \varepsilon$, 则终止退出。
- 更新t : $t = \mu t$

收敛性分析

- 外层循环迭代次数:
$$\left\lceil \frac{\log\left(m/\left(\varepsilon t^{(0)}\right)\right)}{\log \mu} \right\rceil$$
- 中心步骤实质为一个无约束或等式约束优化问题, 其收敛性分析与相应优化问题的收敛性分析结果一致。

例：

- LP问题: $m = 100, n = 50$
- 初始值: $t^{(0)} = 1, \varepsilon = 10^{-6}$



第一阶段方法

- 对于不等式约束的优化问题，如何寻找严格可行解或验证不可解？
- 求解优化问题：

$$\begin{array}{ll} \text{minimize} & s \\ \text{subjectto} & f_i(x) \leq s, i = 1, \ldots, m \\ & Ax = b \end{array}$$

- 该问题最优解存在，假设最优值为$\bar{p}^*$
- 当$\bar{p}^* < 0$ 时，存在严格可行解；
- 当$\bar{p}^* > 0$ 时，原始问题不可解；
- 当$\bar{p}^* = 0$ 时，无法准确确定。

第一阶段方法

- 优化目标为逐项之和：

$$\begin{array}{ll} \text{minimize} & 1^T s \\ \text{subject to} & f_i(x) \le s_i, i = 1, \ldots, m, Ax = b \end{array}$$

- 对于固定的x, $s_i = \max\{f_i(x), 0\}$

寻找严格可行解的方法

- 牛顿法求解优化问题：

$$\begin{array}{ll} \text{minimize} & S \\ \text{subject to} & f_i(x) \leq s, i = 1, \ldots, m, Ax = b \end{array}$$

- 迭代终止条件: 当前解$s^{(k)} < 0$, 即终止迭代, 严格可行解为$x^{(k)}$ 。

凸优化理论与应用

复习

凸集

- 凸集与凸包的概念；
- 几种常见的凸集
  多面体、单纯形、范数球、凸锥、真锥、范数锥、半正定锥等；
- 保持凸集的运算
  集合交运算、仿射变换、透视函数、线性分式函数
- 广义不等式的概念
- 支撑超平面、分割超平面的概念

凸函数

- 凸函数的概念及判定；
- 常见凸函数；
- Jessen不等式；
- 保持函数凸性的算子；
- 共轭函数的概念、常见函数的共轭函数；
- 向量凸函数（广义不等式下的凸性）

优化问题

- 优化问题的基本描述、等价形式；
- 凸优化问题的基本描述、等价形式；
- 常见凸优化问题。

对偶问题

- 优化问题的拉格朗日函数及对偶函数的概念和性质；
- 一些常见问题的对偶函数；
- 对偶函数与共轭函数的关系；
- 对偶问题的描述；
- 强对偶性的概念及Slater条件；
- KKT条件；

常用算法

- 无约束问题常用算法：
  - 下降法
  - 最速下降
  - 牛顿法
- 等式约束问题的常用算法：
  - 等式约束条件的消去
  - 牛顿法
- 内点法