最优化算法作业 2

陈文轩

更新：March 25, 2025

1. 考虑以下问题：

$$\begin{aligned} \min_{x\in\mathbb{R}^2} \quad & x_1^2 + x_2^2 \\ s.t. \quad & (x_1-1)^2 + (x_2-1)^2 \leq 2 \\ & (x_1-1)^2 + (x_2+1)^2 \leq 2 \end{aligned}$$

其中 $x = (x_1, x_2)^\top \in \mathbb{R}^2$。

(a). 这是一个凸优化问题吗？

(b). 写出此问题的拉格朗日函数。使用 Slater 条件验证在这个问题中是否存在强对偶性。

(c). 写出这个优化问题的 KKT 条件。求出 KKT 点和最优点。

解

(a). 由于目标函数与约束函数均为二次函数，且 Hessian 矩阵为 $2I_2$，故是凸优化问题。

(b). 引入非负 Lagrange 乘子 $\lambda_1, \lambda_2 \geq 0$，可以得到问题的 Lagrange 函数为：

$$L(x,\lambda_1,\lambda_2) = x_1^2 + x_2^2 + \lambda_1((x_1-1)^2 + (x_2-1)^2 - 2) + \lambda_2(x_1-1)^2 + (x_2+1)^2 - 2$$

取 $x = (1,0)^\top$，显然这个 x 满足严格不等式，故 Slater 条件成立，即具有强对偶性。

(c). KKT 条件为：

$$\begin{cases} \nabla_x L(x,\lambda_1,\lambda_2) = 2(x_1 + (\lambda_1+\lambda_2)(x_1-1), x_2 + (\lambda_1+\lambda_2)(x_2-1)) = 0 \\ \lambda_1((x_1-1)^2 + (x_2-1)^2 - 2) = 0, \lambda_2((x_1-1)^2 + (x_2+1)^2 - 2) = 0 \\ (x_1-1)^2 + (x_2-1)^2 \leq 2, (x_1-1)^2 + (x_2+1)^2 \leq 2 \\ \lambda_1 \geq 0, \lambda_2 \geq 0 \end{cases}$$

解得 $x_1 = x_2 = \lambda_1 = \lambda_2 = 0$，即 $(0,0)^\top$ 是 KKT 点，也是最优点。

2. 考虑一个凸的分段线性最小化问题，变量为 $x \in \mathbb{R}^n$

$$\min \max_{i=1,\cdots,m} (a_i^\top x + b_i)$$

其中 $a_i \in \mathbb{R}^n, n_i \in \mathbb{R}$。

(a). 考虑原问题的如下等价问题2.1，推导2.1的对偶问题

$$\begin{aligned} \min \max_{i=1,\cdots,m} \quad & y_i \\ s.t. \quad & a_i^\top x + b_i \le y_i, i = 1, \cdots, m \end{aligned} \tag{2.1}$$

变量为 $x \in \mathbb{R}^n, y \in \mathbb{R}^m$。

(b). 假设我们通过平滑函数逼近目标函数, 逼近函数为：

$$f_0(x) = \log \sum_{i=1}^m \exp(a_i^\top x + b_i)$$

现在我们考虑无约束问题，即 $\min f_0(x)$。证明该问题的对偶问题如下：

$$\begin{aligned} \max \quad & b^\top z - \sum_{i=1}^m z_i \log z_i \\ s.t. \quad & A^\top z = 0, \mathbb{1}^\top z = 1, z \succeq 0. \end{aligned} \tag{2.2}$$

其中 $\mathbb{1}$ 表示全是 1 的向量。

(c). 设问题2.1的最优函数值是 p_1^*，问题2.2的最优函数值是 p_2^*，证明 $0 \le p_2^* - p_1^* \le \log m$。

解

(a). 先转化为线性规划问题：

$$\begin{aligned} \min \quad & t \\ s.t. \quad & a_i^\top x + b_i \le t, i = 1, \cdots, m \end{aligned}$$

Lagrange 函数为 $L(x,t,\lambda) = t + \sum_{i=1}^m \lambda_i(a_i^\top x + b_i - t), \lambda_i \ge 0$，目标是 $\max_\lambda \min_{x,t} L(x,t,\lambda)$。

$\nabla_t L(x,t,\lambda) = 1 - \sum_{i=1}^m \lambda_i = 0 \Rightarrow \mathbb{1}^\top \lambda = 1, \nabla_x L(x,t,\lambda) = \sum_{i=1}^m \lambda_i a_i = 0$，

此时 $L(x,y,\lambda) = \sum_{i=1}^m \lambda_i b_i$。因此对偶问题是：

$$\begin{aligned} \max \quad & \sum_{i=1}^m \lambda_i b_i \\ s.t. \quad & \mathbb{1}^\top \lambda = 1, \sum_{i=1}^m \lambda_i a_i = 0 \\ & \lambda_i \ge 0, i = 1, \cdots, m \end{aligned}$$

(b). 记 $A = (a_i, \cdots, a_m), z_j = \dfrac{\exp(a_j^\top x + b_j)}{\sum_{i=1}^m \exp(a_i^\top x + b_i)}$，则 $\sum_{j=1}^m z_j = 1, z_j \ge 0$。

此时 $f_0(x) = \log \sum_{i=1}^m \exp(a_i^\top x + b_i) = \sum_{i=1}^m z_i(a_i^\top x + b_i) - \sum_{i=1}^m z_i \log z_i := L(x,z)$。

$\nabla_x L(x,z) = \sum_{i=1}^m z_i a_i = 0 \Rightarrow A^\top z = 0$，此时 $L(x,z) = b^\top z - \sum_{i=1}^m z_i \log z_i$。

因此对偶问题是：

$$\begin{aligned} \max \quad & b^\top z - \sum_{i=1}^m z_i \log z_i \\ s.t. \quad & A^\top z = 0, \mathbb{1}^\top z = 1, z \succeq 0. \end{aligned}$$

(c). 记 $M=\max\limits_i(a_i^\top x+b_i)$，则 $\exp(M)\le\sum\limits_{i=1}^m\exp(a_i^\top x+b_i)\le m\exp(M)$。

取对数，即有 $M\le\log\sum\limits_{i=1}^m\exp(a_i^\top x+b_i)\le M+\log m,\forall x$。

对问题2.2的最优解 x_2^*，有 $p_2^*=\log\sum\limits_{i=1}^m\exp(a_i^\top x_2^*+b_i)$，由上有 $\max\limits_i(a_i^\top x_2^*+b_i)\le p_2^*$。

又 $p_1^*=\min\limits_x\max\limits_i(a_i^\top x+b_i)\le\max\limits_i(a_i^\top x_2^*+b_i)\le p_2^*$，故 $p_1\le p_2$。

对问题2.1的最优解 x_1^*，有 $\max\limits_i(a_i^\top x_1^*+b_i)=p_1^*$。由上有 $\log\sum\limits_{i=1}^m\exp(a_i^\top x_1^*+b_i)\le p_1^*+\log m$。又 $p_2^*=\min\limits_x\log\sum\limits_{i=1}^m\exp(a_i^\top x+b_i)\le p_1^*+\log m$，故 $p_2^*\le p_1^*+\log m$。

综上所述有 $0\le p_2^*-p_1^*\le\log m$。

3. 若 $f(x)=\|x\|$ 表示任意范数，$x\in\mathbb{R}^n$，证明次微分集合如下：

$$\partial f(x)=\{g\in\mathbb{R}^n:\|g\|_*\le 1,\langle g,x\rangle=\|x\|\},$$

其中 $\|g\|_*$ 表示对偶范数。

解

由次微分定义，有 $g\in\partial f(x)\Rightarrow f(y)\ge f(x)+\langle g,y-x\rangle,\forall y$，即 $\|y\|\ge\|x\|+\langle g,y-x\rangle,\forall y$。

分以下情况讨论：

- 若 $x=0$，则 $\langle g,x\rangle=\|x\|$ 显然成立；
- 对 $x\ne 0$，取 $y=0$，则 $0\ge\|x\|+\langle g,-x\rangle\Leftrightarrow\langle g,x\rangle\ge\|x\|$；

  取 $y=2x$，则 $2\|x\|\ge\|x\|+\langle g,x\rangle\Leftrightarrow\langle g,x\rangle\le\|x\|$。故 $\langle g,x\rangle=\|x\|$。

而 $g\in\partial f(x)\Rightarrow\|y\|\ge\langle g,y\rangle,\forall y$。令 $v=\dfrac{y}{\|y\|},t=\|y\|$，则条件变为：

$t\|v\|\ge t\langle g,v\rangle\Leftrightarrow\langle g,v\rangle\le 1,\forall\|v\|=1$。由对偶范数定义，这等价于 $\|g\|_*=\sup\limits_{\|v\|=1}\langle g,v\rangle\le 1$。

故 $\{g\in\mathbb{R}^n:\|g\|_*\le 1,\langle g,x\rangle=\|x\|\}\subset\partial f(x)$。

反之，若 $g\in\mathbb{R}^n$ 满足 $\|g\|_*\le 1,\langle g,x\rangle=\|x\|$，

则 $\|y\|\ge\|y\|\cdot\|g\|_*\ge\langle g,y\rangle=\langle g,y-x\rangle+\langle g,x\rangle=\|x\|+\langle g,y-x\rangle,\forall y\Rightarrow g\in\partial f(x)$。

故 $\partial f(x)\subset\{g\in\mathbb{R}^n:\|g\|_*\le 1,\langle g,x\rangle=\|x\|\}\Rightarrow\partial f(x)=\{g\in\mathbb{R}^n:\|g\|_*\le 1,\langle g,x\rangle=\|x\|\}$。

4. 对于 $y\in\mathbb{R}^m$，给定 $\mu-$ 强凸，$L-$ 光滑函数 $g(y)$，即 $\nabla g(y)$ 的 Lipchitz 常数为 L。若 $A\in\mathbb{R}^{m\times n},m\le n$ 是行满秩矩阵，证明：

(a). $f(x)=g(Ax)$ 是 $\bar{L}-$ 光滑的，其中

$$\bar{L}=L\|A\|^2$$

$\|A\|$ 表示矩阵 A 的谱范数。

(b). $f(x)=g(Ax)$ 满足规则性条件，即

$$\langle\nabla f(x),x-x_{\text{proj}}\rangle\ge\bar{\mu}\|x-x_{\text{proj}}\|^2$$

其中 x_{proj} 表示 x 到函数 $f(x)$ 最小值解集合的正交投影点，$\bar{\mu}=\mu\lambda_{\min}(AA^\top)$。

解

(a). 由于 $\nabla f(x) = A^\top \nabla g(Ax)$，有：

$$\begin{aligned}\|\nabla f(x_1) - \nabla f(x_2)\| &= \|A^\top \nabla g(Ax_1) - A^\top \nabla g(Ax_2)\| \le \|A^\top\| \cdot \|\nabla g(x_1) - \nabla g(x_2)\| \\ &\le \|A^\top\| \cdot L \cdot \|Ax_1 - Ax_2\| \le \|A^\top\| \cdot L \cdot \|A\| \cdot \|x_1 - x_2\| \\ &= \|A\|^2 \cdot L \cdot \|x_1 - x_2\| := \bar{L}\|x_1 - x_2\|\end{aligned}$$

故 $f(x)$ 是 $\bar{L}-$ 光滑的。

(b). 由 $g(y)\mu-$ 强凸，有 $g(y)$ 有唯一的最小值点 y^*，且 $\nabla f(y^*) = 0$。A 行满秩 $\Rightarrow AA^\top \succ 0$。此时 $\mu\|y - y^*\|^2 \le \langle \nabla g(y) - \nabla g(y^*), y - y^* \rangle = \langle \nabla g(Ax), Ax - y^* \rangle$。
又 $\forall v \in \mathrm{Row}(A), \|Av\|^2 \ge \lambda_{\min}(AA^\top)\|v\|^2$，$x - x_{\text{proj}} \in \mathrm{Col}(A^\top) = \mathrm{Row}(A)$

$$\begin{aligned}\langle \nabla f(x), x - x_{\text{proj}} \rangle &= \langle A^\top \nabla g(Ax), x - x_{\text{proj}} \rangle = \langle \nabla g(Ax), A(x - x_{\text{proj}}) \rangle \\ &= \langle \nabla g(Ax), Ax - y^*) \rangle \ge \mu\|Ax - y^*\|^2 = \mu\|A(x - x_{\text{proj}})\|^2 \\ &\ge \mu\lambda_{\min}(AA^\top)\|x - x_{\text{proj}}\|^2 := \bar{\mu}\|x - x_{\text{proj}}\|^2\end{aligned}$$

故 $\langle \nabla f(x), x - x_{\text{proj}} \rangle \ge \bar{\mu}\|x - x_{\text{proj}}\|^2$。

5. 考虑凸函数 $f(x)$ 的共轭函数

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

证明：

(a). 若 $x \in \partial f(y)$，则 $y \in \partial f^*(x)$；

(b). 若 $f(x)$ 是闭凸函数，利用 $f = f^{**}$ 证明 $x \in \partial f(y)$ 等价于 $y \in \partial f^*(x)$。

解

(a). 对 $x \in \partial f(y), \forall z, f(z) \ge f(y) + x^\top(z - y) \Rightarrow x^\top z - f(z) \le x^\top y - f(y), \forall z$
显然 $y = z$ 时取等，故 $f^*(x) = \sup_z (z^\top x - f(z)) = x^\top y - f(y)$。$\forall w$，
$f^*(x) + y^\top(w - x) = x^\top y - f(y) + y^\top(w - x) = w^\top y - f(y) \le \sup_z (w^\top z - f(z)) = f^*(w)$。
故 $\forall w, f^*(w) \ge f^*(x) + y^\top(w - x)$，即 $y \in \partial f^*(x)$。

(b). 由上，$x \in \partial f(y) \Rightarrow y \in \partial f^*(x)$，$y \in \partial f^*(x) \Rightarrow x \in \partial (f^*)^*(y) = \partial f^{**}(y) = \partial f(y)$。
故 $x \in \partial f(y) \Leftrightarrow y \in \partial f^*(x)$。