

2024 秋季 最优化理论与方法 I(最优化基础)

第 1 次作业

提交日期：2024 年 9 月 30 日 23:00

2024 年 9 月 26 日

提交说明：

1. 将自己的作业保存/扫描成 pdf 文件上传 (代码粘到文档中). 请查云盘上根据上课时间, 选择自己所在班的 excel 文件并找到自己名字的序号. 文件命名规则:

序号-学号-姓名-optihw*(*=1,2,3,4,5,6)

比如序号为 1, 学号为 SY2409107 的陈同学的第一次优化作业, 命名为 1-SY2409107-陈同学-optihw1.pdf.

2. 对照参考解答批阅自己的作业, 需要有红色笔批改痕迹. 并将自己的问题 (新解答-与参考解答不一致的解答、不理解并需要助教解答的地方) 标注出来, 由助教逐一解答. 此外, 务必标注清楚哪些是自己做的; 哪些是自己看过参考解答, 理解后重新写的解答 (可用异色笔).

3. 学院路 1 班 (周三和周五 3-4 节) 同学的第一次作业提交链接如下:

<https://bhpan.buaa.edu.cn/link/AA785C7DAC2CAF44539F84D827645773D4>

文件夹名: 最优化HW_1

有效期限: 2024-09-30 23:59

学院路 2 班 (周三和周五 1-2 节) 同学的第一次作业提交链接如下:

<https://bhpan.buaa.edu.cn/link/AA0851E8FADE1D4307BC25E2F114C034CF>

文件夹名: 最优化HW_1

有效期限: 2024-09-30 23:59

1. (复习并熟悉多元函数梯度和海塞矩阵的概念, 特别注意 (d), 其是非线性最小二乘问题的目标函数, 所得结果是设计求解非线性最小二乘问题算法的基础) 确定下列 n 元函数的梯度向量和 Hesse 阵:

- (a) $\mathbf{a}^T \mathbf{x}$: \mathbf{a} 是已知向量;
- (b) $\mathbf{x}^T \mathbf{A} \mathbf{x}$: \mathbf{A} 是非对称已知矩阵;
- (c) $\frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbf{b}^T \mathbf{x}$: \mathbf{A} 是已知的对称矩阵, \mathbf{b} 是已知向量;
- (d) $\mathbf{r}(\mathbf{x})^T \mathbf{r}(\mathbf{x})$: $\mathbf{r}(\mathbf{x}) = (r_1(\mathbf{x}), \dots, r_m(\mathbf{x}))^T$ 是依赖于 \mathbf{x} 的 m 维向量, 记 $\nabla \mathbf{r}^T$ 为 \mathbf{A}^T , 它一般不是常量.

解:

- (a) $\nabla f = \mathbf{a}$
 $\nabla^2 f = 0$
- (b) $\nabla f = (\mathbf{A} + \mathbf{A}^T) \mathbf{x}$
 $\nabla^2 f = (\mathbf{A} + \mathbf{A}^T)$
- (c) $\nabla f = \mathbf{A} \mathbf{x} - \mathbf{b}$
 $\nabla^2 f = \mathbf{A}$
- (d)

2. (优化问题解的存在性) 一个函数 $f: \mathbb{R}^n \rightarrow \mathbb{R}$ 被称为**强制的** (coercive), 如果 $\lim_{\|\mathbf{x}\| \rightarrow +\infty} f(\mathbf{x}) = +\infty$. 证明强制的连续函数 f 在非空闭集 S 上能取到最小值, 即它在 S 上至少存在一个全局极小点.

3. (本题将第一节课中优化表述技巧延申到非线性规划. 一般地, 在应用中可用来将目标函数非光滑的问题等价表述成光滑的约束最优化问题.) 设 $\mathbf{r}: \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是光滑的向量值函数 (每个分量函数是 n 元的, 其一阶偏导数存在并连续). 考虑 $\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x})$, 其中

- (a) $f(\mathbf{x}) = \|\mathbf{r}(\mathbf{x})\|_\infty$;
- (b) $f(\mathbf{x}) = \max\{r_i(\mathbf{x}), i = 1, \dots, m\}$;
- (c) $f(\mathbf{x}) = \|\mathbf{r}(\mathbf{x})\|_1$.

请将这些 (通常是非光滑的) 问题重新表述成光滑 (目标函数和约束函数的一阶偏导数存在并且连续) 的优化问题.

4. 指出下面哪些集合是凸集. 是凸集的说明理由, 不是凸集的也需要说明理由 (凸集的定义、保凸运算、凸函数的上镜图是凸集、凸函数的下水平集是凸集等). 记以下题目所给集合为 S .

- (a) $\{\mathbf{x} \in \mathbb{R}^2 : x_1 + i^2 x_2 \leq 1, i = 1, \dots, 10\}$
- (b) $\{\mathbf{x} \in \mathbb{R}^2 : x_1^2 + 2ix_1x_2 + i^2x_2^2 \leq 1, i = 1, \dots, 10\}$
- (c) $\{\mathbf{x} \in \mathbb{R}^2 : x_1^2 + ix_1x_2 + i^2x_2^2 \leq 1, i = 1, \dots, 10\}$
- (d) $\{\mathbf{x} \in \mathbb{R}^2 : x_1^2 + 5x_1x_2 + 4x_2^2 \leq 1\}$
- (e) $\{\mathbf{x} \in \mathbb{R}^{10} : x_1^2 + 2x_2^2 + 3x_3^2 + \dots + 10x_{10}^2 \leq 2004x_1 - 2003x_2 + 2002x_3 - \dots + 1996x_9 - 1995x_{10}\}$
- (f) $\{\mathbf{x} \in \mathbb{R}^2 : \exp\{x_1\} \leq x_2\}$
- (j) $\{\mathbf{x} \in \mathbb{R}^2 : \exp\{x_1\} \geq x_2\}$
- (h) $\{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 = 1\}$
- (i) $\{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \leq 1\}$
- (j) $\{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \geq 1\}$
- (k) $\{\mathbf{x} \in \mathbb{R}^n : \max_{i=1, \dots, n} x_i \leq 1\}$
- (l) $\{\mathbf{x} \in \mathbb{R}^n : \max_{i=1, \dots, n} x_i \geq 1\}$
- (m) $\{\mathbf{x} \in \mathbb{R}^n : \max_{i=1, \dots, n} x_i = 1\}$
- (n) $\{\mathbf{x} \in \mathbb{R}^n : \min_{i=1, \dots, n} x_i \leq 1\}$
- (o) $\{\mathbf{x} \in \mathbb{R}^n : \min_{i=1, \dots, n} x_i \geq 1\}$
- (p) $\{\mathbf{x} \in \mathbb{R}^n : \min_{i=1, \dots, n} x_i = 1\}$

5. 判断下面的函数在指定区域上是否是凸函数 (凸函数的定义、函数凸当且仅当上镜图凸、保凸运算、二阶可微函数凸当且仅当任一点的 Hesse 矩阵半正定), 并给出理由:

- (a) $f(x) \equiv 1$ 在 \mathbb{R} 上.
- (b) $f(x) = x$ 在 \mathbb{R} 上.
- (c) $f(x) = |x|$ 在 \mathbb{R} 上.
- (d) $f(x) = -|x|$ 在 \mathbb{R} 上.
- (e) $f(x) = -|x|$ 在 $\mathbb{R}_+ = \{x \in \mathbb{R} : x \geq 0\}$ 上.
- (f) $\exp\{x\}$ 在 \mathbb{R} 上.
- (g) $\exp\{x^2\}$ 在 \mathbb{R} 上.
- (h) $\exp\{-x^2\}$ 在 \mathbb{R} 上.
- (i) $\exp\{-x^2\}$ 在 $\{x \in \mathbb{R} : x \geq 100\}$ 上.
- (j) $\frac{x^2}{y}$ 在 $Q = \{(x, y) \in \mathbb{R}^2 : y > 0\}$ 上.

(l) $\ln(\exp\{x\} + \exp\{y\})$ 在二维平面上.

6. (凸函数及其性质) 完成以下问题:

(a) 证明熵函数

$$f(\mathbf{x}) = - \sum_{i=1}^n x_i \log(x_i)$$

在集合 $S = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x_i > 0, i = 1, \dots, n\}$ 上是凹的.

(b) 设 $f \in C^1$, S 是凸集. 证明 f 在 S 上是凸的当且仅当对所有的 $\mathbf{x}, \mathbf{y} \in S$ 都有

$$(\nabla f(\mathbf{x}) - \nabla f(\mathbf{y}))^T (\mathbf{x} - \mathbf{y}) \geq 0$$

成立. 称这个性质为 ∇f 的单调性.

(c) 给出一个严格凸函数的例子, 它取不到最小值, 即不存在极小点.

(d) (选做题) 证明凸函数在有界多面体 S 上的最大值必定可以在多面体的某极点取到. 提示: 可以使用事实: 一个有界多面体可以表示为其极点集合 $\text{ext } S$ 的凸包, 即 $S = \text{conv}(\text{ext } S)$.

7. (使用 CVX 求解优化问题, 不用交) 复合凸规划 (disciplined convex programming, DCP) 是一个系统, 用于由已知曲率 (函数是常数、仿射的、凸的、凹的) 的基函数库构造数学表达式. 凸优化建模语言 CVX, CVXPY, Convex.jl 和 CVXR 都使用 DCP 以确保特定优化问题是凸的. CVX 是一个出色的用于复合凸规划的系统, 虽然它不总是最快的工具, 但它的适用性很广, 因此是一个很好的工具. 在此练习中, 设置 CVX 环境并求解一个凸优化问题.

一般来说, 对于本课程的作业, 编程问题解答应包括图表以及回答问题所需的解释. 此外, 完整代码应作为作业文件的附件提交.

CVX 的各种变体适用于主要的数值编程语言. 这些变体在语法和功能上有一些微小的差异, 但基本上提供相同的功能. 根据需求下载 CVX 变体:

- Matlab: <http://cvxr.com/cvx/>
- Python: <http://www.cvxpy.org/>
- R: <https://cvxr.rbind.io>
- Julia: <https://github.com/JuliaOpt/Convex.jl>

并查阅文档以了解基本功能. 确保可以求解: 已知任意向量 \mathbf{y} 和矩阵 \mathbf{X} 的最小二乘问题 $\min_{\beta} \|\mathbf{y} - \mathbf{X}\beta\|_2^2$. 提示: 这里可通过与闭式解 $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ 比较来检查 CVX 求解结果的正确性.

- (a) 已知标签 $\mathbf{y} \in \{-1, 1\}^n$ 和特征矩阵 $\mathbf{X} \in \mathbb{R}^{n \times p}$, 其行向量为 $\mathbf{x}_1, \dots, \mathbf{x}_n$, 支持向量机 (support vector machine, SVM) 问题是

$$\begin{aligned} & \underset{\boldsymbol{\beta}, \beta_0, \boldsymbol{\xi}}{\text{minimize}} && \frac{1}{2} \|\boldsymbol{\beta}\|_2^2 + C \sum_{i=1}^n \xi_i \\ & \text{subject to} && \xi_i \geq 0, \quad i = 1, \dots, n \\ & && y_i(\mathbf{x}_i^T \boldsymbol{\beta} + \beta_0) \geq 1 - \xi_i, \quad i = 1, \dots, n. \end{aligned}$$

- (i) 加载训练数据 xy_train.csv. 这是 $n = 200$ 行 3 列的矩阵. 前两列是 $p = 2$ 的特征向量, 第三列是标签. 使用 CVX, 在 $C = 1$ 的情况下求解 SVM 问题. 给出最优目标函数值, 以及最优系数向量 $\boldsymbol{\beta} \in \mathbb{R}^2$ 和截距 $\beta_0 \in \mathbb{R}$.
- (ii) SVM 问题的最优解中的 $\boldsymbol{\beta}$ 和 β_0 定义了一个超平面

$$\beta_0 + \boldsymbol{\beta}^T \mathbf{x} = 0, \quad (1)$$

这是 SVM 分类器的决策边界. 绘制训练数据, 并用不同颜色标记两类点. 在上面绘制决策边界.

- (iii) 现在定义 $\tilde{\mathbf{X}} \in \mathbb{R}^{n \times p}$, 其行向量 $\tilde{\mathbf{x}}_i = y_i \mathbf{x}_i$, $i = 1, \dots, n$, 并使用 CVX 求解

$$\begin{aligned} & \underset{\boldsymbol{\alpha} \in \mathbb{R}^n}{\text{maximize}} && -\frac{1}{2} \boldsymbol{\alpha}^T \tilde{\mathbf{X}} \tilde{\mathbf{X}}^T \boldsymbol{\alpha} + \sum_{i=1}^n \alpha_i \\ & \text{subject to} && 0 \leq \alpha_i \leq C, \quad i = 1, \dots, n, \quad \mathbf{y}^T \boldsymbol{\alpha} = 0. \end{aligned}$$

给出最优函数值; 它应与 (i) 中的最优值对应. 并给出在最优点 $\boldsymbol{\alpha}$ 下的 $\tilde{\mathbf{X}}^T \boldsymbol{\alpha}$; 这应与 (i) 中最优的 $\boldsymbol{\beta}$ 对应. 注意: 这不是巧合, 而是对偶性的一个例子, 后续课程将会进行详细的研究.

- (iv) 对于惩罚参数 $C = 2^a$, a 的值从 -5 到 5 (间隔取 1), 对于每个惩罚参数 C 的取值, 求解 SVM 问题, 得到决策边界(1), 并据此计算 xy_test.csv 中的测试数据的错分率 (预测结果 $\hat{y}_i = \text{sign}(\beta_0 + \boldsymbol{\beta}^T \mathbf{x}_i)$ 与已知标签不一致的频率). 绘制错分率 (y 轴) 与 C (x 轴, 你可能需要使用对数刻度) 的图形.
- (b) 复合凸规划 (disciplined convex programming, DCP) 是一个系统, 用于从已知曲率 (函数是常数、仿射的、凸的、凹的) 的基函数库构造数学表达式. 凸优化建模语言 CVX, CVXPY, Convex.jl, 和 CVXR 都使用 DCP 以确保特定的优化问题是凸的.

DCP 是 CVX 的基础语言. 本质上, 凸表达式的解析树中的每个节点都标有曲率 (凸、凹、仿射、常数) 和符号 (正、负) 的属性, 允许对

整个表达式的凸性进行推理. 网站 <http://dcp.stanford.edu/> 提供简单表达式的可视化和分析.

通常情况下, 以 DCP 形式表述问题是自然的, 但在某些情况下需要进行处理以构建满足规则的表达式. 对于下面的每组数学表达式, 首先简要解释为什么每个表达式定义了一个凸集. 然后, 给出一个等价的 DCP 表达式, 以及为什么 DCP 表达式等同于原表达式的简要解释. DCP 表达式应该以一种形式给出, 该形式在 <http://dcp.stanford.edu/analyzer> 上通过分析的检验 (表达式框左侧是绿勾). 注意: 这个问题实际上可以帮助您更好地理解各种保凸性的函数的复合规则.

以 i) 为例, $\|(x, y, z)\|_2^2 \leq 1$ 表示以原点为中心, 半径为 1 的 2 范数球 (Euclidean 球), 所以是凸的.

DCP 表达式: `pow(norm2(x,y,z),2)<=1.`

- i) $\|(x, y, z)\|_2^2 \leq 1.$
- ii) $\sqrt{x^2 + 1} \leq 3x + y$
- iii) $1/x + 2/y \leq 5, x > 0, y > 0$
- iv) $(x + y)^2 / \sqrt{y} \leq x - y + 5, y > 0$
- v) $(x + z)y \geq 1, x + z \geq 0, y \geq 0$
- vi) $\|(x + 2y, x - y)\|_2 = 0$
- vii) $x\sqrt{y} \geq 1, x \geq 0, y \geq 0$
- viii) $\log(e^{y-1} + e^{x/2}) \leq -e^x$