

最优化算法期末考试

陈士祥

2024.6.27

Problem 1. 解决如下问题: (1) 设 $A = \{x \in \mathbb{R}^n \mid \|x\|_2 \leq 1\}$, 证明其为凸集, 并且说明其为 $B = \{x \in \mathbb{R}^n \mid \|x\|_2 = 1\}$ 的凸包 (5 分)。

(2) 考虑优化问题:

$$\begin{aligned} \min \quad & \frac{1}{2} \|x - c\|_2^2 \\ \text{s.t.} \quad & \|x\|_2 = 1 \end{aligned} \tag{1.1}$$

证明其等价于 (5 分):

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & \|x\|_2 = 1 \end{aligned} \tag{1.2}$$

(3) 利用 (1) 中证明的凸集的性质, 对问题 (1.2) 进行松弛, 并说明松弛问题和原问题等价 (5 分)。

(4) 是否能对问题 (1.1) 做同样的松弛, 为什么 (5 分)。

Problem 2. 考虑优化问题: $\min_x \frac{1}{2} \|Ax - b\|_2^2$ 。

(1) 说明该问题为什么是凸问题 (2 分)。

(2) 估计该问题目标函数的 Lipschitz 常数 (3 分)。

(3) 使用 Armijo 线搜索得到步长相比常数步长有哪些优势, 请说明 (3 分)。

(4) 若要求解是稀疏的, 该优化问题应该如何改进以适配这一条件, 并在改过的问题上使用近似点梯度法求解, 写出迭代公式 (4 分)。

(5) 分别给出 (近似点) 梯度法和加速 (近似点) 梯度法的收敛速度 (4 分)。

Problem 3. 设 $X \in \mathbb{R}^{m \times k}, Y \in \mathbb{R}^{m \times n}$, 考虑如下优化问题:

$$\min_{A \in \mathbb{R}^{n \times r}, B \in \mathbb{R}^{k \times r}} \frac{1}{2} \|Y - XBA^T\|_F^2$$

用交替线性极小化法 (分块坐标下降法的一种) 求解该问题, 写出迭代公式, 并直接估计 A, B 的迭代步长 (即其对应的 Lipschitz 常数) (5 分)。

Problem 4. 我们有如下线性规划问题:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & Gx \geq h \\ & l \leq x \leq u \end{aligned} \tag{4.1}$$

(1) 利用拉格朗日函数，将问题 (4.1) 化为如下鞍点问题 (5 分)：

$$\min_{x \in C_1} \max_{y \in C_2} f(x) - g(y) + y^T Gx$$

(2) 使用 Chambolle-Pock 算法求解问题 (4.1)，写出显式迭代公式 (10 分)。

Problem 5. 一大堆背景，但反正最后是考虑这样一个优化问题：给定矩阵 $M \in \mathbb{R}^{n_1 \times n_2}$ ，解出满足如下问题的 $a \in \mathbb{R}^{n_1}, b \in \mathbb{R}^{n_2}$ ：

$$\min_{a,b} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \left(\ln \left(\frac{M_{ij}}{a_i b_j} \right) - M_{ij} + a_i b_j \right)$$

实际上这样解出来的 a, b 有 $M \approx ab^T$ ，即用低秩矩阵近似 M (10 分)。

Problem 6. 考虑如下线性规划问题：

$$\begin{aligned} \min \quad & -x_1 \\ \text{s.t.} \quad & x \in \Delta_2 \end{aligned} \tag{6.1}$$

其中 $\Delta_2 = \{(x_1, x_2)^T \in \mathbb{R}^2 \mid x_1 + x_2 = 1, x_1 \geq 0, x_2 \geq 0\}$ 是 2 维概率单纯形。

(1) 推导问题 (6.1) 的对偶问题，并说明该问题的强对偶原理是否成立，并说明理由 (5 分)。

(2) 用镜像梯度法求解该问题的迭代公式为：

$$x^{k+1} = \operatorname{argmin}_{y \in \Delta_2} \left\{ t \begin{pmatrix} -1 \\ 0 \end{pmatrix}^T y + D_h(y, x^k) \right\}$$

其中 $D_h(y, x) = h(y) - h(x) - \nabla h(x)^T (y - x)$ ，而 $h(x)_i = x_i \ln x_i$ 。

给定初始点 $x^0 = \begin{pmatrix} \frac{\sqrt{2}\epsilon}{2} e^{-tK} \\ 1 - \frac{\sqrt{2}\epsilon}{2} e^{-tK} \end{pmatrix}$ ，其中 $t > 0$ 是一个常数，而 K 为一正整数。

推导迭代点 x^k ($i = 1, 2, \dots, K$) 的表达式，并证明 (10 分)：

$$\left\| x^{k+1} - \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\|_2 \leq \epsilon$$

(3) 该收敛结论是否与 (2) 中收敛结果矛盾，给出判断并说明理由 (5 分)。

Problem 7. 考虑在机器学习中的一个优化问题：

$$\min_{\theta} l_{\text{val}}(\theta - \alpha \nabla l_{\text{train}}(\theta)) \tag{7.1}$$

其中 α 为给定的常数，而 $l_{\text{val}}(x) = \frac{1}{m} \sum_{i=1}^m \bar{l}_i(x)$ ， $l_{\text{train}}(x) = \frac{1}{n} \sum_{j=1}^n l_j(x)$ 为两个损失函数。且任意的 $\bar{l}_i(x), l_j(x)$ 都是凸可微函数。

(1) 将问题 (7.1) 写成如下双优化问题形式 (6 分)：

$$\begin{aligned} \min \quad & f(x, y^*(x)) \\ \text{s.t.} \quad & y^*(x) = \operatorname{argmin}_y g(x, y) \end{aligned}$$

(2) 使用随机梯度下降法求解问题 (7.1)，要求随机梯度步是真实梯度的无偏估计 (10 分)。