最优化理论

陈鸿峥

2019.04*

目录

1 简介 **1**

1.1 优化概述 1

1.2 分类 2

1.3 历史 2

2 凸集 **3**

3 凸函数 **6**

4 凸优化问题 **11**

4.1 标准型 11

4.2 线性规划 13

5 对偶理论 **17**

1 简介

1.1 优化概述

优化(optimization)：从一个可行解的集合中寻找出最好的元素

例 1.

- 最小二乘线性拟合（凸问题）

---

*Build 20190404

- 深度神经网络（非凸，见下）

$$\begin{aligned}\mathbf{x}_1^{(i)} &= f_1(\mathbf{x}_0^{(i)}, \mathbf{w}_1)\\ \cdots &= \cdots\\ \mathbf{x}_n^{(i)} &= f_n(\mathbf{x}_{n-1}^{(i)}, \mathbf{w}_n)\\ \min \sum_{i=1}^{m} &(\mathbf{y}^{(i)} - \mathbf{x}_n^{(i)})^2\end{aligned}$$

- 图像处理，自然图像通常都是**分块光滑**的，原图Φ_0，有噪声的新图Φ
  全变参*(TV, Total Variation)*范数，计算图像每个像素点左侧和下侧的差异

$$\|\Phi\|_{TV} = \sum_y \sum_x \sqrt{(\Phi(x,y) - \Phi(x,y-1))^2 + (\Phi(x,y) - \Phi(x-1,y))^2}$$

  可得优化目标：近似自然图像，而且跟原图不能差太远

$$\min(\|\Phi\|_{TV} + \lambda\|\Phi - \Phi_0\|_F^2)$$

- 推荐系统：*Netflix*问题
  矩阵横向为用户，纵向为电影，值为评分值$(1 \sim 5)$，问题是把矩阵补全，这样就可以做推荐了$\rightarrow$低秩矩阵补全
  电影很多，但类型不多，关联关系有限$\rightarrow$**近似低秩**[1]
  低秩本来需要最小化$\mathbf{z}$的非零奇异值数目$\|\mathbf{z}\|_0$，但是非凸的；转化为最小化和范数[2]$\|\mathbf{z}\|_\star$

$$\begin{aligned}\min \quad & \|\mathbf{z}\|_\star := \|\mathbf{z}\|_1\\ s.t. \quad & \mathbf{z}_{ij} = \mathbf{M}_{ij},\ (i,j) \in \Omega\end{aligned}$$

1.2 分类

- 线性规划/非线性规划
- 凸规划/非凸规划（更好的分类）

目标函数凸函数，可行解集为凸集则是凸优化，一般容易求解

1.3 历史

- Newton-Raphson算法：求零点，等价于求$\min f^2(x)$
- Gauss-Seidel算法：求解线性方程组$A\mathbf{x} = \mathbf{b}$，等价于求$\min \|A\mathbf{x} - \mathbf{b}\|_2^2$
- Lagrange
- Kantoronc：苏联，线性规划，诺贝尔经济学奖

[1] A的秩等于非零奇异值$\sqrt{eig(A^{\mathrm{T}}A)}$数目

[2] 矩阵所有奇异值之和

- Dantzig：美国，优化决策，线性规划单纯形
- Von Neumann：线性规划问题对偶理论
- Karmarkar：80年代，线性规划内点法
- Nesterov：后80年代，非线性凸优化内点法
- 现代：并行、随机算法

2 凸集

定义 1. 一些集合概念如下

- 仿射集*(affine set)*

$$
\begin{aligned}
\mathcal{C}\text{为仿射集} &\iff \text{过}\mathcal{C}\text{内任意两点的}\textbf{直线}\text{都在}\mathcal{C}\text{内} \\
&\iff \forall x_1, x_2 \in \mathcal{C}, \theta \in \mathbb{R}, \theta x_1 + (1-\theta) x_2 \in \mathcal{C}
\end{aligned}
$$

  例 2. 用定义易证线性方程组的解集$\mathcal{C} = \{\mathbf{x} \mid A\mathbf{x} = \mathbf{b}\}$是仿射集；反过来，每一个仿射集都可以用线性方程组的解集表示

- 仿射组合

$$
\forall x_1, x_2, \ldots, x_k \in \mathcal{C}, \theta_1, \ldots, \theta_k \in \mathbb{R}, \theta_1 + \cdots + \theta_k = 1 :\ \theta_1 x_1 + \cdots + \theta_k x_k \in \mathcal{C}
$$

- 仿射包*(hull)*：所有仿射组合的集合

$$
\operatorname{aff} \mathcal{C} := \{\theta_1 x_1 + \cdots + \theta_k x_k \mid \forall x_1, \ldots, x_k \in \mathcal{C}, \theta_1 + \cdots + \theta_k = 1\}
$$

- 凸集*(convex set)*

$$
\begin{aligned}
\mathcal{C}\text{为凸集} &\iff \text{过}\mathcal{C}\text{内任意两点的}\textbf{线段}\text{都在}\mathcal{C}\text{内} \\
&\iff \forall x_1, x_2 \in \mathcal{C}, \theta \in \mathbf{[0,1]}, \theta x_1 + (1-\theta) x_2 \in \mathcal{C}
\end{aligned}
$$

- 凸组合

$$
\forall x_1, x_2, \ldots, x_k \in \mathcal{C}, \theta_1, \ldots, \theta_k \in [0,1], \theta_1 + \cdots + \theta_k = 1 :\ \theta_1 x_1 + \cdots + \theta_k x_k \in \mathcal{C}
$$

- 凸包：最小的凸集

$$
\operatorname{conv} \mathcal{C} := \{\theta_1 x_1 + \cdots + \theta_k x_k \mid \forall x_1, \ldots, x_k \in \mathcal{C}, \theta_1, \ldots, \theta_k \in [0,1], \theta_1 + \cdots + \theta_k = 1\}
$$

- 凸锥*(convex cone)*

$$\mathcal{C}\text{为凸锥} \iff \forall x_1, x_2 \in \mathcal{C}, \theta_1, \theta_2 \geq 0, \theta_1 x_1 + \theta_2 x_2 \in \mathcal{C}$$

除了**空集**的凸锥都得包含**原点**（取$\theta_1 = \theta_2 = 0$）

- 凸锥组合/非负线性组合：

$$\forall x_1, x_2, \ldots, x_k \in \mathcal{C}, \theta_1, \ldots, \theta_k \geq 0 : \ \theta_1 x_1 + \cdots + \theta_k x_k \in \mathcal{C}$$

- 凸锥包：类似前面定义

由上面的定义易知，仿射组合/凸锥组合（强条件）一定是凸组合。

定义 2 (超平面(hyperplane)与半空间(halfspace))**.** 超平面都是比原空间低一维

$$\{\mathbf{x} \mid \mathbf{a}^{\mathrm{T}}\mathbf{x} = b, \mathbf{x}, \mathbf{a} \in \mathbb{R}^n, b \in \mathbb{R}, \mathbf{a} \neq 0\}$$

超平面将空间划分为两个部分，即半空间

$$\{\mathbf{x} \mid \mathbf{a}^{\mathrm{T}}\mathbf{x} \leq b, \mathbf{a} \neq 0\}$$

若方程特解为$\mathbf{x}_0$，则$\mathbf{a} \perp (\mathbf{x} - \mathbf{x}_0)$

定义 3 (欧式球(Euclidean ball))**.**

$$B(x_c, r) = \{x \mid \|x - x_c\|_2 \leq r\}$$

范数*(norm)*球可类似定义

定义 4 (椭球(ellipsoid))**.**

$$\varepsilon(x_c, P) = \{x \mid (x - x_c)^{\mathrm{T}} P^{-1} (x - x_c) \leq 1\}, P \succ 0$$

其中$P \succ 0$代表P**对称**且正定$(P = P^T)$

分析. 定义内积$\langle x^{\mathrm{T}} P^{-1} y \rangle$（需证满足内积条件），进而$P$-范数$\|x\|_P := \sqrt{x^{\mathrm{T}} P x}$是范数，而椭球不过是$P$-范数意义下的球，由定理得椭球是凸的

定义 5 (多面体(polyhedron))**.**

$$P = \{\mathbf{x} \mid \mathbf{a}_i^{\mathrm{T}}\mathbf{x} \leq b_i, \mathbf{c}_j^{\mathrm{T}}\mathbf{x} = d_j, i = 1, \ldots, m, j = 1, \ldots, p\}$$

例 3.
- 空集、点、$\mathbb{R}^n$空间均为仿射
- 任意直线为仿射；若过原点则为凸锥

- $\mathbb{R}^n$空间的子空间[3]为仿射和凸锥
- 超平面为仿射
- 半空间、欧式球、椭球、多面体为凸集

定义 6 (仿射函数)**.**

$$f:\mathbb{R}^n \mapsto \mathbb{R}^m \quad f(\mathbf{x}) = A\mathbf{x}+\mathbf{b}, A\in\mathbb{R}^{m\times n}, \mathbf{b}\in\mathbb{R}^m$$

性质如下：

- $S\subset\mathbb{R}^n$为凸 $\Longrightarrow f(S)=\{f)\mathbf{x}\mid \mathbf{x}\in S\}$为凸
- $C\subset\mathbb{R}^m$为凸 $\Longrightarrow f^{-1}(C)=\{\mathbf{x}\in\mathbb{R}^n\mid f(\mathbf{x})\in C\}$为凸

例 4. 两个集合的和$S_1+S_2=\{x+y\mid x\in S_1, y\in S_2\}$保凸

分析. 直积$S_1\times S_2=\{(x,y)\mid x\in S_1, y\in S_2\}$显然可以保凸（相当于在两个集合同时画线）
令$A=\begin{bmatrix} I & I\end{bmatrix}, \mathbf{x}=\begin{bmatrix} x & y\end{bmatrix}^{\mathrm{T}}, \mathbf{b}=0$，由仿射函数性质知

定义 7 (透视(perspective)函数[4])**.** 透视函数$P:\mathbb{R}^{n+1}\mapsto\mathbb{R}^n, \operatorname{dom}P=\mathbb{R}^n\times\mathbb{R}_{++}$定义如下

$$P(z,t)=\frac{z}{t}, z\in\mathbb{R}^n, t\in\mathbb{R}_{++}$$

反透视函数

$$P^{-1}(c):=\{(x,t)\in\mathbb{R}^{n+1}\mid \frac{x}{t}\in c, t>0\}$$

若$c\in\operatorname{dom}P$为凸，则$P(c):=\{P(x), x\in c\}$为凸；反透视函数仍保持c的凸性。

考虑$\mathbb{R}^{n+1}$内的线段，$x=(\widetilde{x}\in\mathbb{R}^n, x_{n+1}\in\mathbb{R}_{++}), y=(\widetilde{y}, y_{n+1})$ 则经过透视函数仍是线段

分析.

$$P(\theta x+(1-\theta)y)=\frac{\theta\widetilde{x}+(1-\theta)\widetilde{y}}{\theta x_{n+1}+(1-\theta y_{n+1})}=\frac{\theta x_{n+1}}{\theta x_{n+1}+(1-\theta)y_{n+1}}\frac{\widetilde{x}}{x_{n+1}}+\frac{(1-\theta)y_{n+1}}{\theta x_{n+1}+(1-\theta)y_{n+1}}\frac{\widetilde{y}}{y_{n+1}}$$

定义 8 (线性分数函数)**.** 仿射函数

$$g(x)=\begin{bmatrix} A \\ C^{\mathrm{T}}\end{bmatrix}x+\begin{bmatrix} b \\ d\end{bmatrix}, A\in\mathbb{R}^{m\times n}, b\in, \mathrm{c}\in\mathbb{R}^{\mathrm{n}}, \mathrm{d}\in\mathbb{R}$$

线性分数函数$f:\mathbb{R}^n\mapsto=\mathrm{p}\circ\mathrm{g}$

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

保凸性

- 凸集的交

[3]零元、加法封闭、数乘封闭

[4]$+$代表≥ 0，$++$代表>0

- 仿射、逆仿射
- 透视函数
- 线性分数函数

3 凸函数

定义 9 (凸函数). *1.* $f:\mathbb{R}^n\to\mathbb{R}$为凸 $\iff$ $\operatorname{dom} f$为凸且$\forall x,y\in\operatorname{dom} f,\theta\in[0,1]$

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

- 严格凸：$\theta\in(0,1)$，不等式不能取等
- 凹函数：若$-f$为凸

2. 高维定义：$f:\mathbb{R}^n\mapsto\mathbb{R}$为凸 $\iff$ $\operatorname{dom} f$为凸

$$\forall x\in\operatorname{dom} f,v\in\mathbb{R}^n:g(t):=f(x+tv)\text{为凸，}\quad\operatorname{dom} g=\{t\mid x+tv\in\operatorname{dom} f\}$$

相当于每一个剖面上的低维函数都是凸的

3. 一阶条件*(first-order condition)*[5]

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

4. 二阶条件：$f:\mathbb{R}^n\mapsto\mathbb{R}$为凸 $\iff$ $\operatorname{dom} f$为凸

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

- 凹函数：$\nabla^2 f(x)\preceq 0$
- 严格凸：$\Longleftarrow$ $\nabla^2 f(x)\succ 0$，反例$f(x)=x^4$（在一个点斜率不变并不要紧）

例 5. $f(x)=a^{\mathrm{T}}x+b$

分析. 有$\nabla f(x)=a$，进而

$$f(y)=a^{\mathrm{T}}y+b\ge a^{\mathrm{T}}x+b+a^{\mathrm{T}}(y-x)=a^{\mathrm{T}}y+b$$

定义 10 (凸函数的扩展(extended-value)). 尽管凸函数的定义域为凸，但往往不好处理，那就将其扩展到全空间。$x\in\operatorname{dom} f\subset\mathbb{R}^n,\operatorname{dom}\widetilde{f}=\mathbb{R}^n$，会有

$$\widetilde{f}(x)=\begin{cases}f(x) & x\in\operatorname{dom} f\\ +\infty & x\notin\operatorname{dom} f\end{cases}$$

[5] $\nabla^{\mathrm{T}}f(x)=[\nabla f(x)]^{\mathrm{T}}$

指示/示信(indicator)函数不一定是凸的

$$f(x)=\begin{cases}0 & x\in C\\ +\infty & x\notin C\end{cases}$$

定理 1. 若f为凸，可微，则$\exists x\in \operatorname{dom} f, \nabla f(x)=0$

例 6. 二次函数$f(x)=\frac{1}{2}x^{\mathrm{T}}Px+q^{\mathrm{T}}x+r$，$P\in S^n$（对称矩阵），$q^{\mathrm{T}}\in\mathbb{R}^n$，$r\in\mathbb{R}$

分析. $\nabla^2 f(x)=P$

$P\in S_+^n$凸，$P\in S_{++}^n$严格凸

例 7. $f(x)=\dfrac{1}{x^2}, \operatorname{dom} f=\{x\in\mathbb{R}, x\neq 0\}$

分析. 注意$\operatorname{dom} f$不是凸集

- 指数函数$f(x)=\mathrm{e}^{ax}$
- 幂函数$f(x)=x^a$
- 绝对值的幂函数$f(x)=|x|^p, x\in\mathbb{R}, p>0$：$p\in[1,+\infty)$凸，$p\in(0,1)$既不凸又不凹

  分析.

$$f''(x)=\begin{cases}p(p-1)x^{p-2} & x<0\\ p(p-1)(-x)^{p-2} & x<0\end{cases}$$

- 对数函数$f(x)=\log x$
- 熵$f(x)=-x\log x$
- 极大值函数$f(x)=\max\{x_1,\dots,x_n\}, x\in\mathbb{R}^n$

定义 11 (解析近似)**.** 无穷阶可微

极大值函数的解析近似是$f(x)=\log(\mathrm{e}^{x_1}+\cdots+\mathrm{e}^{x_n})$

$$\max\{x_1,\dots,x_n\}\le f(x)\le \max\{x_1,\dots,x_n\}+\log n$$

分析.

$$\frac{\partial f}{\partial x_i}=\frac{\mathrm{e}^{x_i}}{\mathrm{e}^{x_i}+\cdots+\mathrm{e}^{x_n}}$$

$$\frac{\partial^2 f}{\partial x_i\partial x_j}=\begin{cases}\frac{-\mathrm{e}^{x_i}\mathrm{e}^{x_i}}{i}=j\\ i\neq j\end{cases}$$

$$z:=\begin{bmatrix}\mathrm{e}^{x_1} & \cdots & \mathrm{e}^{x_n}\end{bmatrix}^{\mathrm{T}}$$

求*Hessian*矩阵

$$H=\frac{1}{(\mathbb{1}^{\mathrm{T}}z)^2}(-z\cdot z^{\mathrm{T}}+(\mathbb{1}^{\mathrm{T}}z)\operatorname{diag}(z))$$

将前面常量丢弃 [6]

$$a_i := v_i\sqrt{z_i} = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^{\mathrm{T}}, b_i = \sqrt{z_i}$$

$$\begin{aligned} v^{\mathrm{T}}Hv &= (\mathbb{1}^{\mathrm{T}}z)v^{\mathrm{T}}\operatorname{diag}(z)v - v^{\mathrm{T}}zz^{\mathrm{T}}v \\ &= (\sum_i z_i)(\sum_i v_i^2 z_i) - (\sum_i v_i z_i)^2 \\ &= (b^{\mathrm{T}}b)(a^{\mathrm{T}}a) - (a^{\mathrm{T}}b)^2 \qquad \mathit{Cauchy} \\ &\geq 0 \end{aligned}$$

定义 12 (范数)**.** $p(x)$为范数

1. $p(ax) = |a|p(x)$
2. $p(x+y) \leq p(x) + p(y)$
3. $p(x) = 0 \iff x = 0$

零范数$\|x\|_0$：非零元素数目，是伪范数（不符合第一个定义）

$\mathbb{R}^n$中的范数都是凸函数，正则化!

分析.

$$\forall x, y, \theta \in [0,1] p(\theta x + (1-\theta)y) \leq p(\theta x) + p((1-\theta)y) \leq \theta p(x) + (1-\theta)p(y)$$

行列式的对数$f(x) = \log\det(x), \operatorname{dom} f = S_{++}^n$ $n = 1$凹函数证$n > 1$也为凹，用高维定义

$$\begin{aligned} g(t) &= f(z + tv) \\ &= \log\det(z + tv) \\ &= \log\det(z^{1/2}(I + tz^{1/2}vz^{-1/2})z^{1/2}, \quad z^{1/2} \in S_{++}^n, z^{1/2}z^{1/2} = z \\ &= \log\det(z) + \log\det(I + tz^{1/2}vz^{-1/2}) \\ &= \log\det(z) + \sum_{i=1}^n \log(1 + t\lambda_i), \quad \lambda_i = z^{-1/2}vz^{1/2}\text{的特征值} \end{aligned}$$

$$\begin{aligned} g'(t) &= \sum_{i=1}^n \frac{\lambda_i}{1 + t\lambda_i} \\ g''(t) &= \sum_{i=1}^n -\frac{\lambda_i^2}{(1 + t\lambda_i)^2} \end{aligned}$$

补充证明：对对称阵特征值分解$tz^{1/2}vz^{1/2} = tQ\Lambda Q^{\mathrm{T}}$，对角阵$\Lambda$即为 $QQ^{\mathrm{T}} = I$，Q为酉矩阵

$$I + tz^{-1/2}vz^{-1/2} = QQ^{\mathrm{T}} + tQ\Lambda Q^{\mathrm{T}} = Q(I + t\Lambda)Q^{\mathrm{T}}$$

$$\log\det(I + tz^{-1/2}vz^{-1/2}) = \log\det(Q) + \log\det(I + t\Lambda) + \log\det(Q^{\mathrm{T}})$$

保持函数凸性

[6] H半正定，则$\forall v \in \mathbb{R}^n : v^{\mathrm{T}}Hv \geq 0$

- 非负加权和$f_1,\dots,f_m$为凸，定义域$\mathbb{R}^n$

$$f := \sum_{i=1}^{m} w_i f_i, w_i \geq 0$$

- 非负积分$f(x,y)$对$y \in A$均为凸（A不一定为凸），$w(y) \geq 0$

$$g(x) := \int_{y\in A} w(y) f(x,y)\,\mathrm{d}y$$

- 仿射映射$f : \mathbb{R}^n \mapsto \mathbb{R}$为凸，$A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^n$，$\operatorname{dom} g = \{x \mid Ax + b \in \operatorname{dom} f\}$

$$g(x) := f(Ax + b)$$

分析. – $\operatorname{dom} f$为凸，则$\operatorname{dom} g$为凸

– $\forall x, y \in \operatorname{dom} g, \forall\theta \in [0,1]$

$$\begin{aligned} g(\theta x + (1-\theta)y) &= f(A(\theta x + (1-\theta)y) + b) \\ &= f(\theta(Ax + b) + (1-\theta)(Ay + b)) \\ &\leq \theta f(Ax + b) + (1-\theta) f(Ay + b) \\ &= \theta g(x) + (1-\theta) g(y) \end{aligned}$$

– 其实只是在定义域上改变，而不是改变值域，因而函数凸性不会改变

- 两个函数的极大值函数，f_1, f_2为凸

$$f(x) := \max\{f_1(x), f_2(x)\}, \operatorname{dom} f = \operatorname{dom} f_1 \cap \operatorname{dom} f_2$$

- 任意个凸函数极大值函数为凸

$$f(x) = \max\{a_1^{\mathrm{T}} x + b_1, \dots, a_m^{\mathrm{T}} + b_m\}$$

- 无限个凸函数，$y \in A$，$f(x,y)$对于x为凸，则

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

例 8. 点x到集合C的最远距离

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

位移对于范数凸性不会有影响

例 9. $x\in\mathbb{R}^n$，$x[i]$为第i大元素，$x[1]\geq x[2]\geq\cdots\geq x[r]\geq\cdots\geq x[n]$

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

- $r=1$：$f(x)=x[1]=\max\{x_1,\ldots,x_n\}$，每一项都是$\mathbf{e}_i^{\mathrm{T}}x_i$
- $r>1$：$f(x)=\max\{x_{i_1}+\cdots+x_{i_r}\mid 1\leq i_1<i_2<\cdots<i_r\leq n\}$

- 函数的组合：$h:\mathbb{R}^k\mapsto\mathbb{R},g:\mathbb{R}^n\mapsto\mathbb{R}^k$

$$f:=h\circ g:\mathbb{R}^n\mapsto\mathbb{R}$$

先考虑$n=k=1,\operatorname{dom}g=\mathbb{R}^n,\operatorname{dom}h=\mathbb{R}^k,\operatorname{dom}f=\mathbb{R}$，$h,g$二阶可微

$$\begin{aligned}&f'(x)=h'(g(x))\cdot g'(x)\\&f''(x)=h''(g(x))(g'(x))^2+h'(g(x))g''(x)>0\end{aligned}$$

即当g为凸，h为凸且不降；g为凹，h为凸且不增时，$f(x)$为凸
（若定义域非全空间）当g为凸，h为凸，扩展值函数$\tilde{h}$不降；g为凹，h为凸，$\tilde{h}$不增时，$f(x)$为凸

例 10. g为凸，$\exp g(x)$为凸；g为凹，$g>0$，$\log g(x)$为凹；g为凸，$g>0$，$1/g(x)$为凸

例 11. $g(x)=x^2,\operatorname{dom}g=\mathbb{R}$，$h(y)=0,\operatorname{dom}h=[1,2]$，$f=h\circ g$，注意$\tilde{h}$并非不降!

- 函数透视：$P:\mathbb{R}^{n+1}\mapsto\mathbb{R}^n,\operatorname{dom}P\in\mathbb{R}^n\times\mathbb{R}_{++},P(z,t)=\dfrac{z}{t}$

$$f:\mathbb{R}^n\mapsto\mathbb{R},g(x,t)=tf(\frac{x}{t}),\operatorname{dom}g=\{(x,t)\mid\frac{x}{t}\in\operatorname{dom}f\},g:\mathbb{R}^n\times\mathbb{R}_{++}\mapsto\mathbb{R}$$

若$f(x)$为凸，则$g(x,t)$相对于(x,t)联合凸

例 12.
- $f(x)=x^{\mathrm{T}}x$, $g(x,t)=x^{\mathrm{T}}x/t$
- $f(x)=-\log x$, $g(x,t)=t\log(t/x)$
- $u,v\in\mathbb{R}^n_{++}$, $g(u,v)=\sum_{i=1}^n u_i\log(u_i/v_i)$，信息论常用，衡量相似性，$KL$散度

$$D_{KL}:=\sum_{i=1}^{n}\left(u_i\log\frac{u_i}{v_i}-u_i+v_i\right)$$

定义 13 (α次水平集(α-sub level set))**.** $f:\mathbb{R}^n\mapsto\mathbb{R}$，$C_\alpha=\{x\in\operatorname{dom}f\mid f(x)\leq\alpha\}$

定义 14 (拟凸函数(quasi-convex))**.** α次水平集为凸集$\iff f$为拟凸函数

拟凸函数有很好的性质→单模态/单峰函数
凸函数与凸集联系

- 凸函数定义域为凸集
- 凸函数的α次水平集为凸集

4 凸优化问题

4.1 标准型

广义定义：极小化凸函数，约束为凸集

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

- 优化变量$\mathbf{x} \in \mathbb{R}^n$
- 目标/损失函数$f_0 : \mathbb{R}^n \mapsto \mathbb{R}$
- 不等式约束函数$f_i : \mathbb{R}^n \mapsto \mathbb{R}$
- 等式约束函数$h_j : \mathbb{R}^n \mapsto \mathbb{R}$
- 域$\mathcal{D} = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i$
- 可行解$\mathcal{X} = \{\mathbf{z} \mid f_i(\mathbf{z}) \le 0, h_j(\mathbf{z}) = 0, i = 1, \ldots, m, j = 1, \ldots, p\}$
- 最优值$P^\star = \inf\{f_0(\mathbf{x}) \mid x \in \mathcal{X}\}$
- 最优解$\mathbf{x}^\star \iff \forall \mathbf{z} \in \mathbb{R}^n, \mathbf{z} \in \mathcal{X} : \ f_0(\mathbf{z}) \ge f_0(\mathbf{x}^\star)$
- 最优解集$X^\star = \{x^\star \mid f_0(\mathbf{x}^\star) = P^\star, \mathbf{x}^\star \in \mathcal{S}\}$
- ε-次优解集$X_\varepsilon = \{\mathbf{x} \mid f_0(\mathbf{x}) \le P^\star + \varepsilon, \mathbf{x} \in \mathcal{X}\}$
- 局部最优$\exists R > 0, f_0(x) = \inf\{f_0(\mathbf{z}) \mid \mathbf{x} \in \mathcal{X}, \mathbf{z} \in \mathcal{X}, \|\mathbf{x} - \mathbf{z}\| \le R\}$
- 局部最优解集$x_{local} = \{\mathbf{x} \mid \mathbf{x}\text{为局部最优}\}$

狭义定义：$f_i(x), i = 0, 1, \ldots$为凸函数，$h_i(x)$为仿射函数

例 13.

$$
\begin{aligned}
\min \quad & f_0(x) = x_1^2 + x_2^2 \\
s.t. \quad & f_1(x) = \frac{x_1}{1 + x_2^2} \le 0 \implies x_1 \le 0 \\
& h_1(x) = (x_1 + x_2)^2 = 0 \implies x_1 + x_2 = 0
\end{aligned}
$$

定理 2. 凸问题局部最优等价于全局最优

分析. 若x为局部最优

$$\exists R > 0 : \ f_0(x) = \inf\{f_0(z) \mid z \in \mathcal{X}, x \in \mathcal{X}, \|x - z\|_2 \le R\}$$

反证法，设x不是全局最优，y为全局最优，$f_0(x) > f_0(y)$
$z = \theta x + (1 - \theta) y, \theta = \frac{R}{2\|y - x\|_2}$

$$\|z - x\|_2 = \frac{R\|y - x\|_2}{2\|y - x\|_2} = \frac{R}{2}$$

由$\|z-x\|_2 \leq R \implies f_0(x) \leq f_0(z)$，有

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

矛盾

可微凸目标函数

无约束$\min f_0(x), \nabla f_0^\star(x) = 0$

$$\forall x, y : f_0(y) \geq f_0(x) + \langle \nabla f_0(x), y-x \rangle$$
$$f_0(y) \geq f_0(x^\star) + \langle \nabla f_0(x^\star), y-x \rangle = f_0(x^\star)$$

有约束$\min f_0(x),\ s.t. x \in \mathcal{X}$

$$x^\star \in \mathcal{X}, \langle \nabla f_0(x^\star), y - x^\star \rangle \geq, \forall y \in x$$

例 14. 等式约束 $\min f_0(x), \operatorname{dom} f_0 \subset \mathbb{R}^n$，$f_0$可微，使得$Ax = b$

分析. $x^\star$最优，$Ax^\star = b, \forall y Ay = b$

$$\langle \nabla f_0(x^\star), y - x^\star \rangle \geq 0$$
$$\begin{cases} y = x^\star + v \\ Av = 0 \end{cases}, v \in \operatorname{Nul} A$$
$$\forall v \in \operatorname{Nul} A, \langle \nabla f_0(x^\star), v \rangle \geq 0$$

1. $\operatorname{Nul} A = \{0\}$
2. A不可逆，$\nabla f_0(x^\star) \perp \operatorname{Nul} A$

例 15. 正约束$\min f_0(x), s.t. x \geq 0$

分析. 若$x^\star$最优，$\iff x^\star \geq 0, \forall y \geq 0, \langle \nabla f_0(x^\star), y - x^\star \rangle \geq 0$

$$\iff \langle \nabla f_0(x^\star, y) \rangle \geq \langle \nabla f_0(x^\star, x^\star) \rangle$$

1. 若$\nabla f_0(x^\star) \ngeq 0$有矛盾（负数行乘上正无穷），故$\nabla f_0(x^\star) \geq 0$
2. 令$y = 0$，有$0 \geq \langle \nabla f_0(x^\star), x^\star \rangle \implies \sum_{i=1} n(\nabla f_0(x^\star)_i x^\star) \leq 0$
   前面≥ 0，进而互补松弛条件
3. $x^\star \geq 0$

4.2 线性规划

$$\begin{aligned} \min \quad & c^{\mathrm{T}}\mathbf{x}+\mathbf{d} \\ \text{s.t.} \quad & G\mathbf{x} \leq \mathbf{h} \\ & A\mathbf{x}=\mathbf{b} \end{aligned}$$

$$\begin{aligned} \min \quad & c^{\mathrm{T}}\mathbf{x}+\mathbf{d} \\ \text{s.t.} \quad & G\mathbf{x}+\mathbf{s}=\mathbf{h} \\ & A\mathbf{x}=\mathbf{b} \\ & \mathbf{s} \geq 0 \end{aligned}$$

$\mathbf{s}$为松弛变量(slack variable)
用$\mathbf{x}^+$和$\mathbf{x}^-$拆分，得到$\mathbf{x}=\mathbf{x}^+-\mathbf{x}^-,\mathbf{x}^+\geq 0,\mathbf{x}^-\geq 0,\mathbf{s}\geq 0$

例 16 (食谱问题)**.** m种营养元素不小于$b_1,\dots,b_m$，n种食物，单位含量$a_{1j},\dots,a_{mj}$，食物量$x_1,\dots,x_n$，价格$c_1,\dots,c_n$

$$\begin{aligned} \min \quad & \sum_{i=1}^{n} c_j x_j \\ s.t. \quad & \sum_{j=1}^{n} a_{ij}x_j \geq b_i \\ & x_j \geq 0 \end{aligned}$$

其中$i=1,\dots,m,j=1,\dots,n$

线性分数规划

$$\begin{aligned} \min \quad & f_0(x)=\frac{c^{\mathrm{T}}x+d}{e^{\mathrm{T}}x+f}, \operatorname{dom} f=\{x \mid e^{\mathrm{T}}x+f>0\} \\ \text{s.t.} \quad & Gx \leq h \\ & Ax=b \end{aligned}$$

等价于

$$\begin{aligned} \min \quad & c^{\mathrm{T}}y+dz \\ \text{s.t.} \quad & Gy-hz \leq 0 \\ & Ay-bz=0 \\ & e^{\mathrm{T}}y+fz=1 \\ z \geq 0 \end{aligned}$$

分析. 证明两个问题等价，P_0与P_1
若x在P_0内可行

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

若(y,z)在P_1中可行

$$x=\frac{y}{z}(z\neq 0)$$

若$z=0$，x_0为P_0的可行解

$$x=x_0+ty, t\geq 0$$

$$\lim_{t\to\infty}\frac{c^{\mathrm{T}}(x_0+ty)+d}{e^{\mathrm{T}}(x_0+ty)+f}=c^{\mathrm{T}}y$$

代入看所有条件结论都相同

二次规划(Quadratic Programming)

$$\begin{aligned}\min\quad & \frac{1}{2}x^{\mathrm{T}}px+q^{\mathrm{T}}+r,\ p\succ 0\\ \text{s.t.}\quad & Gx\leq h\\ & Ax=b\end{aligned}$$

二次约束二次规划(QCQP)

$$\begin{aligned}\min\quad & \frac{1}{2}p_0x+q_0^{\mathrm{T}}x+r_0, p\succ 0\\ \text{s.t.}\quad & \frac{1}{2}x^{\mathrm{T}}p_ix+q_i^{\mathrm{T}}x+r_i\leq 0, i=1,\ldots,m, p_i\succ 0\\ & Ax=b\end{aligned}$$

最小二乘问题

$$\begin{aligned}\min_x\quad & \frac{1}{2}\|Ax-b\|_2^2\\ \text{s.t.}\quad & b=Ax+e\end{aligned}$$

$$\frac{1}{2}(x^{\mathrm{T}}A^{\mathrm{T}}Ax-2b^{\mathrm{T}}Ax+b^{\mathrm{T}}b)$$

一范数规范化最小二乘

$$\min\frac{1}{2}\|Ax-b\|_2^2+\lambda_1\|x\|_1$$

本来用零范数，但用一范数拟合

改写

$$\|x\|_1=\mathbb{1}^{\mathrm{T}}\mathbf{x}^{+}+\mathbb{1}^{\mathrm{T}}\mathbf{x}^{-}$$

Basic Pursuit

$$\begin{aligned}\min\quad & \frac{1}{2}\|Ax-b\|_2^2\\ \text{s.t.}\quad & \|x\|_1\leq\varepsilon_1\end{aligned}$$

原式很难平衡两者，下式只需考虑$\|x\|_1$的影响

岭回归(Ridge)：所有x差距不要太大

$$\min\frac{1}{2}\|Ax-b\|_2^2+\frac{1}{2}\lambda_2\|x\|_2^2$$

$$\begin{aligned} \min \quad & \frac{1}{2}\|Ax-b\|_2^2 \\ \text{s.t.} \quad & \|x\|_2^2 \le \varepsilon_2 \end{aligned}$$

投资组合问题(portfolio optimization)：初始价格$x_1,\dots,x_n$，最终价格$P_1x_1,\dots,P_nx_n$

$$\begin{aligned} \max \quad & P_1x_1+\cdots+P_nx_n \\ \text{s.t.} \quad & x_1+\cdots+x_n=B \\ & x_1,\dots,x_n \ge 0 \end{aligned}$$

$\bar{P}=\mathbb{E}(P)$已知，$\Sigma=\mathbb{D}(P)$

$$\begin{aligned} \min \quad & x^{\mathrm{T}}\Sigma x \\ \text{s.t.} \quad & p^{\mathrm{T}}x \ge r_{\min} \\ & x_1+\cdots+x_n=B \\ & x_1,\dots,x_n \ge 0 \end{aligned}$$

半定规划(semi-definite programming, SDP)（矩阵意义下的线性规划问题）：$X\in\mathbb{R}^{n\times n}, C\in\mathbb{R}^{n\times n}, A_i\in\mathbb{R}^{n\times n}, b_i\in\mathbb{R}$

$$\begin{aligned} \min \quad & \mathrm{tr}(CX) \\ \text{s.t.} \quad & \mathrm{tr}(A_iX)=b_i, i=1,\dots,p \\ & X\succeq 0 \end{aligned}$$

例 17 (谱范数极小化问题)**.** 矩阵多项式$A(x)=A_0+x_1A_1+\cdots+x_nA_n, A_i\in\mathbb{R}^{p\times q}$

$$\min_x \|A(x)\|_2$$

谱范数代表$A(x)$的最大奇异值[7]

$$\begin{aligned} \min_{x,s} \quad & S \\ \text{s.t.} \quad & A^{\mathrm{T}}(x)A(x)\preceq SI \end{aligned}$$

例 18 (最快分布式线性平均)**.**

$$x(t)=Px(t-1)$$

$$P=\begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & & \vdots \\ P_{n1} & \cdots & P_{nn} \end{bmatrix}, P\mathbb{1}=\mathbb{1}$$

其中$(i,j)\in E$或$i=j$，$P_{ij}\neq 0$；否则$P_{ij}=0$

$P=P^{\mathrm{T}}, P_{ij}=P_{ji}$，$P_{ij}>0$

$P\succeq 0, P_{ij}\ge 0$

[7]谱范数是诱导范数，F-范数(Frobenias)$\|A(x)\|_F$才算是矩阵意义下的2范数

只要图是连通图，则一定会收敛

收敛速度与第二大|特征值绝对值|有关

$$1 = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq -1$$

即收敛速度由$\max\{|\lambda_2|, |\lambda_n|\}$决定

$$\min \max\{|\lambda_2|, |\lambda_n|\}$$
$$\max\{|\lambda_2|, |\lambda_n|\} = \left\| P - \frac{1}{n}\mathbb{1}\mathbb{1}^{\mathrm{T}} \right\|$$

$$\begin{aligned}
\min \quad & t := \left\| P - \frac{1}{n}\mathbb{1}\mathbb{1}^{\mathrm{T}} \right\|_2 \\
\text{s.t.} \quad & P\mathbb{1} = \mathbb{1} \\
& P = P^{\mathrm{T}} \\
& P \succeq 0 \\
& P_{ij} = 0, \quad (i,j) \neq E \wedge i \neq j \\
& -tI \preceq P - \frac{1}{n}\mathbb{1}\mathbb{1}^{\mathrm{T}} \preceq tI
\end{aligned}$$

多目标优化问题：帕累托最优解

若有另一解在某个指标上更好，则必有指标更差

帕累托最优值/帕累托最优面

$\min f_{01}$与$\min f_{02}$的交点为理想点(oracle)

若$f_{01}(x), \ldots, f_{0q}(x)$为凸，$\mathcal{X}$为凸

$$\begin{aligned}
\min \quad & \lambda_1 f_{01}(x) + \cdots + \lambda_q f_{0q}(x), \quad \lambda_1, \ldots, \lambda_q \geq 0 \\
\text{s.t.} \quad & x \in \mathcal{X}
\end{aligned}$$

1. 能找到一个Pareto最优解
2. 遍历$\lambda_1, \ldots, \lambda_q$，可找到全部

岭回归的多目标优化表示

$$\begin{cases} \min \frac{1}{2}\left\| Ax - b \right\|_2^2 \\ \min \frac{1}{2}\left\| x \right\|_2^2 \end{cases}$$

5 对偶理论

拉格朗日函数(Lagrangian function)

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

拉格朗日乘子(multiplier)

- 原变量(primal variable)：$\lambda=\begin{bmatrix}\lambda_1 & \cdots & \lambda_m\end{bmatrix}^{\mathrm{T}}$
- 对偶变量(dual variable)：$v=\begin{bmatrix}v_1 & \cdots & v_p\end{bmatrix}^{\mathrm{T}}$

拉格朗日对偶函数

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

注意遍历域是$\mathcal{D}=\bigcap_{i=0}^{m}\operatorname{dom} f_i\cap\bigcap_{i=1}^{p}\operatorname{dom} h_i$，而不是可行解集$\mathcal{X}$

- $g(\lambda,v)$一定是关于λ和v的凹函数（关于λ和v的仿射函数，注意x为常数）
- $\forall\lambda\geq 0,\forall v,g(\lambda,v)\leq P^\star$

  对偶(dual)问题

$$\begin{aligned}\max\quad & g(\lambda,v)\\ \text{s.t.}\quad & \lambda\geq 0\end{aligned}$$

  其最优解记为$D^\star$，则$D^\star\leq P^\star$，即给出了原问题的一个最优下界

$x^\star$原问题最优解

$$\sum_{i=1}^{m}\lambda_i f_0(x^\star)+\sum_{i=1}^{p}v_i h_i(x^\star)\leq 0$$

$$L(x^\star,\lambda,v)=f_0(x^\star)+(\cdots)\leq P^\star$$

$$g(\lambda,v)=\inf_{x\in\mathcal{D}}L(x,\lambda,v)\leq L(x^\star,\lambda,v)\leq P^\star$$

例 19.

$$\begin{aligned}\min\quad & x^{\mathrm{T}}x\\ \text{s.t.}\quad & Ax=b\end{aligned}$$

分析.

$$L(x,v)=x^{\mathrm{T}}x+v^{\mathrm{T}}(Ax-b)$$

$$\begin{aligned}
g(v) &= \inf_{x\in\mathcal{D}} L(x,v) \\
&= \inf_{x\in\mathcal{D}} x^{\mathrm{T}}x + v^{\mathrm{T}}Ax - v^{\mathrm{T}}b \\
&= (-\frac{A^{\mathrm{T}}v}{2})^{\mathrm{T}}(-\frac{A^{\mathrm{T}}v}{2}) + v^{\mathrm{T}}A(-\frac{A^{\mathrm{T}}v}{2}) - v^{\mathrm{T}}b \\
&= -\frac{1}{4}v^{\mathrm{T}}AA^{\mathrm{T}}v - b^{\mathrm{T}}v
\end{aligned}$$

补充求梯度：$2x + A^{\mathrm{T}}v = 0 \implies x = -\frac{A^{\mathrm{T}}v}{2}$

因而得到对偶问题

$$\max_v -\frac{1}{4}v^{\mathrm{T}}AA^{\mathrm{T}}v - b^{\mathrm{T}}v$$

例 20.

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

分析. 注意λ前面符号，要化为一般形式

$$L(x,\lambda,v) = c^{\mathrm{T}}x - \lambda^{\mathrm{T}}x + v^{\mathrm{T}}(Ax-b)$$

$$\begin{aligned}
g(\lambda,v) &= \inf_x L(x,\lambda,v) \\
&= \inf_x (c-\lambda+A^{\mathrm{T}}v)^{\mathrm{T}}x - v^{\mathrm{T}}b \\
&= \begin{cases} -\infty & c-\lambda+A^{\mathrm{T}}v \neq 0 \\ -v^{\mathrm{T}}b & c-\lambda+A^{\mathrm{T}}v = 0 \end{cases}
\end{aligned}$$

对偶问题，由于要极大，故不考虑负无穷部分

$$\begin{aligned}
\max_{\lambda,v} \quad & -v^{\mathrm{T}}b \\
\text{s.t.} \quad & c-\lambda+A^{\mathrm{T}}v = 0 \\
& \lambda \geq 0
\end{aligned}$$

逆过来求解

$$\begin{aligned}
\min \quad & b^{\mathrm{T}}v \\
\text{s.t.} \quad & A^{\mathrm{T}}v + c \geq 0
\end{aligned}$$

$$L(v,\lambda) = b^{\mathrm{T}}v - \lambda^{\mathrm{T}}(A^{\mathrm{T}}v + c)$$

$$
\begin{aligned}
g(\lambda) &= \inf_v L(v,\lambda) \\
&= \inf (b - A\lambda)^{\mathrm{T}} v - \lambda^{\mathrm{T}} c \\
&= \begin{cases} -\lambda^{\mathrm{T}} c & b - A\lambda = 0 \\ -\infty & b - A\lambda \neq 0 \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
\max \quad & -\lambda^{\mathrm{T}} c \\
\text{s.t.} \quad & b - A\lambda = 0 \\
& \lambda \geq 0
\end{aligned}
$$

对偶的对偶不一定回去，线性规划才满足

例 21.

$$
\begin{aligned}
\min \quad & x^{\mathrm{T}} w x \\
\text{s.t.} \quad & x_i = \pm 1, \quad i = 1, \ldots, n
\end{aligned}
$$

分析.

$$
L(x,v) = x^{\mathrm{T}} w x + \sum_{i=1}^{n} v_i (x_i^2 - 1)
$$

$$
\begin{aligned}
g(v) &= \inf_x L(x,v) \\
&= \inf_x x^{\mathrm{T}} w x + \sum_{i=1}^{n} v_i x_i^2 - \sum_{i=1}^{n} v_i \\
&= \inf_x x^{\mathrm{T}} \left(w + \operatorname{diag} v\right) x - \mathbb{1}^{\mathrm{T}} v \quad = \begin{cases} -\mathbb{1}^{\mathrm{T}} v & w + \operatorname{diag}(v) \succeq 0 \\ -\infty & \textit{otherwise} \end{cases}
\end{aligned}
$$

补充求梯度：$2(w + \operatorname{diag}(v))x = 0$

$$
\begin{aligned}
\max_v \quad & -\mathbb{1}^{\mathrm{T}} v \\
\text{s.t.} \quad & w + \operatorname{diag}(v) \succeq 0
\end{aligned}
$$

定义 15 (函数的共轭)**.** $f : \mathbb{R}^n \mapsto \mathbb{R}, f^\star(y) = \sup_{x \in \operatorname{dom} f} (y^{\mathrm{T}} x - f(x))$，几何意义即到不同斜率直线的距离最大值

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & Ax \leq b \\
& cx = d
\end{aligned}
$$

$$
\begin{aligned}
L(x,\lambda,v) &= f_0(x) + \lambda^{\mathrm{T}}(Ax-b) + v^{\mathrm{T}}(cx-d) \\
&= f_0(x) + (A^{\mathrm{T}}\lambda + c^{\mathrm{T}}v)^{\mathrm{T}}x - \lambda^{\mathrm{T}}b - v^{\mathrm{T}}d \\
g(\lambda,v) &= \inf_x f_0(x) + (A^{\mathrm{T}}\lambda + c^{\mathrm{T}}v)^{\mathrm{T}}x - \lambda^{\mathrm{T}}b - v^{\mathrm{T}}d \\
&= -\sup_x -(A^{\mathrm{T}}\lambda + c^{\mathrm{T}}v)^{\mathrm{T}}x - f_0 \\
&= -f_0^{\star}(-(A^{\mathrm{T}}\lambda + c^{\mathrm{T}}v)) - \lambda^{\mathrm{T}}b - v^{\mathrm{T}}d
\end{aligned}
$$

对偶间隙(duality gap)：$p^{\star} - d^{\star} \geq 0$

- 弱对偶：严格大于0
- 强对偶：对偶间隙为0

1. 对于非凸问题，**通常**$p^{\star} \neq d^{\star}$
2. 对于凸问题，若slater条件满足，$p^{\star} = d^{\star}$

定义 16 (相对内点(relative interior))**.**

$$
relint\, D = \{x \in D \mid B(x,r) \cap \operatorname{aff} D \subset v, \exists r > 0\}
$$

定理 3 (Slater条件)**.**

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

$\exists x \in relint\, D$使得$f_i(x) < 0, i = 1,\ldots,m, Ax = b$

例 22. *二次规划(QP)*

$$
\begin{aligned}
\min \quad & x^{\mathrm{T}}x \\
\text{s.t.} \quad & Ax = b
\end{aligned}
$$

Slater条件$\{x \mid Ax = b\}$*非空*

例 23. 二次约束二次规划*(QCQP)*

$$\begin{aligned} \min \quad & \frac{1}{2}x^{\mathrm{T}}P_0x + q_0^{\mathrm{T}} + r_0 \\ \text{s.t.} \quad & \frac{1}{2}x^{\mathrm{T}}P_ix + q_i^{\mathrm{T}}x + r_i \le 0, \quad i = 1, \ldots, m \end{aligned}$$

$P_0, \ldots, P_i$半正定

凸问题+Slater条件 $\implies p^\star = d^\star$，但有可能不满足Slater条件也依然强对偶

例 24.

$$\begin{aligned} \min \quad & x, x \in \mathbb{R} \\ \text{s.t.} \quad & xleq0 \\ & -x \le 0 \end{aligned}$$

分析.

$$L(x, \lambda_1, \lambda_2) = x + \lambda_1 x - \lambda_2 x = (1 + \lambda_1 - \lambda_2)x$$

$$g(\lambda_1, \lambda_2) = \inf_{x \in \mathbb{R}} (1 + \lambda_1 - \lambda_2)x = \begin{cases} 0 & 1 + \lambda_1 - \lambda_2 = 0 \\ -\infty & otherwise \end{cases}$$

$$\begin{aligned} \max_{\lambda_1, \lambda_2} \quad & 0 \\ \text{s.t.} \quad & 1 + \lambda_1 - \lambda_2 = 0 \end{aligned}$$

$$\implies p^\star = d^\star = 0$$

置信域问题

$$\begin{aligned} \min \quad & x^{\mathrm{T}}Ax + b^{\mathrm{T}}x \\ \text{s.t.} \quad & x^{\mathrm{T}}x \le 1 \\ & A \not\succeq 0 \end{aligned}$$

依然可以得到$p^\star = d^\star$

几何解释

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \end{aligned}$$

$$G = \{(f_1(x), f_0(x)) \mid x \in \mathcal{D}\}$$

$$g(\lambda) = \inf\{t + \lambda u \mid (u, t) \in G\}$$

$$L(x, \lambda) = f_0(x) + \lambda f_1(x)$$

$$g(\lambda) = \inf_{x \in \mathcal{D}} \{f_0(x) + \lambda f_1(x)\}$$

$$p^\star = \inf\{t \mid (u, t) \in G, u \leq 0\}$$

$$\lambda \geq 0, \max g(\lambda)$$

注意问题必须要有可行解

经济学解释：满足原材料约束下，利润最多价格$\lambda_i \geq 0$

$$g(\lambda) = \inf_x f_(x) + \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x) = \inf_x L(x, \lambda)$$

则$g(\lambda)$为对偶函数，市场$p^\star$损失最小（$g(\lambda) \leq p^\star$）

$$d^\star = \sup_{lambda \geq 0} g(\lambda)$$

市场平衡点，均衡市场$p^\star = d^\star$，最优/影子价格$\lambda^\star$

多目标优化解释

$$\begin{cases} \min f_0(x) & 1 \\ \min f_1(x) & \lambda_1 \\ \vdots & \vdots \\ \min f_m(x) & \lambda_m \end{cases}$$

$$\min_x f_0(x) + \lambda_1 f_1(x) + \cdots + \lambda_m f_m(x)$$

鞍点(saddle point)解释

$$f(w, z), w \in S_w, z \in S_z$$

极小极大不等式

$$\sup_{z \in S_z} \inf_{w \in S_w} f(wz) \leq \inf_{w \in S_w} \sup_{z \in S_z} f(w, z)$$

若有$(\tilde{w}, \tilde{z})$使得

$$(\tilde{w}, \tilde{z}) = \arg \max_{z \in S_z} \min_{w \in S_w} f(w, z)(\tilde{w}, \tilde{z}) \quad = \arg \min_{w \in S_w} \max_{z \in S_z} f(w, z)$$

则$(\tilde{w}, \tilde{z})$为鞍点

有下面不等式成立

$$f((\tilde{w}, z)) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z}), \forall z \in S_z, w \in S_w$$

即从一个方向望过去是最小，从另一个方向望过去是最大

$$
\begin{aligned}
L(x,\lambda) &= f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x) \\
\Longrightarrow \sup_{\lambda\geq 0} L(x,\lambda) &= \sup_{\lambda\geq 0}\{f_0(x)+\sum_{i=1}^{m}\lambda_i f_i(x)\} \\
&= \begin{cases} f_0(x) & f_i(x)\leq 0, i=1,\ldots,m \\ +\infty & \text{otherwise} \end{cases} \\
\Longrightarrow p^\star &= \inf_x\{f_0(x)\mid f_i(x)\leq 0, i=1,\ldots,m\} = \inf_x \sup_{\lambda\geq 0} L(x,\lambda)
\end{aligned}
$$

$$
d^\star = \sup_{\lambda\geq 0} g(\lambda) = \sup_{\lambda\geq 0}\inf_x L(x,\lambda) \implies p^\star \geq d^\star
$$

如果$L(x,\lambda)$有鞍点，则必有$p^\star = d^\star$

鞍点在无约束优化问题中是很糟糕的点（所有方向上梯度为0），但是有约束优化问题则是非常好的点

若$(\tilde{x},\tilde{\lambda})$为$L(x,\lambda)$鞍点$\iff p^\star = d^\star$且$\tilde{x},\tilde{\lambda}$为原对偶问题最优解 $\Longrightarrow$ 若为鞍点，$p^\star = d^\star$

$$
\sup_{\lambda\geq 0}\inf_x L(x,\lambda) = \inf_x \sup_{\lambda\geq 0} L(x,\lambda)
$$

已知$(\tilde{x},\tilde{\lambda})$为左边最优

$$
\begin{aligned}
\tilde{\lambda} &= \arg\max_{\lambda\geq 0}\inf_x L(x,\lambda) \\
\tilde{x} &= \arg\inf_x \sup_{\lambda\geq 0} L(x,\lambda)
\end{aligned}
$$

则$\tilde{\lambda}$对偶最优，$\tilde{x}$为原问题最优

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x)\leq 0, \quad i=1,\ldots,m
\end{aligned}
$$

定理 4. $(\tilde{x},\tilde{\lambda})$为拉格朗日函数鞍点$\iff p^\star = d^\star$，且$(\tilde{x},\tilde{\lambda})$为原对偶的最优解

分析. 右推左，$(\tilde{x},\tilde{\lambda})$原对偶可行

$$
f_i(\tilde{x})\leq 0, i=1,\ldots,m, \tilde{\lambda}\geq 0
$$

因$p^\star = d^\star$，有

$$
\begin{aligned}
f_0(\tilde{x}) &= g(\tilde{\lambda}) \\
&= \inf_x\{f_0(x)+\sum_{i=1}^{m}\tilde{\lambda}_i f_i(x)\} \\
&\leq f_0(\tilde{x})+\sum_{i=1}^{m}\tilde{\lambda}_i f_i(\tilde{x}) \\
&\leq f_0(\tilde{x})
\end{aligned}
$$

进而不等号都得为等号

1. $\inf_x L(x,\tilde{\lambda}) = L(\tilde{x},\tilde{\lambda})$
2. $f_0(\tilde{x}) = \sup_{\lambda\geq 0}\{f_0(\tilde{x}) + \sum_{i=1}^m \lambda_i f_i(\tilde{x})\} = \sup_{\lambda\geq 0} L(\tilde{x},\lambda)$

$$\Longrightarrow\ L(\tilde{x},\tilde{\lambda}) = \sup_{\lambda\geq 0} L(\tilde{x},\lambda)$$

$$\Longrightarrow\ (\tilde{x},\tilde{\lambda})\text{是}L(x,\lambda)\text{的鞍点}$$

一般优化问题的对偶理论

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \quad i = 1,\dots,m \\ & h_i(x) = 0, \quad i = 1,\dots,p \end{aligned}$$

不一定是凸问题，但$p^\star = d^\star$，最优解满足什么条件？

对偶问题

$$\begin{aligned} \max \quad & g(\lambda, v) \\ \text{s.t.} \quad & \lambda \geq 0 \end{aligned}$$

分析. 设$(x^\star,\lambda^\star,v^\star)$为原对偶最优解，则$(x^\star,\lambda^\star,v^\star)$为原对偶可行解

$$f_i(x^\star) \leq 0, i = 1,\dots,m, \quad h_i(x^\star) = 0, i = 1,\dots,p, \quad \lambda^\star \geq 0$$

$$\begin{aligned} p^\star = d^\star \implies f_0(x^\star) &= g(\lambda^\star, v^\star) \\ &= \inf_x\{f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p v_i^\star h_i(x)\} \\ &\leq f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star f_i(x^\star) + \sum_{i=1}^p v_i^\star h_i(x^\star) \\ &\leq f_0(x^\star) \end{aligned}$$

同上理，不等号全取等

1. $\lambda_i^\star f_i(x^\star) = 0, \forall i = 1,\dots,m$
2. $x^\star = \arg\min_x L(x,\lambda^\star,v^\star)$

若f_0, f_i, h_i均可微，则必要条件为

$$\left.\frac{\partial L(x,\lambda^\star,v^\star)}{\partial x}\right|_{x=x^\star} = 0$$

可微优化问题的KKT(Karush-Kuhn-Tucker)条件

- $f_i(x^\star) \leq 0, i = 1,\dots,m$ primal feasibility
- $h_i(x^\star) = 0, i = 1,\dots,p$ primal feasibility

- $\lambda^\star \geq 0$ dual feasibility
- $\lambda_i^\star f_i(x^\star) = 0, i = 1, \ldots, m$ complementarity slackness(对偶互斥条件)
- $\left.\dfrac{\partial L(x, \lambda^\star, v^\star)}{\partial x}\right|_{x=x^\star} = 0$ stablity

定理 5. 若原问题为凸，则KKT条件为**充要条件**

分析. 必要性已证，证明充分性

若$(\tilde{x}, \tilde{\lambda}, \tilde{v})$满足$KKT$条件 $\Longrightarrow$ $(\tilde{x}, \tilde{\lambda}, \tilde{v})$最优 $\tilde{x}$为原问题可行解，$(\tilde{\lambda}, \tilde{v})$为对偶问题可行解

证明思路：$g(\tilde{\lambda}, \tilde{v}) = f_0(\tilde{x})$

$L(x, \tilde{\lambda}, \tilde{v})$为$x$的凸函数，则$\tilde{x}$使$L(x, \tilde{\lambda}, \tilde{v})$最小

$$\begin{aligned}
g(\tilde{\lambda}, \tilde{v}) &= \inf_x L(x, \tilde{\lambda}, \tilde{v}) \\
&= L(\tilde{x}, \tilde{\lambda}, \tilde{v}) \\
&= f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i f_i(\tilde{x}) + \sum_{i=1}^p \tilde{v}_i h_i(\tilde{x}) \\
&= f_0(\tilde{x})
\end{aligned}$$

例 25 (Waterfilling算法)**.** 共n个信道*(channel)*

$$source \longleftrightarrow destination$$

$$\begin{aligned}
\min \quad & -\sum_{i=1}^n \log(\alpha_i + x_i) \\
\text{s.t.} \quad & x \geq 0 \\
& \mathbb{1}^{\mathrm{T}} = 1
\end{aligned}$$

分析. KKT条件

- $x^\star \geq 0$
- $\mathbb{1}^{\mathrm{T}} x^\star = 1$
- $\lambda^\star \geq 0$
- $x_i^\star \lambda_i^\star = 0, \forall i$

$$L(x, \lambda, v) = -\sum_{i=1}^n \log(\alpha_i + x_i) - \lambda^{\mathrm{T}} x + v(\mathbb{1}^{\mathrm{T}} x - 1)$$

$$\left(\frac{\partial L(x, \lambda, v)}{\partial x}\right)_i = -\frac{1}{\alpha_i + x_i} - \lambda_i + v$$

$$-\frac{1}{\alpha_i + x_i^\star} - \lambda_i^\star + v^\star = 0, \forall i$$

$$\Longrightarrow \; v^\star \frac{1}{\alpha_i + x_i^\star}, i = 1, \ldots, n$$

$$x_i^\star\left(v^\star - \frac{1}{\alpha_i + x_i^\star}\right) = 0, i = 1, \ldots, n$$

若$v^\star > \frac{1}{\alpha_i} \implies x_i^\star = 0$
若$v^\star < \frac{1}{\alpha_i}$

$$\frac{1}{\alpha_i} > v^\star \geq \frac{1}{\alpha_i + x_i^\star}$$

进而

$$x_i^\star > 0$$
$$v^\star = \frac{1}{\alpha_i + x_i^\star}$$
$$x_i^\star = \frac{1}{v^\star} - \alpha_i$$

$$\implies x_i^\star = \max\{0, \frac{1}{v^\star} - \alpha_i\}$$

结合$\sum_i x_i^\star = 1$，即注水算法