最优化计算方法（简化版）
参考答案

丁思哲　　邓展望　　李天佑　　陈铖　　谢中林　　俞建江
刘浩洋　　文再文

版本：v1.12 （更新于 2022.10.20）

教材链接：https://bicmr.pku.edu.cn/~wenzw/optbook.html

目录

第一章　最优化简介

1.1 考虑稀疏优化问题，我们已经直观地讨论了在 ℓ_0，ℓ_1，ℓ_2 三种范数下问题的解的可能形式．针对一般的 ℓ_p “范数”：

$$\|x\|_p \overset{\text{def}}{=\!=} (\sum_{i=1}^{n} |x|^p)^{1/p}, \quad 0 < p < 2,$$

我们考虑优化问题：

$$\begin{aligned} \min \quad & \|x\|_p, \\ \text{s.t.} \quad & Ax = b. \end{aligned}$$

试着用几何直观的方式（类似于图 1.2）来说明当 $p \in (0, 2)$ 取何值时，该优化问题的解可能具有稀疏性.

解 (丁思哲). 在 $\mathbb{R}^2$ 空间中，不同 p 的范数球情形如图 1.1 所示. □

1.2 给定一个函数 $f(x) : \mathbb{R}^n \to \mathbb{R}$ 及其一个局部最优点 x^*，则该点沿任何方向 $d \in \mathbb{R}^n$ 也是局部最优的，即 0 为函数 $\phi(\alpha) \overset{\text{def}}{=\!=} f(x^* + \alpha d)$ 的一个局部最优解．反之，如果 x^* 沿任何方向 $d \in \mathbb{R}^n$ 都是局部最优解，则 x^* 是否为 $f(x)$ 的一个局部最优解？若是，请给出证明；若不是，请给出反例.

解 (俞建江). 反例：考虑极坐标表示的函数

$$f(r, \theta) = \begin{cases} r \sin(\dfrac{r}{\theta}), & \theta \in (0, 2\pi), \\ 0, & \theta = 0. \end{cases}$$ □

1.3 考虑函数 $f(x) = x_1^2 + x_2^2$，$x = (x_1, x_2) \in \mathbb{R}^2$，以及迭代点列 $x^k = (1 + \dfrac{1}{2^k})(\cos k, \sin k)^{\mathrm{T}}, k = 1, 2, \cdots$，请说明

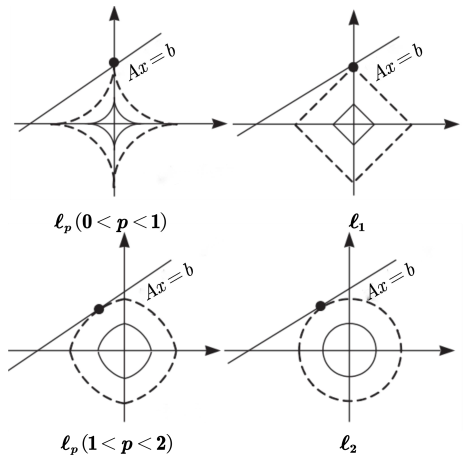

图 1.1 ℓ_p 范数优化问题的求解

(a) $\{f(x^{k+1})\}$ 是否收敛？若收敛，给出 Q-收敛速度；

(b) $\{x^{k+1}\}$ 是否收敛？若收敛，给出 Q-收敛速度.

解 (俞建江).

(a) 该点列 Q-线性收敛.

(b) 该点列不收敛.

□

第二章　基础知识

2.1 证明：矩阵 A 的 2 范数等于其最大奇异值，即

$$\sigma_1(A) = \max_{\|x\|_2=1} \|Ax\|_2.$$

解 (俞建江，丁思哲). 考虑

$$\|Ax\|_2 = \sqrt{(Ax)^{\mathrm{T}}(Ax)} = \sqrt{x^{\mathrm{T}}(A^{\mathrm{T}}A)\,x}$$

且 $A^{\mathrm{T}}A$ 是实对称矩阵. □

2.2 证明如下有关矩阵范数的不等式：

(a) $\|AB\|_F \leqslant \|A\|_2\|B\|_F$,

(b) $|\langle A, B\rangle| \leqslant \|A\|_2\|B\|_*$.

解 (俞建江).

(a) 对 $A^{\mathrm{T}}A$ 正交对角化，利用 F 范数的定义证明.

(b) 对 B 作 SVD 分解，利用矩阵内积的定义证明.

□

2.3 假设 A 和 B 均为半正定矩阵，求证：$\langle A, B\rangle \geqslant 0$. 提示：利用对称矩阵的特征值分解.

解 (俞建江). 利用矩阵内积的定义证明. □

2.4 计算下列矩阵变量函数的导数.

(a) $f(X)=a^{\mathrm{T}}Xb$，这里 $X\in\mathbb{R}^{m\times n}$，$a\in\mathbb{R}^m,b\in\mathbb{R}^n$ 为给定的向量；

(b) $f(X)=\mathrm{Tr}(X^{\mathrm{T}}AX)$，其中 $X\in\mathbb{R}^{m\times n}$ 是长方形矩阵，A 是方阵（但不一定对称）；

解（俞建江，丁思哲）.

(a) $\nabla f(X)=ab^{\mathrm{T}}$.

(b) $\nabla f(X)=(A+A^{\mathrm{T}})X$.

(c) $\nabla f(X)=X^{-\mathrm{T}}$. □

2.5 考虑二次不等式

$$x^{\mathrm{T}}Ax+b^{\mathrm{T}}x+c\leqslant 0,$$

其中 A 为 n 阶对称矩阵，设 C 为上述不等式的解集.

(a) 证明：当 A 正定时，C 为凸集；

(b) 设 C' 是 C 和超平面 $g^{\mathrm{T}}x+h=0$ 的交集 $(g\neq 0)$，若存在 $\lambda\in\mathbb{R}$，使得 $A+\lambda gg^{\mathrm{T}}$ 半正定，证明：C' 为凸集.

解（俞建江）.

(a) 考察 $f(x)=x^{\mathrm{T}}Ax+b^{\mathrm{T}}x+c$ 的凸性.

(b) 利用上一小问的结论，注意点在 C' 上的条件. □

2.6 利用凸函数二阶条件证明如下结论：

(a) ln-sum-exp 函数：$f(x)=\ln\sum\limits_{k=1}^{n}\exp x_k$ 是凸函数；

(b) 几何平均：$f(x)=(\prod\limits_{k=1}^{n}x_k)^{1/n}$ $(x\in\mathbb{R}^n_{++})$ 是凹函数；

(c) 设 $f(x)=\left(\sum\limits_{i=1}^{n}x_i^p\right)^{1/p}$，其中 $p\in(0,1)$，定义域为 $x>0$，则 $f(x)$ 是凹函数.

解（俞建江）． 求海瑟矩阵后，证明矩阵半正定即可. □

2.7 证明定理 2.12.

解 (俞建江). 充分性证明：考虑凸函数的定义和上方集的凸性.
必要性证明：利用凸集定义. □

2.8 求下列函数的共轭函数：

(a) 负熵：$\sum_{i=1}^{n} x_i \ln x_i$;

(b) 矩阵对数：$f(x) = -\ln\det(X)$;

(c) 最大值函数：$f(x) = \max_i x_i$;

(d) 二次锥上的对数函数：$f(x,t) = -\ln(t^2 - x^{\mathrm{T}}x)$，注意这里 f 的自变量是 (x,t).

解 (丁思哲).

(a) $f^*(y) = \sum_{i=1}^{n} e^{y_i - 1}$.

(b) $f^*(Y) = -n - \ln\det(-Y)$, 其中 Y 的定义域是 $\{Y \in \mathbb{R}^{n\times n} \mid \det(-Y) > 0\}$.

(c) 若 $\|y\|_1 \leqslant 1$ 且 $y \geqslant 0$，$f^*(y) = 0$.
若不然，$f^*(y)$ 不存在.

(d) $f^*(y,q) = -2 + \ln(\frac{4}{q^2 - y^{\mathrm{T}}y})$, 定义域为 $\{(y,q) \mid q^2 - y^{\mathrm{T}}y > 0\}$. □

2.9 求下列函数的一个次梯度：

$$f(x) = \|Ax - b\|_2 + \|x\|_2.$$

解 (俞建江). 一个次梯度为：

$$g(x) = \begin{cases} \dfrac{A^{\mathrm{T}}(Ax-b)}{\|Ax-b\|_2} + \dfrac{x}{\|x\|_2}, & Ax - b \neq 0,\ x \neq 0, \\ \dfrac{A^{\mathrm{T}}(Ax-b)}{\|Ax-b\|_2}, & Ax - b \neq 0,\ x = 0, \\ \dfrac{x}{\|x\|_2}, & Ax - b = 0,\ x \neq 0, \\ 0, & Ax - b = 0,\ x = 0. \end{cases}$$

□

2.10 设 $f(x)$ 为 m-强凸函数，求证：对于任意的 $x \in \mathbf{int\,dom}\, f$，

$$f(x) - \inf_{y \in \mathbf{dom}\, f} f(y) \leqslant \frac{1}{2m} \mathrm{dist}^2(0, \partial f(x)),$$

其中 $\mathrm{dist}(z, S)$ 表示点 z 到集合 S 的欧几里得距离.

解 (俞建江). 利用引理 2.2 和 m-强凸函数的性质证明. $\square$

第三章　典型优化问题

3.1 将下面的问题转化为线性规划：给定 $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^n$,

(a) $\displaystyle\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_1, \quad \text{s.t.} \quad \|x\|_\infty \leqslant 1;$

(b) $\displaystyle\min_{x\in\mathbb{R}^n} \quad \|x\|_1, \quad \text{s.t.} \quad \|Ax-b\|_\infty \leqslant 1;$

(c) $\displaystyle\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_1 + \|x\|_\infty;$

(d) $\displaystyle\min_{x\in\mathbb{R}^n} \quad \sum_{i=1}^m \max\{0, a_i^\mathrm{T} x + b_i\}.$

解 (丁思哲). 考虑等价转化目标函数或条件中的非线性项.

(a)

$$\begin{aligned} \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^m_+} \quad & \sum_{i=1}^n z_i \\ \text{s.t.} \quad & -z \leqslant Ax-b \leqslant z, \\ & -\mathbf{1} \leqslant x \leqslant \mathbf{1}. \end{aligned}$$

(b)

$$\begin{aligned} \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^m_+} \quad & \sum_{i=1}^m z_i \\ \text{s.t.} \quad & -z \leqslant x \leqslant z, \\ & -\mathbf{1} \leqslant Ax-b \leqslant \mathbf{1}. \end{aligned}$$

(c)

$$\begin{aligned} \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^m_+, t\in\mathbb{R}_+} \quad & \sum_{i=1}^m z_i + t \\ \text{s.t.} \quad & -z \leqslant Ax-b \leqslant z, \\ & -t\mathbf{1} \leqslant x \leqslant t\mathbf{1}. \end{aligned}$$

(d)

$$\begin{aligned} \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^m_+} \quad & \sum_{i=1}^m z_i \\ \text{s.t.} \quad & z \geqslant Ax + b. \end{aligned}$$

□

3.2 求解下面的线性规划问题：给定向量 $c \in \mathbb{R}^n$，

(a) $\min\limits_{x\in\mathbb{R}^n} \quad c^{\mathrm{T}}x, \quad \text{s.t.} \quad \mathbf{0} \leqslant x \leqslant \mathbf{1};$

(b) $\min\limits_{x\in\mathbb{R}^n} \quad c^{\mathrm{T}}x, \quad \text{s.t.} \quad -\mathbf{1} \leqslant x \leqslant \mathbf{1};$

(c) $\min\limits_{x\in\mathbb{R}^n} \quad c^{\mathrm{T}}x, \quad \text{s.t.} \quad -1 \leqslant \mathbf{1}^{\mathrm{T}}x \leqslant 1;$

(d) $\min\limits_{x\in\mathbb{R}^n} \quad c^{\mathrm{T}}x, \quad \text{s.t.} \quad \mathbf{1}^{\mathrm{T}}x = 1, \quad x \geqslant 0;$

解 (邓展望).

(a) 解为

$$x = (x_1, ..., x_n)^{\mathrm{T}}, \quad x_i = \frac{1}{2} - \frac{1}{2}\text{sign}\,(c_i)\,.$$

(b) 解为

$$x = (x_1, ..., x_n)^{\mathrm{T}}, \quad x_i = -\text{sign}\,(c_i)\,.$$

(c) 若 $c_i \neq c_j$，原问题无界．若 $m > 0$，解在 $x = \sum\limits_i x_i = -1$ 处取得，否则解在 $x = \sum\limits_i x_i = 1$ 处取得.

(d) 设 c_j 为 $c_i(i = 1...n)$ 中最小的项，则解为 $x = (0, ...1, ...0)$，其中第 j 个分量取 1.

□

3.3 在数据插值中，考虑一个简单的复合模型（取 ϕ 为恒等映射，两层复合模型）：

$$\min_{X_1\in\mathbb{R}^{q\times p}, X_2\in\mathbb{R}^{q\times q}} \quad \sum_{i=1}^m \|X_2X_1a_i - b_i\|_2^2,$$

其中 $a_i \in \mathbb{R}^p, b_i \in \mathbb{R}^q, i = 1, 2, \cdots, m$.

(a) 试计算目标函数关于 X_1, X_2 的导数；

(b) 试给出该问题的最优解.

解 (俞建江). 将 a_i, b_i 整合成矩阵 A, B:

$$A = (a_1, a_2, \cdots, a_m) \in \mathbb{R}^{p\times m},\ B = (b_1, b_2, \cdots, b_m) \in \mathbb{R}^{q\times m},$$

(a) $\dfrac{\partial f}{\partial X_1} = 2X_2^{\mathrm{T}}(X_2X_1A - B)A^{\mathrm{T}}$, $\dfrac{\partial f}{\partial X_2} = 2(X_2X_1A - B)A^{\mathrm{T}}X_1^{\mathrm{T}}$.

(b) 令 $X = X_2X_1$, $g(X) = \|XA - B\|_F^2$, 考虑 $\dfrac{\partial g}{\partial X} = 0$. □

3.4 给定数据点 $a_i \in \mathbb{R}^n, b_i \in \mathbb{R}^n, i = 1, 2, \cdots, m$, 我们用二次函数拟合, 即求 $X \in \mathcal{S}^n, y \in \mathbb{R}^n, z \in \mathbb{R}$ 使得

$$b_i \approx f(a_i) = a_i^{\mathrm{T}} X a_i + y^{\mathrm{T}} a_i + z, \quad i = 1, 2, \cdots, m.$$

这里假设数据点 $a_i \in \mathcal{B} = \{a \in \mathbb{R}^n \mid l \leqslant a \leqslant u\}$. 相应的最小二乘模型为

$$\sum_{i=1}^m (f(a_i) - b_i)^2.$$

此外, 对函数 f 有三个额外要求: (1) f 是凹函数; (2) f 在集合 $\mathcal{B}$ 上是非负的, 即 $f(a) \geqslant 0, \forall a \in \mathcal{B}$; (3) f 在 $\mathcal{B}$ 上是单调非减的, 即对任意的 $a, \hat{a} \in \mathcal{B}$ 且满足 $a \leqslant \hat{a}$, 有 $f(a) \leqslant f(\hat{a})$.

请将上述问题表示成一个凸优化问题, 并尽可能地简化.

解 (邓展望). 凸优化问题可写为

$$\begin{aligned}
\min_{X\in\mathcal{S}, y\in\mathbb{R}^n, z\in R} \quad & \sum_{i=1}^m (a_i^{\mathrm{T}} X a_i - y^{\mathrm{T}} a_i - z + b_i)^2. \\
\text{s.t.} \quad & X \geqslant 0, \\
& x^{\mathrm{T}} X x^{\mathrm{T}} + y^{\mathrm{T}} x + z \geqslant 0, \quad x_i \in \{l_i, u_i\}, \\
& 2Xa + y \geqslant 0, \\
& a \in \mathcal{B}.
\end{aligned}$$

□

3.5 考虑下面的复合优化问题:

$$\min_{x\in\mathbb{R}^n} \quad \|x\|_1 + \|Dx - a\|_2^2,$$

其中 $a \in \mathbb{R}^n, D = \mathrm{Diag}(d_1, d_2, \cdots, d_n)$ 均已知. 试给出最优解 x^* 的表达式.

解 (邓展望). x^* 各分量的表达式[1]如下.

(a) 若 $-2d_ia_i+1\leqslant 0$，$x_i^*=\dfrac{2d_ia_i-1}{2d_i^2}$.

(b) 若 $-2d_ia_i-1\geqslant 0$，$x_i^*=\dfrac{1+2d_ia_i}{2d_i^2}$.

(c) 否则 $x_i^*=0$. □

3.6 考虑下面的复合优化问题：

$$\min_{x\in\mathbb{R}^n}\quad \|x\|_0+\|Dx-a\|_2^2,$$

其中 $a\in\mathbb{R}^n, D=\mathrm{Diag}(d_1,d_2,\cdots,d_n)$ 均已知. 试给出最优解 x^* 的表达式.

解 (邓展望). 按 **4.5** 题的方法. 最优解分量 x_i^* 的表达式为：

(a) 若 $d_i=0$，则 $x_i^*=0$.

(b) 若 $d_i\neq 0$，

   i. 若 $|a_i|\leqslant 1$，取 $x_i^*=0$ 较小，此时满足

$$g(0)\leqslant g(x_i).\quad (x_i\neq 0)$$

   ii. 若 $|a_i|>1$，取使 $(d_ix_i-a_i)^2$ 最小的非零 x_i，使得 $x_i^*=\dfrac{a_i}{d_i}$. □

3.7 将不等式形式的半定规划问题 (4.5.1) 转化成标准形式 (4.5.2).

解 (俞建江). 将原问题化为

$$\begin{aligned}\min_{y\in\mathbb{R}^m}\quad & -b^{\mathrm{T}}y,\\ \text{s.t.}\quad & \sum_{i=1}^m y_iA_i\preceq C.\end{aligned}\tag{3.1}$$

，根据对偶函数求对偶问题. □

[1] 可参考教材 8.4.12

3.8 对于对称矩阵 $C \in \mathcal{S}^n$，记其特征值分解为 $C = \sum_{i=1}^{n} \lambda_i u_i u_i^{\mathrm{T}}$，假设

$$\lambda_1 \geqslant \cdots \geqslant \lambda_m > 0 > \lambda_{m+1} \geqslant \cdots \geqslant \lambda_n,$$

考虑如下半定规划问题:

$$\begin{aligned}
\min_{X \in \mathcal{S}^n} \quad & \langle C, X \rangle, \\
\text{s.t.} \quad & u_i^{\mathrm{T}} X u_i = 0, i = m+1, m+2, \cdots, n, \\
& X \succeq 0.
\end{aligned}$$

试给出该问题最优解的表达式.

解 (俞建江). 考虑

$$\langle C, X \rangle = \mathrm{Tr}(C^{\mathrm{T}} X),$$

再由约束条件，可推出 $\mathrm{Tr}(X) = 0$ 和 $X = 0$. □

3.9 如果在最大割问题 (4.5.6) 中，约束 $x_j \in \{-1, 1\}$ 改为 $x_j \in \{0, 1\}$，即对应优化问题

$$\begin{aligned}
\max \quad & \frac{1}{2} \sum_{i<j} w_{ij}(1 - x_i x_j), \\
\text{s.t.} \quad & x_j \in \{0, 1\},\ j = 1, 2, \cdots, n.
\end{aligned}$$

试给出其一个半定规划松弛.

解 (俞建江). 考虑作某变换 $y = h(x)$ 将题目中问题的形式转化为标准的半定规划原问题形式，类似 (4.5.5) 和 (4.5.6) 进行松弛. □

第四章　最优性理论

4.1 考虑优化问题

$$\min_{x\in\mathbb{R}^n} \quad x^{\mathrm{T}}Ax + 2b^{\mathrm{T}}x,$$

其中 $A \in \mathcal{S}^n$, $b \in \mathbb{R}^n$. 为了保证该问题最优解存在，A, b 需要满足什么性质?

解 (邓展望). A, b 需要满足 A 为半正定矩阵并且 $b \in \mathcal{R}(A)$. □

4.2 试举例说明对无约束光滑优化问题，二阶必要条件不是充分的，二阶充分条件也不是必要的（见定理 5.4).

解 (陈铖). 举例某类多项式函数. □

4.3 证明下列锥是自对偶锥：

(a) 半正定锥 $\{X \mid X \succeq 0\}$（全空间为 $\mathcal{S}^n$）;

(b) 二次锥 $\{(x,t) \in \mathbb{R}^{n+1} \mid t \geqslant \|x\|_2\}$ （全空间为 $\mathbb{R}^{n+1}$).

解 (陈铖).

(a) 等价于证明 Y 对于任意半正定矩阵 X 有 $\langle X, Y\rangle \geqslant 0$ 成立 $\Leftrightarrow Y$ 是半正定矩阵.

(b) 令 $\mathcal{I} = \{(x,t) \in \mathbb{R}^{n+1} \mid t \geqslant \|x\|_2\}$，证明

$$(x', t') \in \mathcal{I} \quad \Leftrightarrow \quad \langle x', x\rangle + t't \geqslant 0, \forall (x,t) \in \mathcal{I}.$$

$\Rightarrow$: 利用柯西-施瓦茨不等式.

$\Leftarrow$: 利用柯西-施瓦茨不等式反证. □

4.4 考虑优化问题

$$\min_{x\in\mathbb{R}^n} \quad x^{\mathrm{T}}Ax+2b^{\mathrm{T}}x, \quad \text{s.t.} \quad \|x\|_2\leqslant \Delta,$$

其中 $A\in\mathcal{S}^n_{++}$，$b\in\mathbb{R}^n$，$\Delta>0$．求出该问题的最优解.

解 (陈铖)．验证满足 Slater 条件，利用 KKT 条件写出最优解满足的必要条件. □

4.5 考虑函数 $f(x)=2x_1^2+x_2^2-2x_1x_2+2x_1^3+x_1^4$，求出其所有一阶稳定点，并判断它们是否为局部最优点 (极小或极大)、鞍点或全局最优点?

解 (陈铖)．$(0,0)$ 和 $(-1,-1)$ 是全局最优点，$(-0.5,-0.5)$ 是鞍点. □

4.6 给出下列优化问题的显式解:

(a) $\displaystyle\min_{x\in\mathbb{R}^n} \quad c^{\mathrm{T}}x, \quad \text{s.t.} \quad Ax=b$，其中 $A\in\mathbb{R}^{m\times n}$，$b\in\mathbb{R}^m$;

(b) $\displaystyle\min_{x\in\mathbb{R}^n} \quad \|x\|_2, \quad \text{s.t.} \quad Ax=b$;

(c) $\displaystyle\min_{x\in\mathbb{R}^n} \quad c^{\mathrm{T}}x, \quad \text{s.t.} \quad \mathbf{1}^{\mathrm{T}}x=1,\ x\geqslant 0$;

(d) $\displaystyle\min_{X\in\mathbb{R}^{m\times n}} \quad \|X\|_*+\frac{1}{2}\|X-Y\|_F^2$，其中 $Y\in\mathbb{R}^{m\times n}$ 是已知的.

解 (陈铖，邓展望).

(a) 将 $Ax=b$ 的解分解为特解和对应齐次方程解的形式，讨论齐次方程的解的存在性.

(b) 构造拉格朗日函数，由 KKT 条件求出全局最优解．若 A 不是行满秩的，参考上一小节.

(c) 设 i 为 c 的最小分量的下标，则 $x=e_i$，即 x 为第 i 个分量为 1 的单位向量.

(d) 利用 Y 和 X 的奇异值分解和正交变换的性质将原问题中的矩阵替换成它们对应的奇异值矩阵. □

4.7 计算下列优化问题的对偶问题.

(a) $\displaystyle\min_{x\in\mathbb{R}^n} \quad \|x\|_1, \quad \text{s.t.} \quad Ax=b$;

(b) $\min\limits_{x\in\mathbb{R}^n} \quad \|Ax-b\|_1$;

(c) $\min\limits_{x\in\mathbb{R}^n} \quad \|Ax-b\|_\infty$;

(d) $\min\limits_{x\in\mathbb{R}^n} \quad x^{\mathrm{T}}Ax+2b^{\mathrm{T}}x, \quad \text{s.t.} \quad \|x\|_2^2 \leqslant 1$, 其中 A 为正定矩阵.

解 (陈铖). 构造拉格朗日函数，对拉格朗日函数取极小，即可得到对偶问题.

(a) 对偶问题为

$$\begin{aligned} \max \quad & -b^{\mathrm{T}}\lambda, \\ \text{s.t.} \quad & \|A^{\mathrm{T}}\lambda\|_\infty \leqslant 1. \end{aligned}$$

(b) 对偶问题为

$$\begin{aligned} \max \quad & -b^{\mathrm{T}}\lambda, \\ \text{s.t.} \quad & \|\lambda\|_\infty \leqslant 1, \\ & A^{\mathrm{T}}\lambda = 0. \end{aligned}$$

(c) 对偶问题为

$$\begin{aligned} \max \quad & -b^{\mathrm{T}}(\lambda_1+\lambda_2), \\ \text{s.t.} \quad & A^{\mathrm{T}}(\lambda_1+\lambda_2) = 0, \\ & -\lambda_1^{\mathrm{T}}\mathbf{1}+\lambda_2^{\mathrm{T}}\mathbf{1}+1 = 0. \end{aligned}$$

(d) 对偶问题为

$$\begin{aligned} \max \quad & -b^{\mathrm{T}}(A+\lambda I)^{-1}b-\lambda, \\ \text{s.t.} \quad & \lambda \geqslant 0. \end{aligned}$$

□

4.8 如下论断正确吗？为什么？

对等式约束优化问题

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & c_i(x) = 0,\ i\in\mathcal{E}. \end{aligned}$$

考虑与之等价的约束优化问题：

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & c_i^2(x) = 0,\ i\in\mathcal{E}. \end{aligned} \tag{4.1}$$

设 $x^\sharp$ 是上述问题的一个 KKT 点，根据 (5.5.8) 式，$x^\sharp$ 满足

$$\begin{aligned}
0 &= \nabla f(x^\sharp) + 2\sum_{i\in\mathcal{E}} \lambda_i^\sharp c_i(x^\sharp)\nabla c_i(x^\sharp),\\
0 &= c_i(x^\sharp), \quad i\in\mathcal{E},
\end{aligned}$$

其中 $\lambda_i^\sharp$ 是相应的拉格朗日乘子. 整理上式得 $\nabla f(x^\sharp)=0$. 这说明对等式约束优化问题，我们依然能给出类似无约束优化问题的最优性条件.

解 (邓展望). 不正确. KKT 条件所需的约束品性不满足. □

4.9 证明：若在点 x 处线性约束品性（见定义 5.11）满足，则有 $T_{\mathcal{X}}(x)=\mathcal{F}(x)$.

解 (陈铖). 只需证明 $\mathcal{F}(x)\subseteq T_{\mathcal{X}}(x)$. 取 $z_k = x+t_k d$，$\{t_k\}$ 为一组正标量且 $\lim\limits_{k\to\infty} t_k = 0$，证明 $z_k\in\mathcal{X}$. □

4.10 考虑优化问题

$$\begin{aligned}
\min_{x\in\mathbb{R}^2} \quad & x_1,\\
\text{s.t.} \quad & 16-(x_1-4)^2-x_2^2 \geqslant 0,\\
& x_1^2+(x_2-2)^2-4=0.
\end{aligned}$$

求出该优化问题的 KKT 点，并判断它们是否是局部极小点、鞍点以及全局极小点?

解 (陈铖). 局部极小点包括 $(0,0)$ 和 $(\frac{8}{5},\frac{16}{5})$，全局极小点是 $(0,0)$. 所有的 KKT 点均非鞍点. □

解 (陈铖). 对 X 和 S 分别进行对角化 $X=Q\Lambda_1Q^{\mathrm{T}}$，$S=R\Lambda_2R^{\mathrm{T}}$，证明 $\langle X,S\rangle=0$. □

4.11 考虑等式约束的最小二乘问题

$$\min_{x\in\mathbb{R}^n} \quad \|Ax-b\|_2^2, \quad \text{s.t.} \quad Gx=h,$$

其中 $A\in\mathbb{R}^{m\times n}$ 且 $\mathrm{rank}(A)=n$，$G\in\mathbb{R}^{p\times n}$ 且 $\mathrm{rank}(G)=p$.

(a) 写出该问题的对偶问题；

(b) 给出原始问题和对偶问题的最优解的显式表达式.

解 (陈铖). (a) 构造拉格朗日函数，对变量 x 取极小可得对偶问题.

(b) 原始问题最优解由 KKT 条件解出，对偶问题的最优解直接给出. □

4.12 考虑优化问题

$$\min_{x\in\mathbb{R}^n} \quad x^{\mathrm{T}}Ax + 2b^{\mathrm{T}}x, \quad \text{s.t.} \quad \|x\|_2 \leqslant 1,$$

其中 $A \in \mathcal{S}^n$，$b \in \mathbb{R}^n$. 写出该问题的对偶问题，以及对偶问题的对偶问题.

解 (陈铖). 利用对偶问题的定义. □

4.13 考虑支持向量机问题

$$\begin{aligned}
\min_{x\in\mathbb{R}^n,\xi} \quad & \frac{1}{2}\|x\|_2^2 + \mu\sum_{i=1}^m \xi_i, \\
\text{s.t.} \quad & b_i a_i^{\mathrm{T}} x \geqslant 1 - \xi_i,\ i = 1, 2, \cdots, m, \\
& \xi_i \geqslant 0,\ i = 1, 2, \cdots, m,
\end{aligned}$$

其中 $\mu > 0$ 为常数且 $b_i \in \mathbb{R}$，$a_i \in \mathbb{R}^n$，$i = 1, 2, \cdots, m$ 是已知的. 写出该问题的对偶问题.

解 (陈铖). 利用对偶问题的定义. □

4.14 考虑优化问题

$$\min_{x\in\mathbb{R},y>0} \quad e^{-x}, \quad \text{s.t.} \quad \frac{x^2}{y} \leqslant 0.$$

(a) 证明这是一个凸优化问题，求出最小值并判断 Slater 条件是否成立；

(b) 写出该问题的对偶问题，并求出对偶问题的最优解以及对偶间隙.

解 (丁思哲).

(a) 目标和约束函数均是凸函数，因此是凸优化问题.
最小值是 $x=0$，Slater 条件不成立.

(b) 原问题等价于

$$\min_{x\in\mathbb{R}} \quad e^{-x}, \quad \text{s.t.} \quad x=0,$$

其对偶问题的最优解为 $v=1$，对偶间隙为 0. □

解 (邓展望). 解不唯一. 利用 KKT 条件，一组解满足

$$ZV = X - \mathbf{1}\mathbf{1}^{\mathrm{T}}\frac{X}{n}.$$

□

第五章　无约束优化算法

5.1 设 $f(x)$ 是连续可微函数，d^k 是一个下降方向，且 $f(x)$ 在射线 $\{x^k + \alpha d^k \mid \alpha > 0\}$ 上有下界．求证：当 $0 < c_1 < c_2 < 1$ 时，总是存在满足 Wolfe 准则 (6.1.4a) (6.1.4b) 的点.并举一个反例说明当 $0 < c_2 < c_1 < 1$ 时，满足 Wolfe 准则的点可能不存在.

解 (谢中林)．利用 $\phi(\alpha) = f(x^k + \alpha d^k)$ 在 $\alpha = 0$ 处的一阶泰勒展开和(**??**)证明. □

5.2 f 为正定二次函数 $f(x) = \dfrac{1}{2}x^{\mathrm{T}}Ax + b^{\mathrm{T}}x$，$d^k$ 为下降方向，x^k 为当前迭代点．试求出精确线搜索步长

$$\alpha_k = \mathop{\arg\min}_{\alpha > 0} f(x^k + \alpha d^k),$$

并由此推出最速下降法的步长满足 (6.2.2) 式（见定理 6.2).

解 (谢中林)．$f(x^k + \alpha d^k)$ 关于 α 强凸，利用一阶条件可以导出精确线搜索步长. □

5.3 利用定理 6.5 证明推论 6.2.

解 (谢中林)．只说明 (2) 的证明思路. 注意到 $\alpha_i\|g^i\|$ 是常数, 且 $\|g^i\| \leqslant G$ 可对 $\displaystyle\sum_{i=0}^{k}\alpha_i$ 建立估计. □

5.4 考虑非光滑函数

$$f(x) = \max_{1 \leqslant i \leqslant K} x_i + \frac{1}{2}\|x\|^2,$$

其中 $x \in \mathbb{R}^n$，$K \in [1, n]$ 为一个给定的正整数.

(a) 求出 $f(x)$ 的最小值点 x^* 和对应的函数值 f^*；

(b) 证明 $f(x)$ 在区域 $\{x \mid \|x\| \leqslant R \overset{\text{def}}{=\!=} 1/\sqrt{K}\}$ 上是 G-利普希茨连续的，其中 $G = 1 + \dfrac{1}{\sqrt{K}}$；

(c) 设初值 $x^0 = 0$，考虑使用次梯度算法 (6.3.1) 对 $\min f(x)$ 进行求解，其中 x 处的次梯度取为 $g = x + e_j$，j 为使得 $x_j = \max\limits_{1 \leqslant i \leqslant K} x_i$ 成立的最小整数，步长 α_k 可任意选取，证明：在 k（$k < K$）次迭代后，

$$\hat{f}^k - f^* \geqslant \frac{GR}{2(1+\sqrt{K})},$$

其中 $\hat{f}^k$ 的定义和定理 6.5 相同．并根据此例子推出次梯度算法的收敛速度 $\mathcal{O}\left(\dfrac{GR}{\sqrt{K}}\right)$ 是不能改进的．

解 (谢中林).

(a) 最小值点为

$$x_i^* = -\frac{1}{K}, i = 1, \ldots, K; \quad x_i^* = 0, i = K+1, \ldots, n.$$

(b) 只需注意对 $\max\limits_{1 \leqslant i \leqslant K} x_i$ 取合适的放缩估计.

(c) 利用 x^k 的具体取值和 $f(x^k)$、f^* 的表达式进行放缩估计. □

5.5 考虑非平方 ℓ_2 正则项优化问题

$$\min \quad f(x) = \frac{1}{2}\|Ax - b\|_2^2 + \mu\|x\|_2,$$

其中 $A \in \mathbb{R}^{m \times n}$，注意这个问题并不是岭回归问题.

(a) 若 A 为列正交矩阵，即 $A^{\mathrm{T}}A = I$，利用不可微函数的一阶最优性条件求出该优化问题的显式解；

(b) 对一般的 A 我们可以使用迭代算法来求解这个问题．试设计出不引入次梯度的一种梯度类算法求解该优化问题．提示：$f(x)$ 仅在一点处不可导，若这个点不是最小值点，则次梯度算法和梯度法等价.

解 (谢中林).

(a) $\|A^{\mathrm{T}}b\|_2 > \mu$ 时，存在唯一最优解

$$x = \left(1 - \frac{\mu}{\|A^{\mathrm{T}}b\|_2}\right) A^{\mathrm{T}}b.$$

$\mu \geqslant \|A^{\mathrm{T}}b\|_2$ 时，$x = 0$ 是最优解.

(b) $\mu < \|A^{\mathrm{T}}b\|_2$ 时证明 $x = 0$ 不是 $g_\lambda(x)$ 的最小值点即可. □

5.6 设函数 $f(x) = \|x\|^\beta$，其中 $\beta > 0$ 为给定的常数. 考虑使用经典牛顿法 (6.4.2) 对 $f(x)$ 进行极小化，初值 $x^0 \neq 0$. 证明：

(a) 若 $\beta > 1$ 且 $\beta \neq 2$，则 x^k 收敛到 0 的速度为 Q-线性的；

(b) 若 $0 < \beta < 1$，则牛顿法发散；

(c) 试解释定理 6.6 在 (a) 中不成立的原因.

解 (谢中林).

(a) 利用牛顿方程考察 $\dfrac{\|x+d\|^2}{\|x\|^2}$ 的值.

(b) 同上.

(c) 利用 $f(x)$ 的海瑟矩阵证明. □

5.7 设矩阵 A 为 n 阶对称矩阵，d^k 为给定的非零向量. 若对任意满足 $\|d\| = \|d^k\|$ 的 $d \in \mathbb{R}^n$，均有 $(d - d^k)^{\mathrm{T}}A(d - d^k) \geqslant 0$，证明：$A$ 是半正定矩阵.

解 (谢中林). 利用半正定矩阵的定义，考虑设 $d = d^k + \alpha x$ 证明. □

5.8 设 $f(x)$ 为正定二次函数，且假定在迭代过程中 $(s^k - H^k y^k)^{\mathrm{T}} y^k > 0$ 对任意的 k 均满足，其中 H^k 由 SR1 公式 (6.5.10) 产生的拟牛顿矩阵. 证明：

$$H^k y^j = s^j, \quad j = 0, 1, \cdots, k-1,$$

其中 k 是任意给定的整数. 这个结论说明对于正定二次函数，SR1 公式产生的拟牛顿矩阵在当前点处满足割线方程，且历史迭代产生的 (s^j, y^j) 也满足割线方程.

解 (谢中林). 利用归纳法证明. □

5.9 仿照 BFGS 公式的推导过程, 利用待定系数法推导 DFP 公式 (6.5.15).

解 (谢中林). 设出 DFP 的秩二修正:

$$H^{k+1} = H^k + auu^{\mathrm{T}} + bvv^{\mathrm{T}},$$

代入割线方程推导出各系数. □

5.10 (小样本问题) 设 $J(x) \in \mathbb{R}^{m\times n}$ 为最小二乘问题 (6.7.1) 中 $r(x)$ 在点 x 处的雅可比矩阵，其中 $m \ll n$. 设 $J(x)$ 行满秩，证明:

$$\hat{d} = -J(x)^{\mathrm{T}}(J(x)J(x)^{\mathrm{T}})^{-1}r(x)$$

给出了高斯 – 牛顿方程 (6.7.3) 的一个 ℓ_2 范数最小解.

解 (谢中林). 考虑证明解 d 满足 $\hat{d} \leqslant d$. 进一步可证 $(d-\hat{d})^{\mathrm{T}}\hat{d} \geqslant 0$. □

第六章　约束优化算法

6.1 构造一个等式约束优化问题，使得它存在一个局部极小值，但对于任意的 $\sigma > 0$，它的二次罚函数是无界的.

解 (谢中林). 例如

$$\min_{x,y} \quad -e^x, \quad \text{s.t. } x^2 + y^2 = 1. \qquad \square$$

6.2 考虑等式约束优化问题

$$\begin{aligned} \min \quad & -x_1x_2x_3, \\ \text{s.t.} \quad & x_1 + 2x_2 + 3x_3 = 60. \end{aligned}$$

使用二次罚函数求解该问题，当固定罚因子 σ_k 时，写出二次罚函数的最优解 x^{k+1}. 当 $\sigma_k \to +\infty$ 时，写出该优化问题的解并求出约束的拉格朗日乘子. 此外，当罚因子 σ 满足什么条件时，二次罚函数的海瑟矩阵 $\nabla^2_{xx} P_E(x, \sigma)$ 是正定的?

解 (谢中林). 最优解为

$$x_1 = 20, \quad x_2 = 10, \quad x_3 = \frac{20}{3},$$

拉格朗日乘子为 $-\dfrac{200}{3}$.
再由 $\nabla^2_{xx} P_E(x(\sigma), \sigma)$ 的正定性确定 σ 的值. $\square$

6.3 考虑等式约束优化问题

$$\min \quad f(x), \quad \text{s.t.} \quad c_i(x) = 0, \ i \in \mathcal{E},$$

定义一般形式的罚函数

$$P_E(x,\sigma)=f(x)+\sigma\sum_{i\in\mathcal{E}}\varphi(c_i(x)),$$

其中 $\varphi(t)$ 是充分光滑的函数，且 $t=0$ 是其 s 阶零点（$s\geqslant 2$），即

$$\varphi(0)=\varphi'(0)=\cdots=\varphi^{(s-1)}(0)=0,\quad \varphi^{(s)}(0)\neq 0.$$

设 x^k,σ_k 的选取方式和算法 7.1 的相同，且 $\{x^k\}$ 存在极限 x^*，在点 x^* 处 LICQ （见定义 5.9）成立.

(a) 证明：$\sigma_k(c_i(x^k))^{s-1},\ \forall\, i\in\mathcal{E}$ 极限存在，其极限 λ_i^* 为约束 $c_i(x^*)=0$ 对应的拉格朗日乘子；

(b) 求 $P_E(x,\sigma)$ 关于 x 的海瑟矩阵 $\nabla^2_{xx}P_E(x,\sigma)$；

(c) 设在 (a) 中 $\lambda_i^*\neq 0,\ \forall\, i\in\mathcal{E}$，证明：当 $\sigma_k\to+\infty$ 时，$\nabla^2_{xx}P_E(x^k,\sigma_k)$ 有 m 个特征值的模长与 $\sigma_k^{1/(s-1)}$ 同阶，其中 $m=|\mathcal{E}|$.

解 (陈铖，丁思哲).

(a) 利用定理 7.2.

(b) 海瑟矩阵为

$$\begin{aligned}\nabla^2_{xx}P_E(x,\sigma)=\ &\nabla^2_{xx}f(x)+\sigma\sum_{i\in\mathcal{E}}\varphi'(c_i(x))\nabla^2_{xx}c_i(x)\\&+\sigma\sum_{i\in\mathcal{E}}\varphi''(c_i(x))\nabla_x c_i(x)\nabla_x c_i(x)^{\mathrm{T}}.\end{aligned}$$

(c) 考虑 k 较大时用拉格朗日函数近似海瑟矩阵的前 2 项，利用泰勒展开分析海瑟矩阵的阶数. □

6.4 考虑不等式约束优化问题 (7.1.12)，其中 f 在可行域 $\mathcal{X}$ 上有下界，现使用对数罚函数法进行求解（算法 7.4）. 假设在算法 7.4 的每一步子问题能求出罚函数的全局极小值点 x^{k+1}，证明：算法 7.4 在有限次迭代后终止，或者

$$\lim_{k\to\infty}\sigma_k\sum_{i\in\mathcal{I}}\ln(-c_i(x^{k+1}))=0,$$

并且

$$\lim_{k\to\infty}f(x^k)=\inf_{x\in\mathbf{int}\mathcal{X}}f(x).$$

解 (谢中林). 若算法不能在有限步终止，利用下确界的性质分别证明等式成立. 应注意讨论对数罚函数中约束函数的取值. □

6.5 考虑一般约束优化问题 (7.1.15)，现在针对等式约束使用二次罚函数，对不等式约束使用对数罚函数：

$$P(x,\sigma) = f(x) + \frac{\sigma}{2}\sum_{i\in\mathcal{E}} c_i^2(x) - \frac{1}{\sigma}\sum_{i\in\mathcal{I}} \ln(-c_i(x)),$$

其中 $\mathbf{dom}\, P = \{x \mid c_i(x) < 0,\ i \in \mathcal{I}\}$. 令罚因子 $\sigma_k \to +\infty$，定义

$$x^{k+1} = \arg\min_x P(x, \sigma_k).$$

假定涉及的所有函数都是连续的，$\{x \mid c_i(x) \leqslant 0,\ i \in \mathcal{I}\}$ 是有界闭集，x^* 为问题 (7.1.15) 的解. 试证明如下结论：

(a) $\lim\limits_{k\to\infty} P(x^{k+1}, \sigma_k) = f(x^*)$;

(b) $\lim\limits_{k\to\infty} \sigma_k \sum\limits_{i\in\mathcal{E}} c_i^2(x^{k+1}) = 0$;

(c) $\lim\limits_{k\to\infty} \dfrac{1}{\sigma_k} \sum\limits_{i\in\mathcal{I}} \ln(-c_i(x^{k+1})) = 0$.

解 (陈铖、丁思哲). (a) 利用 $P(x^{k+1},\sigma_k) \geqslant f(x^*)$ 和 $P(x^{k+1},\sigma_k) \leqslant P(x^*,\sigma_k)$ 证明.

(c) 考虑构造固定罚因子且罚函数为对数函数的新函数，对新函数再构造二次罚函数，由二次罚函数的收敛性证明.

(b) 由 (a) 和 (c) 推导 (b) 成立. □

6.6 (Morrison 方法) 考虑等式约束优化问题 (7.1.1)，设其最优解为 x^*. 令 M 是最优函数值 $f(x^*)$ 的一个下界估计（即 $M \leqslant f(x^*)$），构造辅助函数

$$v(M, x) = [f(x) - M]^2 + \sum_{i\in\mathcal{E}} c_i^2(x),$$

Morrison 方法的迭代步骤如下：

$$x^k = \arg\min_x\ v(M_k, x),$$
$$M_{k+1} = M_k + \sqrt{v(M_k, x^k)}.$$

试回答以下问题：

(a) 证明：$f(x^k) \leqslant f(x^*)$；

(b) 若 $M_k \leqslant f(x^*)$，证明：$M_{k+1} \leqslant f(x^*)$；

(c) 证明：$\lim\limits_{k\to\infty} M_k = f(x^*)$；

(d) 求 $v(M,x)$ 关于 x 的海瑟矩阵，并说明 Morrison 方法和算法 7.1 的联系.

解 (丁思哲).

(a) 从 $f(x^k) > f(x^*)$ 时 $v(M,x^*)$ 和 $v(M,x^k)$ 的比较可以推出矛盾.

(b) 利用 $v(M_k,x^k) \leqslant v(M_k,x^*)$.

(c) 先证明 $\lim\limits_{k\to\infty} M_k$ 存在，然后在题设式中令 $k\to\infty$.

(d) 海瑟矩阵略. Morrison 方法在 k 足够大时，相当于对问题

$$\min_x \{|f(x) - f(x^*)|^2 + 2\sum_{i\in\mathcal{E}} c_i^2\,(x)\}$$

使用算法 7.1 求解. □

6.7 考虑不等式约束优化问题

$$\min \quad f(x), \quad \text{s.t.} \quad c_i(x) \leqslant 0,\ i\in\mathcal{I}.$$

(a) 定义函数 $F(x) = \sup\limits_{\lambda_i\geqslant 0}\left\{ f(x) + \sum\limits_{i\in\mathcal{I}} \lambda_i c_i(x) \right\}$，证明：原问题等价于无约束优化问题 $\min\limits_x F(x)$；

(b) 定义函数

$$\hat{F}(x,\lambda^k,\sigma_k) = \sup_{\lambda_i\geqslant 0}\left\{ f(x) + \sum_{i\in\mathcal{I}} \lambda_i c_i(x) - \frac{\sigma_k}{2}\sum_{i\in\mathcal{I}} (\lambda_i - \lambda_i^k)^2 \right\},$$

求 $\hat{F}(x,\lambda^k,\sigma_k)$ 的显式表达式；

(c) 考虑如下优化算法：

$$\begin{aligned} x^k &= \arg\min_x \hat{F}(x,\lambda^k,\sigma_k), \\ \lambda^{k+1} &= \arg\max_{\lambda\geqslant 0}\left\{ \sum_{i\in\mathcal{I}} \lambda_i c_i(x^k) - \frac{\sigma_k}{2}\sum_{i\in\mathcal{I}} (\lambda_i - \lambda_i^k)^2 \right\}, \\ \sigma_{k+1} &= \min\{\rho\sigma_k, \bar{\sigma}\}, \end{aligned}$$

试说明其与算法 7.5 的区别和联系.

解 (丁思哲). (a) 利用原问题的广义拉格朗日函数证明，注意对 x 取值的讨论.

(b) 适当取 λ_i 的值，使得 $\hat{F}(x,\lambda^k,\sigma_k)$ 取极值.

(c) 迭代格式中的 $1/\sigma_k$ 对应与算法 7.5 中的 σ_k. □

6.8 对于 LASSO 问题

$$\min_{x\in\mathbb{R}^n} \quad \frac{1}{2}\|Ax-b\|_2^2+\mu\|x\|_1,$$

写出该问题及其对偶问题的增广拉格朗日函数法.

解 (丁思哲).

LASSO 问题的增广拉格朗日函数法为

$$\begin{aligned}
\left(x^{k+1},z^{k+1}\right) &= \operatorname*{argmin}_{x,z}\left\{L_{\sigma_k}\left(x,z,\lambda^k\right)\right\},\\
\lambda^{k+1} &= \lambda^k+\sigma_k\left(Ax^{k+1}-z^{k+1}\right),\\
\sigma_{k+1} &= \rho\sigma_k,\quad \sigma_k\uparrow\sigma_\infty<\infty.
\end{aligned}$$

LASSO 对偶问题的增广拉格朗日函数法为

$$\begin{aligned}
\left(y^{k+1},s^{k+1}\right) &= \arg\min_{y,s}\left\{L_{\sigma_k}\left(y,s;\lambda^k\right)\right\},\\
\lambda^{k+1} &= \lambda^k+\sigma_k\left(A^{\mathrm{T}}y^{k+1}-s^{k+1}\right),\\
\sigma_{k+1} &= \rho\sigma_k,\quad \sigma_k\uparrow\sigma_\infty\leqslant\infty.
\end{aligned}$$

□

6.9 考虑线性规划问题

$$\min_{x\in\mathbb{R}^n} \quad c^{\mathrm{T}}x,\quad \text{s.t.}\quad Ax=b,\ x\geqslant 0.$$

写出该问题以及其对偶问题的增广拉格朗日函数法.

解 (丁思哲). 方法同上. □

第七章　复合优化算法

7.1 证明例 8.2 中的运算法则成立.

解 (丁思哲).　证明等式两边各自的最优性条件相同. □

7.2 求下列函数的邻近算子：

(a) $f(x) = I_C(x)$，其中 $C = \{(x,t) \in \mathbb{R}^{n+1} | \|x\|_2 \leqslant t\}$；

(b) $f(x) = \inf\limits_{y \in C} \|x - y\|$，其中 C 是闭凸集；

(c) $f(x) = \dfrac{1}{2}(\inf\limits_{y \in C} \|x - y\|)^2$，其中 C 是闭凸集.

解 (丁思哲).

(a) $u = \mathcal{P}_{\|x\|_2 \leqslant t}(x)$.

(b) 若 C 是闭凸集且 $\|x - \mathcal{P}_C(x)\| > 1$，则 $u = x + \dfrac{\mathcal{P}_C(x) - x}{\|\mathcal{P}_C(x) - x\|}$；反之则 $u = \mathcal{P}_C(x)$.

(c) $u = \dfrac{1}{2}x + \dfrac{1}{2}\mathcal{P}_C(x)$. □

7.3 对一般复合优化问题的加速算法（算法 8.9），试证明：

(a) 当 $t_k = \gamma_k \lambda_k$ 且 $h(x) = 0$ 时，算法 8.9 等价于第二类 Nesterov 加速算法；

(b) 当 $t_k = \lambda_k$ 时，算法 8.9 等价于近似点梯度法.

解 (丁思哲).

(a) 将条件代入算法 8.9，只需证明 $x^k = (1 - \gamma_k)x^{k-1} + \gamma_k y^k$.

(b) 将条件代入算法 8.9，只需证明 $y^k = x^k$. □

7.4 假设 f 是闭凸函数，证明 Moreau 分解的推广成立，即对任意的 $\lambda > 0$ 有

$$x = \mathrm{prox}_{\lambda f}(x) + \lambda\, \mathrm{prox}_{\lambda^{-1} f^*}\left(\frac{x}{\lambda}\right) \quad \forall x.$$

提示：利用 *Moreau* 分解的结论.

解 (邓展望). 由 Moreau 分解的结论可知

$$\mathrm{prox}_{\lambda f}(x) = x - \mathrm{prox}_{\lambda f^*(\cdot/\lambda)}(x).$$ □

7.5 写出关于 LASSO 问题的鞍点问题形式，并写出原始 – 对偶混合梯度算法和 Chambolle-Pock 算法.

解 (丁思哲). LASSO 鞍点问题的形式为

$$\min_x \max_z \left\{ \mu \left\| x \right\|_1 + z^{\mathrm{T}} A x - \frac{1}{2} \left\| z \right\|_2^2 - b^{\mathrm{T}} z \right\}. \tag{7.1}$$

照此设计相应算法. LASSO 鞍点问题的形式不唯一. □

7.6 设函数 $f(x_1, x_2) = |x_1 - x_2| - \min\{x_1, x_2\}$，其定义域为 $[0,1] \times [0,1]$. 试推导基于格式 (8.4.3) 的分块坐标下降法（x_1 和 x_2 分别看做一个变量块），此算法是否收敛?

解 (丁思哲). 函数可写成

$$f(x_1, x_2) = \frac{3}{2} |x_1 - x_2| - \frac{1}{2}(x_1 + x_2).$$

将 x_1, x_2 视为 2 个变量块，进行分块下降，算法收敛. □

7.7 试对分组 LASSO 问题（即例 8.7）推导出基于格式 (8.4.4) 的分块坐标下降法.

解 (邓展望). 问题的形式为

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2} \|Ax - b\|_2^2 + \mu_1 \sum_{\ell=1}^{G} \sqrt{n_\ell} \|x_{\mathcal{I}_\ell}\|_2, \tag{7.2}$$

其中第 i 块变量为 x 的第 i 组分量. 据此设计分块下降算法. □

7.8 考虑约束优化问题

$$\begin{aligned} \min \quad & \max\{e^{-x}+y, y^2\}, \\ \text{s.t.} \quad & y \geqslant 2, \end{aligned}$$

其中 $x, y \in \mathbb{R}$ 为自变量.

(a) 通过引入松弛变量 z，试说明该问题等价于

$$\begin{aligned} \min \quad & \max\{e^{-x}+y, y^2\} + I_{\mathbb{R}_+}(z), \\ \text{s.t.} \quad & y - z = 2; \end{aligned}$$

(b) 推导 (a) 中问题的对偶问题，并求出原始问题的最优解；

(c) 对 (a) 中的问题形式，使用 ADMM 求解时可能会遇到什么问题?

解 (邓展望).

(a) 将条件 $z \geqslant 0$ 加在目标函数上.

(b) 最优解为 $z = 0, x = \infty, y = 2, \lambda = -4$.

(c) 若初始条件为 $z = 0, \lambda = 0$，ADMM 算法无法在下一步产生关于 (x, y) 的最小值点. □

7.9 写出对于线性规划对偶问题运用 ADMM 的迭代格式，以及与之等价的对于原始问题的 DRS 格式，并指出 ADMM 和 DRS 算法更新变量之间的关系.

解 (陈铖). ADMM 迭代格式略. 将上述问题改写为可分的凸问题的形式，进一步得到对偶问题，并据此设计 DRS 算法.
ADMM 算法与 DRS 算法中的变量存在一一对应的关系. □

7.10 考虑 ℓ_0 范数优化问题的罚函数形式：

$$\min \quad \lambda\|x\|_0 + \frac{1}{2}\|Ax-b\|^2,$$

其中 $A \in \mathbb{R}^{m\times n}, m < n$ 为实矩阵，$\|\cdot\|_0$ 为 ℓ_0 范数，即非零元素的个数. 试针对 ℓ_0 范数优化问题形式化推导具有两个变量块的 ADMM 格式. 在算法中每个子问题是如何求解的?

解 (陈铖). 考虑上述问题的等价形式：

$$\begin{aligned} \min \quad & \lambda\|z\|_0 + \frac{1}{2}\|Ax-b\|^2, \\ \text{s.t.} \quad & x = z. \end{aligned}$$

写出增广拉格朗日函数，利用 ADMM 法分别求解相关变量的子问题为更新方式. □

7.11 试说明 LASSO 对偶问题中，若在问题 (8.6.27) 中对约束

$$A^{\mathrm{T}}y + z = 0$$

引入乘子 x，则 x 恰好对应 LASSO 原始问题 (8.6.26) 中的自变量.

解 (邓展望). 对偶问题为

$$\max \quad -\mu\|x\|_1 - \frac{1}{2}\|Ax-b\|^2, \tag{7.3}$$

它与 LASSO 问题等价. □

7.12 实现关于 LASSO 问题使用以下算法的程序，并比较它们的效率

(a) 近似点梯度算法；

(b) Nesterov 加速算法；

(c) 交替方向乘子法；

(d) Chambolle-Pock 算法；

(e) 分块坐标下降法；

(f) 随机近似点梯度算法.

解. 见教材代码主页，此处从略. □

7.13 设 $f(x) = \dfrac{1}{N}\displaystyle\sum_{i=1}^{N} f_i(x)$，其中每个 $f_i(x)$ 是可微函数，且 $f(x)$ 为梯度 L-利普希茨连续的. $\{x^k\}$ 是由随机梯度下降法产生的迭代序列，s_k 为第 k 步随机抽取的下标. 证明：

$$\mathbb{E}[\|\nabla f_{s_k}(x^k)\|^2] \leqslant L^2\mathbb{E}[\|x^k - x^*\|^2] + \mathbb{E}[\|\nabla f_{s_k}(x^k) - \nabla f(x^k)\|^2].$$

（请注意可能与教材不同，此处为**订正**版本）其中 x^* 是 $f(x)$ 的一个最小值点，α_k 为第 k 步的步长.

解 (邓展望). 只需注意题设条件中的梯度 L-利普希茨连续即可进行合适的放缩估计. □

7.14 在 SAGA 算法中，每一步的下降方向取为:

$$v^k = \nabla f_{s_k}(x^k) - g_{s_k}^{k-1} + \frac{1}{N}\sum_{i=1}^{N} g_i^{k-1},$$

假设初值 $g_i^0 = 0, i = 1, 2, \cdots, N$，证明:

$$\mathbb{E}[v^k | s_1, s_2, \cdots, s_{k-1}] = \nabla f(x^k).$$

解 (邓展望). 利用随机梯度下降中随机梯度的期望收敛于梯度. □

更新历史

2022.04.04-2022.10.20

- 版本 v1.00 更新. 本次更新主要提供教材全部习题的参考解答（简短版本），是正式发布的第一版.
- 版本 v1.10 更新. 本次更新进一步精简了答案内容.
- 版本 v1.11 更新. 本次更新修改了部分题目的答案.
- **（最新）** 版本 v1.12 更新. 本次更新修改了 5.10 题的答案.

致谢

本书内容在北京大学数学科学学院多次开设的“凸优化”和“大数据分析中的算法”课程中使用，感谢选课同学的反馈和支持.