最优化算法

孙天阳

中国科学技术大学数学科学学院

tysun@mail.ustc.edu.cn

2025 年 4 月 30 日

目录

1 凸集

凸集分离定理, 从比较几何的观点看优化问题

2 锥与广义不等式

定义 2.1. *我们称集合* $C \subset \mathbb{R}^n$ *是一个锥, 如果对于任意* $x \in C$ *和任意* $t \geqslant 0$ *有* $tx \in C$.

直观来说, 锥是一些射线的并集. 如果锥 C 还是凸集, 我们称它为凸锥.

定义 2.2. *称凸锥* K *是适当锥, 如果 1.*K *是闭集 2.*$\operatorname{Int} K \neq \varnothing$*3. 若* $x, -x \in K$ *则* $x = 0$.

3 凸函数

在最优化领域, 经常会涉及对某个函数其中的一个变量取 inf 或 sup 的操作, 这导致函数的取值可能为无穷, 于是有了下面的定义

定义 3.1. 令 $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ 为广义实数空间, 则映射 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ 称为广义实值函数.

定义 3.2. 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 $f(x) < +\infty$ 且对任意的 $x \in \mathcal{X}$ 有 $f(x) > -\infty$, 那么称函数 f 关于集合 $\mathcal{X}$ 是适当的.

对于最优化问题 $\min f(x)$, 适当函数可以帮助我们去掉一些不感兴趣的函数, 从而在一个比较合理的函数类中考虑最优化问题. 规定适当函数的定义域为 $\operatorname{dom} f = \{x | f(x) < +\infty\}$.

定义 3.3. 设 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, 称 $C_\alpha = \{x | f(x) \leqslant \alpha\}$ 为 f 的 α-下水平集.

定义 3.4. 设 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, 称

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

为 f 的上方图. 若 $\operatorname{epi} f$ 为闭集, 则称 f 为闭函数.

定义 3.5. 设 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, 若对任意的 $x \in \mathbb{R}^n$ 有 $\liminf\limits_{y \to x} f(y) \geqslant f(x)$, 则称 $f(x)$ 为下半连续函数.

命题 3.1. 设 $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ 是适当函数, 则以下命题等价

(1) $f(x)$ 是下半连续的

(2) $f(x)$ 是闭函数

(3) $f(x)$ 的任意 α-下水平集都是闭集

凸函数, 如果想找 f(x) 由 f(y) 的下界估计，就在 f(y) 处画切线，让切线在 x 处取值；如果想找 f(x) 由 f(y) 的上界估计，就在 f(y) 处由 f(x) 的斜率画割线，在 x 处取值。

4 最优化问题解的存在性

考虑优化问题

$$\min f(x), \quad x \in \mathcal{X}$$

其中 $\mathcal{X} \subset \mathbb{R}^n$ 为可行域. 首先要考虑的是最优解的存在性, 其次是唯一性, 然后考虑如何求出最优解. 我们知道定义在紧集上的连续函数一定存在最大值点和最小值点, 但在许多实际问题中, 定义域可能不是紧的, 目标函数也不一定连续, 因此我们对此命题进行推广

定理 4.1. 设 $\mathcal{X}$ 是闭集, 函数 $f\colon \mathcal{X} \to (-\infty, +\infty]$ 适当且闭, 且以下条件中任意一个成立,

(1) $\operatorname{dom} f := \{x \in \mathcal{X}\colon f(x) < +\infty\}$ 是有界的.

(2) 存在一个常数 $\bar{\gamma}$ 使得下水平集 $C_{\bar{\gamma}} := \{x \in \mathcal{X}\colon f(x) \leqslant \bar{\gamma}\}$ 是非空且有界的.

(3) 对于任何满足 $\|x^k\| \to +\infty$ 的点列 $\left\{x^k\right\} \subset \mathcal{X}$, 都有 $\lim\limits_{k\to\infty} f(x^k) = +\infty$.

则函数 f 的最小值点集 $\{x \in \mathcal{X} | f(x) \leqslant f(y), \forall y \in X\}$ 非空且紧.

证明.

(2) 设 $t = \inf\limits_{x\in\mathcal{X}} f(x) = -\infty$, 存在点列 $\left\{x^k\right\}_{k=1}^{\infty} \subset C_{\bar{\gamma}}$ 使得 $\lim\limits_{k\to\infty} f(x^k) = -\infty$, 因为 $C_{\bar{\gamma}}$ 有界, 所以 $\left\{x^k\right\}_{k=1}^{\infty}$ 必有收敛子列, 不妨仍记为 $\left\{x^k\right\}_{k=1}^{\infty}$, 设它的序列极限是 x^*, 因为 $\mathcal{X}$ 是闭集, 所以 $x^* \in \mathcal{X}$. 易知 $(x^k, f(x^k))$ 收敛于 (x^*, t), 因为 $\operatorname{epi} f$ 是闭集, 所以 $(x^*, t) \in \operatorname{epi} f$, 所以 $f(x^*) \leqslant t = -\infty$, 这与 f 是适当的矛盾, 故 t 是有限值. 考察集合 $f^{-1}(t)$, 因为闭集的原像是闭集, 而 $f^{-1}(t) \subset C_{\bar{\gamma}}$ 是有界集, 所以是紧集.

(1) 由 f 的适当性与条件 (1) 显然能推出条件 (2).

(3) 假设某个下水平集无界, 由条件 (3) 显然能导出矛盾.

$\square$

5 无约束可微问题的最优性理论

6 对偶理论

在本节中, 我们介绍对偶理论, 他是一套宽泛而强大的工具, 其威力我们会逐渐体会. 考虑最一般的优化问题, 即带约束 + 非凸 + 非光滑,

$$\min_{x\in\mathbb{R}^n} f(x),\qquad \text{s.t. } c_i(x)\leqslant 0, i\in\mathcal{I},\quad c_i(x)=0, i\in\mathcal{E}.$$

构造它的 Lagrange 函数

$$L(x,\lambda,\nu)=f(x)+\sum_{i\in\mathcal{I}}\lambda_i c_i(x)+\sum_{i\in\mathcal{E}}\nu_i c_i(x),\quad \lambda_i\geqslant 0.$$

则我们有如下对偶问题

$$\max_{\lambda\geqslant 0,\nu} g(\lambda,\nu)=\max_{\lambda\geqslant 0,\nu}\inf_{x\in\mathbb{R}^n} L(x,\lambda,\nu).$$

任给可行域 $\mathcal{X}$ 中的 x_0 和对偶可行域中的 $\lambda_0\geqslant 0,\nu_0$, 我们有

$$g(\lambda_0,\nu_0)=\inf_{x\in\mathbb{R}^n} L(x,\lambda_0,\nu_0)\leqslant L(x_0,\lambda_0,\nu_0)\leqslant f(x_0)$$

右侧取下确界, 左侧取上确界, 则得到对偶问题的上确界 d^* 小于等于原问题的下确界 p^*. 我们称这件事为弱对偶原理, 称 p^*-d^* 为对偶间隙. 如果对于某个优化问题有 $d^*=p^*$, 我们称该优化问题满足强对偶原理. 本节接下来的内容按如下逻辑展开:1. 计算若干具体例子的对偶问题; 2. 证明如果优化问题满足强对偶原理, 那么最优解有一阶必要条件 KKT 条件; 3. 证明如果一个凸问题的某个点对满足 KKT 条件那么这个点就是最优点; 4. 证明满足 Slater 条件的凸问题满足强对偶原理, 并指出非凸问题也可能满足强对偶原理, 但例子比较复杂.

例 6.1.

$$\min\|x\|,\qquad \text{s.t. } Ax=b.$$

解. 拉格朗日函数为

$$L(x,\nu)=\|x\|-\nu^T(Ax-b)=\|x\|-\nu^TAx+\nu^Tb.$$

对偶函数为

$$g(\nu)=\inf_x(\|x\|-\nu^TAx+\nu^Tb)=-\sup_x(\nu^TAx-\|x\|)+\nu^Tb.$$

利用 $\|x\|$ 的共轭函数的结论

$$g(\nu)=\begin{cases}\nu^Tb & \|\nu^TA\|_*\leqslant 1\\ -\infty & \|\nu^TA\|_*>1\end{cases}.$$

□

例 6.2.

$$\min_x\|Ax-b\|$$

解. 将原问题改写为

$$\min_{x,y}\|y\|,\qquad \text{s.t. } y=Ax-b.$$

拉格朗日函数为

$$L(x,y,\nu)=\|y\|+\nu^T(y-Ax+b).$$

对偶函数为

$$g(\nu)=\inf_{x,y}(\|y\|+\nu^Ty-\nu^TAx+\nu^Tb)=\inf_{y}(\|y\|+\nu^Ty)+\inf_{x}(-\nu^TAx)+\nu^Tb.$$

$$g(\nu)=\begin{cases}\nu^Tb & \|\nu\|_*\leqslant 1,\nu^TA=0\\ -\infty & \text{otherwise}\end{cases}$$

所以对偶问题为

$$\max b^T\nu,\quad \text{s.t. } A^T\nu=0,\|\nu\|_*\leqslant 1.$$

□

假设某个优化问题满足强对偶原理即 $d^*=p^*$, 并且为有限值, 并且二者均可以取到, 即存在位于可行域中的 λ_0,ν_0,x_0 满足 $d^*=g(\lambda_0,\nu_0),p^*=f(x_0)$. 根据上面的不等式我们现在有

$$g(\lambda_0,\nu_0)=\inf_{x\in\mathbb{R}^n}L(x,\lambda_0,\nu_0)=L(x_0,\lambda_0,\nu_0)=f(x_0)$$

不等号变成了等号. 等号 $\inf\limits_{x\in\mathbb{R}^n}L(x,\lambda_0,\nu_0)=L(x_0,\lambda_0,\nu_0)$ 带来稳定性条件, 等号 $L(x_0,\lambda_0,\nu_0)=f(x_0)$ 带来互补松弛条件. λ_0,ν_0,x_0 位于相应的可行域带来原始可行性条件和对偶可行性条件.

命题 6.1. *考虑一个凸问题, 设 (x_0,λ_0,ν_0) 满足 KKT 条件, 则 (x_0,λ_0,ν_0) 是相应问题的最优值点.*

证明. 因为是凸问题, 所以稳定性条件现在能够说明 x_0 是 $L(x,\lambda_0,\nu_0)$ 的最小点,

$$f(x_0)=L(x_0,\lambda_0,\nu_0)\leqslant L(x,\lambda_0,\nu_0)\leqslant f(x).$$

而 $d^*\leqslant p^*$, 现在 $g(\lambda_0,\nu_0)=L(x_0,\lambda_0,\nu_0)=f(x_0)=p^*$, 说明 (λ_0,ν_0) 也是对偶问题的最大点. □

7 共轭函数

函数 f 的共轭函数定义为

$$f^*(y) = \sup_x (y^T x - f(x)).$$

我们来理解一下这个函数, 给定 y, 决定了关于 x 的以 y 为斜率的线性函数 $y^T x$, 比函数 $f(x)$ 高的最大距离, 就是 $f^*(y)$. 因此一个具有明显几何意义的结论是

$$f(x) \geqslant y^T x - f^*(y)$$

即函数 $f(x)$ 全部在直线 $y^T x - f^*(y)$ 的上方, 且这件事是刚好成立的. 所以或许可以将 $f^*(y)$ 诠释为 $f(x)$ 上以 y 为斜率的位置的切线的负截距. 所以变量 y 应该被理解为函数 $f(x)$ 的斜率. 回忆 Legendre 变换,

$$f^*(y) = y^T x - f(x),$$

其中将 x 通过隐式方程 $y = \nabla f(x)$ 视作关于 y 的函数.

例 7.1.

$$f(x) = \|x\|, \quad f^*(y) = \begin{cases} 0 & \|y\|_* \leqslant 1 \\ +\infty & \|y\|_* > 1 \end{cases}$$

证明.

$$f^*(y) = \sup_x (y^T x - \|x\|) = \max\left\{0, \sup_{\lambda > 0} \lambda \cdot \sup_{\|x\|=1} (y^T x - 1)\right\}.$$

□

8 切锥与线性化可行方向锥

考虑一个带约束的优化问题, 设 $\mathcal{X} \subset \mathbb{R}^n$ 是该问题的可行域.

定义 8.1. 称 $d \neq 0 \in \mathbb{R}^n$ 是 $x \in \mathcal{X}$ 的可行方向, 若存在 $\delta > 0$ 使得对任意 $0 < \lambda < \delta$ 有 $x + \lambda d \in \mathcal{X}$.

带约束的优化问题才需要讨论可行方向, 无约束的优化问题每个方向都是可行方向. 回忆无约束的可微的优化问题的极小值的一阶必要性条件是无下降方向. 那么类似的有

命题 8.1. 设 x^* 是带约束的可微的优化问题的极小值, 那么 x^* 的可行方向中无下降方向.

定义 8.2. 称 $d \neq 0 \in \mathbb{R}^n$ 是 $x \in \mathcal{X}$ 的切向量, 若存在 $\delta > 0$ 使得对任意 $0 < \lambda < \delta$ 有 $x + \lambda d \in \mathcal{X}$.

9 梯度下降法

梯度下降法在整个框架中属于求解无约束光滑问题的算法. 如果去掉光滑的要求, 它可以衍生为次梯度算法; 如果去掉无约束的要求, 它可以衍生为投影梯度法和条件梯度法. 从研究梯度下降法本身的角度来看, 我们关心三个问题, 一是下降方向的选取（梯度下降法就是下降方向选为负梯度, 但也可以替换成其他的得到别的算法), 二是步长的选择, 三是在对函数加额外条件下的收敛性.

步长的选择

第一种要介绍的选择方式叫做精确线搜索. 在确定好下降方向 d 之后, 可以考虑辅助函数 $\phi(\alpha) = f(x+\alpha d)$, 其中 $\alpha > 0$, 一个自然的想法是选择 $\alpha_0 = \operatorname{argmin}_{\alpha>0}\phi(\alpha)$, 这显然是最优的一个步长, 这也是为什么这种方法叫做精确线搜索. 但可想而知这种方法需要很大计算量, 因此实际中较少使用.

接下来介绍一类选择方式, 回溯法 + 某个准则. 准则用来判断一个步长是不是比较好的步长, 回溯法从一个较大的初始步长开始, 判断它是否在该准则下是好的步长, 若不是则以指数方式缩小.

收敛速度

命题 9.1. *设 f 为凸的 L-光滑函数, 最小值可以取到. 取步长为常数 α 且满足 $0<\alpha<1/L$, 则点列 $\{x^k\}$ 的函数值 $f(x^k)$ 收敛到最优值 f^*, 且在函数值的意义下收敛速度为 $O(1/k)$.*

证明. 因为要研究函数值如何收敛到最优值, 肯定要从 $f(x-\alpha\nabla f(x))$ 出发,

$$f(x-\alpha\nabla f(x)) \leqslant f(x) + \nabla f(x)\cdot(x-\alpha\nabla f(x)-x) + \frac{L}{2}\|x-\alpha\nabla f(x)-x\|^2$$

$$f(x-\alpha\nabla f(x)) \leqslant f(x) - \alpha(1-\frac{L\alpha}{2})\|\nabla f(x)\|^2$$

因为 $0<\alpha<1/L$, 所以 $1/2 < 1-L\alpha/2 < 1$, 所以

$$f(x-\alpha\nabla f(x)) \leqslant f(x) - \frac{\alpha}{2}\|\nabla f(x)\|^2$$

这也说明在这个步长的选择下, 迭代点的函数值确实是在减小的. 接下来想把右边跟 f^* 建立联系,

$$f(x-\alpha\nabla f(x)) \leqslant f^* + \nabla f(x)\cdot(x-x^*) - \frac{\alpha}{2}\|\nabla f(x)\|^2$$

$$\nabla f(x)\cdot(x-x^*) - \frac{\alpha}{2}\|\nabla f(x)\|^2 = \frac{1}{2\alpha}(\|x-x^*\|^2 - \|x-x^*-\alpha\nabla f(x)\|^2)$$

$$f(\tilde{x}) \leqslant f^* + \frac{1}{2\alpha}(\|x-x^*\|^2 - \|\tilde{x}-x^*\|^2)$$

这样一来就在右边构造出了我们比较喜欢的前一项减后一项的结构, 当然要累加.

$$\sum_{i=1}^{k}(f(x^i)-f^*) \leqslant \frac{1}{2\alpha}(\|x^0-x^*\|^2 - \|x^k-x^*\|^2) \leqslant \frac{1}{2\alpha}\|x^0-x^*\|^2$$

由于迭代值是非增的, 所以 $f(x^k)$ 是离 f^* 最近的,

$$f(x^k)-f^* \leqslant \frac{1}{k}\sum_{i=1}^{k}(f(x^i)-f^*) \leqslant \frac{1}{2k\alpha}\|x^0-x^*\|^2.$$

□

命题 9.2. 设 f 为 m-强凸的 L-光滑函数, 最小值可以取到. 取步长为常数 α 且满足 $0 < \alpha <\leqslant 2/(m+L)$, 则点列 $\left\{x^k\right\}$ 的收敛到 x^*, 且为 Q-线性收敛.

证明.

$$\|x^{k+1}-x^*\|^2 = \|x^k - \alpha\nabla f(x^k) - x^*\|^2 = \|x^k - x^*\|^2 - 2\alpha\nabla f(x^k)\cdot(x^k - x^*) + \alpha^2\|\nabla f(x^k)\|^2$$

如果这时候我们只用上 f 的 L-光滑性,

$$(\nabla f(x) - \nabla f(y))^T(x-y) \geqslant \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2 \Longrightarrow -2\alpha(\nabla f(x^k))^T(x - x^*) \leqslant -\frac{2\alpha}{L}\|\nabla f(x^k)\|^2$$

$$\|x^{k+1}-x^*\|^2 \leqslant \|x^k - x^*\|^2 + \alpha(\alpha - \frac{2}{L})\|\nabla f(x^k)\|^2$$

只要步长 α 取的足够小, 距离就是递减的.

$$\|x^{k+1}-x^*\|^2 \leqslant \|x^0 - x^*\|^2 + \alpha(\alpha - \frac{2}{L})(\|\nabla f(x^k)\|^2 + \cdots + \|\nabla f(x^0)\|^2).$$

如果我们还同时利用上 f 的 m-强凸性, 则函数 $g(x) = f(x) - m/2x^T x$ 是 $L-m$-光滑的,

$$(\nabla g(x) - \nabla g(y))^T(x-y) \geqslant \frac{1}{L-m}\|\nabla g(x) - \nabla g(y)\|^2, \quad \nabla g(x) = \nabla f(x) - mx$$

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

$$\frac{L+m}{L-m}(\nabla f(x) - \nabla f(y))^T(x-y) \geqslant \frac{1}{L-m}\|\nabla f(x) - \nabla f(y)\|^2 + \frac{mL}{L-m}\|x-y\|^2$$

$$(\nabla f(x) - \nabla f(y))^T(x-y) \geqslant \frac{1}{L+m}\|\nabla f(x) - \nabla f(y)\|^2 + \frac{mL}{L+m}\|x-y\|^2$$

从这个出发再应用上面的过程, 得到

$$\|x^{k+1}-x^*\|^2 \leqslant (1 - \frac{2\alpha mL}{L+m})\|x^k - x^*\|^2 + \alpha(\alpha - \frac{2}{L+m})\|\nabla f(x^k)\|^2$$

可以看到本质的区别发生在现在右侧 $\|x^k - x^*\|$ 的系数不再是 1 而是一个小于 1 的正数. 当 α 的取值不太大即小于等于 $2/(m+L)$ 时, 可以直接丢掉右侧的梯度项, 这样一来

$$\|x^{k+1}-x^*\|^2 \leqslant (1 - \frac{2\alpha mL}{L+m})\|x^k - x^*\|^2 \Longrightarrow \|x^k - x^*\|^2 \leqslant (1 - \frac{2\alpha mL}{L+m})^k\|x^0 - x^*\|^2$$

□

10 Barzilar-Borwein 方法

11 局部结构性条件

到目前为止, 我们在强凸性和梯度 Lipschitz 性的条件下建立了梯度法的线性收敛. 强凸性要求往往可以放宽, 不再假设问题是凸问题.

局部强凸性

规则性条件

Polyak-Lojasiewicz 条件

12 非凸问题的梯度方法

13 次梯度方法

14 投影梯度法

到简单集合的投影

例 14.1. $C = \left\{x \mid a^T x = b\right\}$, 则 $P_C(x) = x + \dfrac{b - a^T x}{\|a\|_2^2} a$

解. $P_C(x)$ 是点 x 到平面 $a^T x = b$ 的垂直投影, 该平面的法向量是 a, 所以可以设 $P_C(x) = x + \lambda a$, 其中 λ 是待定的常数. 利用 $P_C(x)$ 在平面上可以把常数确定出来. $\square$

15 第一次作业

问题 1 (判断集合是否是凸集)

1. 考虑这样点的集合，这些点离给定点 x_0 比离给定集合 S 中的任何点都更近，即集合 $\{x \mid \|x-x_0\|_2 \le \|x-y\|_2 \text{ for all } y \in S\}$, where $S \subseteq \mathbb{R}^n$.
2. 记 $n\times n$ 的对称矩阵集合为 $\mathbb{S}^n$，集合 $\{X \in \mathbb{S}^n \mid \lambda_{\min}(X) \ge 1\}$.

证明.

1. $$\{x \mid \|x-x_0\|_2 \le \|x-y\|_2 \text{ for all } y \in S\} = \bigcap_{y\in S}\{x \mid \|x-x_0\|_2 \le \|x-y\|_2\}$$
   而凸集的任意交是凸集, 所以问题中集合是凸集.
2. 设 $X_1, X_2 \in \{X \in \mathbb{S}^n \mid \lambda_{\min}(X) \ge 1\}$, 设 $\theta \in [0,1]$, 则
   $$\lambda_{\min}(\theta X_1 + (1-\theta)X_2) = \min_{\|x\|=1} x^T(\theta X_1 + (1-\theta)X_2)x \geqslant \theta\lambda_{\min}X_1 + (1-\theta)\lambda_{\min}X_2 \geqslant 1.$$

□

问题 2 (判断是否是凸函数)

1. 函数 $f(x) = \sum_{i=1}^{r} |x|_{[i]}$ 在 $\mathbb{R}^n$ 上定义，其中向量 $|x|$ 的分量满足 $|x|_i = |x_i|$（即，$|x|$ 是 x 的每个分量的绝对值），而 $|x|_{[i]}$ 是 $|x|$ 中第 i 大的分量。换句话说，$|x|_{[1]} \ge |x|_{[2]} \ge \ldots \ge |x|_{[n]}$ 是 x 的分量的绝对值按非增序排序。
2. 若 f, g 都是凸函数，并且都非递减，而且 f, g 函数值都是正的。那么他们的乘积函数 $h = fg$ 是否为凸函数？

证明.

1. 考虑两个向量 x, y 满足 $x_i \leqslant y_i$, 则 $x_{[i]} \leqslant y_{[i]}$, 下证明之. 对指标进行轮换, 不妨设 y 已经降序排列. 设 $x_{[i]} = x_j$, 分类讨论 $j < i$ 和 $j \geqslant i$
   - 当 $j \geqslant i$ 时 $x_{[i]} = x_j \leqslant y_j \leqslant y_i = y_{[i]}$.
   - 当 $j < i$ 时, 考虑 $x_{[1]}, \cdots, x_{[i-1]}$, 其中一定有一项 x_k 满足 $k \geqslant i$. 则
     $$x_{[i]} = x_j < x_k \leqslant y_k \leqslant y_i = y_{[i]}.$$

   $$\theta f(x) + (1-\theta)f(y) = \sum_{i=1}^{r} \theta|x|_{[i]} + (1-\theta)|y|_{[i]} \geqslant \sum_{i=1}^{r} |\theta x + (1-\theta)y|_{[i]} = f(\theta x + (1-\theta)y)$$

   中间不等号成立的原因是
   $$|\theta x + (1-\theta)y|_j \leqslant \theta|x|_j + (1-\theta)|y|_j$$

2. 按照定义验证 h 是凸函数.

$$
\begin{aligned}
h(\theta x+(1-\theta)y) &= f(\theta x+(1-\theta)y)g(\theta x+(1-\theta)y)\\
&\leqslant (\theta f(x)+(1-\theta)f(y))(\theta g(x)+(1-\theta)g(y))\\
&= \theta f(x)g(x)+(1-\theta)f(y)g(y)+\theta(1-\theta)(f(y)-f(x))(g(x)-g(y))\\
&\leqslant \theta f(x)g(x)+(1-\theta)f(y)g(y)\\
&= \theta h(x)+(1-\theta)h(y)
\end{aligned}
$$

□

问题 3
对于最大分量函数 $f(x)=\max\limits_{i=1,\cdots,n} x_i, x\in\mathbb{R}^n$, 证明其共轭函数为

$$
f^*(y)=\begin{cases}0 & y_i\geqslant 0, \sum\limits_i y_i=1\\ +\infty & \text{otherwise}\end{cases}.
$$

证明. 当 $y_i\geqslant 0, \sum\limits_i y_i=1$ 时, $y^Tx\leqslant f(x)$, 且等号在 $x=0$ 处可以取到, 所以 $f^*(y)=0$.

当存在 $y_i<0$ 时, 构造 $x_i<0$ 其余 $x_j=0$, 则 $y^Tx>0$ 且可趋于无穷但 $f(x)=0$.

当 $y_i\geqslant 0$ 且 $\sum\limits_i y_i>1$ 时, 构造 $x=(t,\cdots,t)$ 其中 $t>0$, 所以 $f^*(y)=+\infty$.

当 $y_i\geqslant 0$ 且 $\sum\limits_i y_i<1$ 时, 构造 $x=(t,\cdots,t)$ 其中 $t<0$, 所以 $f^*(y)=+\infty$. □

问题 4 对于分式线性问题

$$
\min f_0(x)=\frac{c^Tx+d}{e^Tx+f},\quad s.t.\ Gx\leqslant h, Ax=b
$$

其中 $\operatorname{dom} f_0(x)=\{x\mid e^Tx+f>0\}$. 证明该问题等价于一个线性规划问题

$$
\min c^Ty+dz,\quad s.t.\ Gy\leqslant hz, Ay=bz, e^Ty+fz=1, z\geqslant 0.
$$

证明. 在原问题中作变量代换 $x=y/z$, 其中要求 $z>0$, 这样一来目标函数和约束变成

$$
\min \frac{c^Ty+dz}{e^Ty+fz},\quad s.t.\ Gy\leqslant hz, Ay=bz, z>0
$$

定义域为 $e^Ty+fz>0$. 因为分子分母同时乘大于零的数并不影响目标函数的值, 所以可以仅仅考虑 $e^Ty+fz=1$ 的情形. 即将问题修改为

$$
\min c^Ty+dz,\quad s.t.\ Gy\leqslant hz, Ay=bz, e^Ty+fz=1, z>0.
$$

如何证明最小值不在 $z=0$ 时取到? □

问题 5 对于 $i=1,\cdots,m$, 令 B_i 是 $\mathbb{R}^n$ 中的球体, 它的球心和半径分别是 x_i 和 ρ_i. 我们希望找到 $B_i, i=1,\cdots,m$ 的最小外接球, 即找到一个球 B, 使得 B 包含所有 B_i, 并且 B 的半径最小. 将整个问题写为一个 SOCP 问题.

解. 设计变量为 x_0,ρ_0, 代表 B 的球心和半径, 目标函数为极小化 ρ_0, 约束为

$$
\|x_0-x_i\|_2\leqslant\rho_0-\rho_i,\quad i=1,\cdots,m.
$$

□