最优化算法作业 4

陈文轩

更新：May 20, 2025

1. 对于可逆矩阵 A，求解如下方程的零点

$$F(X) = X^{-1} - A = 0$$

可以得到 A^{-1}。

(a). 使用牛顿法，写出迭代公式。

(b). 实现该算法，随机生成 100×100 维可逆矩阵 A，作出误差随着迭代数的收敛图像。

提示：根据 $DF(X)[B] = -X^{-1}BX^{-1}$，计算 $DF(X)^{-1}[B]$

解

(a). 记 Newton 法迭代格式为 $X_{k+1} = X_k + \Delta X$，则 ΔX 满足 $DF(X_k)[\Delta X] = -F(X_k)$，即 $X_k^{-1}\Delta X X_k^{-1} = X_K^{-1} - A, \Delta X = X_k - X_k A X_k$，迭代公式为 $X_{k+1} = 2X_k - X_k A X_k$。

(b). 算法代码如下：

Newton 法代码

```python
import numpy as np
import matplotlib.pyplot as plt
np.random.seed(42); epsilon = 0.09
A = np.random.randn(100, 100)
while np.linalg.det(A) == 0:
    A = np.random.randn(100, 100)
A_inv = np.linalg.inv(A)
X = A_inv + np.eye(100) * epsilon
max_iter = 50; errors = []
for k in range(max_iter):
    X_new = 2 * X - np.dot(np.dot(X, A), X)
    error = np.linalg.norm(X_new - A_inv, 'fro')
    errors.append(error); X = X_new
plt.plot(range(max_iter), errors)
plt.yscale('log'); plt.xlabel('Iteration')
plt.ylabel('Frobenius Norm of Error')
plt.title('Convergence of Newton\'s Method'); plt.show()
```

测试时发现算法对初值极其敏感，`epsilon`=0.1 时便无法收敛。以下是 `epsilon`=0.09 时的收敛图像：

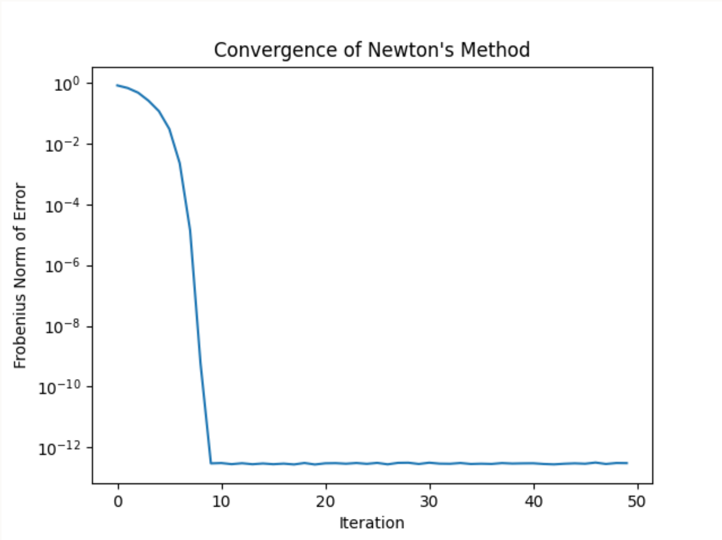

图 1: 牛顿法收敛图像

2. 给定集合 $C_i, i=1,2,\cdots,m$ 为闭凸集，且易于计算投影，考虑投影问题：

$$\begin{aligned} \min \quad & \frac{1}{2}\|x-c\|^2 \\ \text{s.t.} \quad & x \in C_1 \cap C_2 \cap \cdots \cap C_m \end{aligned}$$

请使用 ADMM 算法求解此问题，说明是否收敛（无需证明收敛）。

解

把问题重写为以下形式：

$$\begin{aligned} \min_{x,z} \quad & \frac{1}{2}\|z-c\|^2 + \sum_{i=1}^{m} \mathcal{I}_{C_i}(x_i) \\ \text{s.t.} \quad & x_i - z = 0, i=1,2,\cdots,m \end{aligned}$$

增广 Lagrange 函数为 $L(x,z,\lambda) = \frac{1}{2}\|z-c\|^2 + \sum_{i=1}^{m} \mathcal{I}_{C_i}(x_i) + \sum_{i=1}^{m} \lambda_i^\top (x_i - z) + \frac{\rho}{2}\sum_{i=1}^{m}\|x_i - z\|^2$，其中 λ_i 是 Lagrange 乘子，ρ 是正的罚参数。对 x, z 做交替极小化：

$$\begin{aligned} x_i^{k+1} =& \arg\min_{x_i} \left(\mathcal{I}_{C_i}(x_i) + \lambda_i^{k^\top}(x_i - z^k) + \frac{\rho}{2}\|x_i - z^k\|^2 \right) \\ =& \arg\min_{x_i} \left(\mathcal{I}_{C_i}(x_i) + \lambda_i^{k^\top} x_i - \lambda_i^{k^\top} z^k + \frac{\rho}{2}\left(\|x_i\|^2 + \|z^k\|^2 - 2x_i^\top z^k \right) \right) \\ =& \arg\min_{x_i} \left(\mathcal{I}_{C_i}(x_i) + \frac{\rho}{2}\left(\|x_i\|^2 + \left\| z^k - \frac{1}{\rho}\lambda_i^{k^\top} \right\|^2 - 2x_i^\top \left(z^k - \frac{1}{\rho}\lambda_i^{k^\top} \right) \right) \right) \\ =& \arg\min_{x_i} \left(\mathcal{I}_{C_i}(x_i) + \frac{\rho}{2}\left\| x_i - \left(z^k - \frac{1}{\rho}\lambda_i^{k^\top} \right) \right\| \right) \\ =& \mathcal{P}_{C_i}\left(z^k - \frac{1}{\rho}\lambda_i^{k^\top} \right) \end{aligned}$$

$$z^{k+1} = \arg\min_z \left(\frac{1}{2}\|z-c\|^2 + \sum_{i=1}^m \lambda_i^{k\top}(x_i^{k+1}-z) + \frac{\rho}{2}\sum_{i=1}^m \|x_i^{k+1}-z\|^2 \right) := \arg\min_z \varphi(z)$$

$$\nabla\varphi(z) = (z-c) - \sum_{i=1}^m \lambda_i^{k\top} + \rho\sum_{i=1}^m (z - x_i^{k+1}) = (1+m\rho)z - c - \sum_{i=1}^k \left(\lambda_i^k + \rho x_i^{k+1}\right) = 0$$

$$\Longrightarrow z^{k+1} = \frac{c + \sum_{i=1}^k \left(\lambda_i^k + \rho x_i^{k+1}\right)}{1+m\rho}, \lambda_i^{k+1} = \lambda_i^k + \tau\rho(x_i^{k+1} - z_i^{k+1}), \tau \in \left(0, \frac{1+\sqrt{5}}{2}\right]$$

这个算法是收敛的。

3. 相关系数矩阵的逼近问题的定义为：

$$\begin{aligned} \min \quad & \frac{1}{2}\|X-G\|_F^2 \\ \text{s.t.} \quad & X_{ii} = 1, i = 1,2,\cdots,n \\ & X \succeq 0 \end{aligned}$$

其中自变量 X 取值于对称矩阵空间 S_n，G 为给定的实对称矩阵。这个问题在金融领域中有重要的应用。由于误差等因素，根据实际观测得到的相关系数矩阵的估计 G 往往不具有相关系数矩阵的性质（如对角线为 1，正定性），我们的最终目标是找到一个和 G 最接近的相关系数矩阵 X。试给出满足如下要求的算法：

(a). 对偶近似点梯度法，并给出化简后的迭代公式；

(b). 针对原始问题的 ADMM，并给出每个子问题的显式解。

解

(a). Lagrange 函数为 $L(X,y,Z) = \frac{1}{2}\|X-G\|_F^2 - y^\top(\mathrm{diag}(X) - \mathbb{1}) - \mathrm{tr}(ZX)$。

又 $y^\top \mathrm{diag}(X) = \mathrm{tr}(\mathrm{Diag}(Y)X)$，其中 $Diag(y)$ 是对角元素为 y 的元素的对角矩阵，

故 $L(X,y,Z) = \frac{1}{2}\|X-G\|_F^2 - \mathrm{tr}(\mathrm{Diag}(Y)X) - \mathbb{1}^\top y - \mathrm{tr}(ZX)$。

$\nabla_X L(X,y,Z) = X - G - \mathrm{Diag}(y) - Z = 0 \Rightarrow X = G + Z + \mathrm{Diag}(Y)$，

代回 $L(X,y,Z)$ 得到对偶问题目标函数：

$$\begin{aligned} g(y,Z) =& \frac{1}{2}\|G + \mathrm{Diag}(y) + Z - G\|_F^2 - \mathrm{tr}(\mathrm{Diag}(y)(G + \mathrm{Diag}(y) + Z)) + y^\top\mathbb{1} \\ & - \mathrm{tr}(Z(G + \mathrm{Diag}(y) + Z)) \\ =& \frac{1}{2}\|\mathrm{Diag}(y) + Z\|_F^2 - \mathrm{tr}(\mathrm{Diag}(y)G) - \mathrm{tr}(\mathrm{Diag}(y)^2) - \mathrm{tr}(\mathrm{Diag}(y)Z) + y^\top\mathbb{1} \\ & - \mathrm{tr}(ZG) - \mathrm{tr}(Z\,\mathrm{Diag}(y)) - \mathrm{tr}(Z^2) \\ =& -\frac{1}{2}\|\mathrm{Diag}(y)\|_F^2 - \frac{1}{2}\|Z\|_F^2 - \mathrm{tr}(\mathrm{Diag}(y)Z) - \mathrm{tr}(\mathrm{Diag}(y)G) - \mathrm{tr}(ZG) \\ & + y^\top\mathbb{1} \\ =& -\frac{1}{2}\|\mathrm{Diag}(y) + Z\|_F^2 - \mathrm{tr}(G(\mathrm{Diag}(y) + Z)) + y^\top\mathbb{1} \\ =& -\frac{1}{2}\|\mathrm{Diag}(y) + Z + G\|_F^2 + \frac{1}{2}\|G\|_F^2 + y^\top\mathbb{1} \end{aligned}$$

故对偶问题是 $\max\limits_{Z\succeq 0,y}\left(-\frac{1}{2}\|\operatorname{Diag}(y)+Z+G\|_F^2+\frac{1}{2}\|G\|_F^2+y^\top\mathbb{1}\right)$。

等价写为 $\min\limits_{y,Z}(f(y,Z)+h(Z)):=\min\limits_{y,Z}\left(\frac{1}{2}\|\operatorname{Diag}(y)+Z+G\|_F^2-\mathbb{1}^\top y+\mathcal{I}_{S_+^n}(Z)\right)$

此时迭代可写为 $(y^{k+1},Z^{k+1})=\operatorname{prox}_{\alpha_k h}\left((y^k,Z^k)-\alpha_k\nabla f(y^k,Z^k)\right)$，$\alpha_k$ 是步长。

先求 ∇f，记 $\Lambda^k=G+\operatorname{Diag}(y^k)+Z^k, \nabla_y f(y^k,Z^k)=\operatorname{diag}(\Lambda^k)-\mathbb{1}$，$\nabla_Z f(y^k,Z^k)=\Lambda^k$。

记 $y_{mid}^{k+1}=y^k-\alpha_k(\operatorname{diag}(\Lambda^k)-\mathbb{1}), Z_{mid}^{k+1}=Z^k-\alpha_k\Lambda^k$，考虑临近点算子的作用：

由于 h 只和 Z 有关，故 $y^{k+1}=y_{mid}^{k+1}, Z^{k+1}=\operatorname{prox}_{\alpha_k h}(Z_{mid}^{k+1})=\mathcal{P}_{S_+^n}(Z_{mid}^{k+1})$。

即 $Z_{mid}^{k+1}=V\operatorname{diag}(\lambda_1,\cdots,\lambda_n)V^\top$ 时 $Z^{k+1}=V\operatorname{diag}(\max\{0,\lambda_1\},\cdots,\max\{0,\lambda_n\})V^\top$。

以上给出了一个完整的对偶近似点梯度迭代公式。

(b). 把问题重写为以下形式：

$$\begin{aligned}\min_{X,Z}\quad & \frac{1}{2}\|X-G\|_F^2+\mathcal{I}_{\operatorname{diag}(X)=\mathbb{1}}(X)+\mathcal{I}_{S_+^n}(Z)\\ \text{s.t.}\quad & X-Z=0\end{aligned}$$

此时增广 Lagrange 函数为

$$L(X,Z,\Lambda)=\frac{1}{2}\|X-G\|_F^2+\mathcal{I}_{\operatorname{diag}(X)=\mathbb{1}}(X)+\mathcal{I}_{S_+^n}(Z)+\operatorname{tr}(\Lambda^\top(X-Z))+\frac{\rho}{2}\|X-Z\|_F^2$$

其中 Λ 是 Lagrange 乘子，ρ 是正的罚参数。对 X,Z 做交替极小化：

X_{ii}^{k+1} 直接取为 1，只需 $\arg\min\limits_X\left(\frac{1}{2}\|X-G\|_F^2+\operatorname{tr}(\Lambda^{k\top}(X-Z^k))+\frac{\rho}{2}\|X-Z^k\|_F^2\right)$。

又由于展开式中所有 X_{ij} 均为变量分离形式，故只需按以下方式更新：

$$X_{ij}^{k+1}=\arg\min_{X_{ij}}\left(\frac{1}{2}(X_{ij}-G_{ij})^2+\Lambda_{ij}^kX_{ij}+\frac{\rho}{2}(X_{ij}-Z_{ij}^k)^2\right)=\frac{G_{ij}+\rho Z_{ij}^k-\Lambda_{ij}^k}{1+\rho}$$

$$\begin{aligned}Z^{k+1}&=\arg\min_Z\left(\mathcal{I}_{S_+^n}(Z)+\operatorname{tr}(\Lambda^\top(X^{k+1}-Z))+\frac{\rho}{2}\|X^{k+1}-Z\|_F^2\right)\\&=\arg\min_Z\left(\mathcal{I}_{S_+^n}(Z)+\frac{\rho}{2}\left\|X^{k+1}-Z+\frac{\Lambda^k}{\rho}\right\|_F^2\right)=\mathcal{P}_{S_+^n}\left(X^{k+1}+\frac{\Lambda^k}{\rho}\right)\end{aligned}$$

若有特征值分解 $X^{k+1}+\frac{\Lambda^k}{\rho}=V\operatorname{diag}(\lambda_1,\cdots,\lambda_n)V^\top$，则 Z 的更新为：

$Z^{k+1}=V\operatorname{diag}(\max\{0,\lambda_1\},\cdots,\max\{0,\lambda_n\})V^\top,\Lambda^{k+1}=\Lambda^k+\rho(X^{k+1}-Z^{k+1})$。

4. 给定算子 $A:\mathbb{R}^n\to\mathbb{R}^n$，称 A 是单调算子，如果其满足：

$$\langle Ax-Ay,x-y\rangle\geq 0,\forall x,y$$

证明：

(a). 给定闭凸函数 $f(x)$，次微分算子 ∂f 是单调算子。

(b). 给定闭凸函数 $f(x)$，若 $f(x)$ 是 μ-强凸的，那么 ∂f 是强单调算子，即：

$$\langle\partial f(x)-\partial f(y),x-y\rangle\geq\mu\|x-y\|^2,\forall x,y$$

(c). A 是非扩张的，等价于 $\frac{1}{2}(I+A)$ 是固定非扩张的。

(d). 给定凸函数 $f(x)$，且 $f(x)$ 一阶光滑，$\nabla f(x)$ 是 L-Lipchitz 连续的。定义 $G = I - t\nabla f$，$t \in \left(0, \frac{1}{L}\right]$

- 验证 G 是固定非扩张的。
- 若 $f(x)$ 是 μ-强凸的，那么 G 是压缩算子，即存在 $\rho \in (0,1)$ 使得

$$\|G(x) - G(y)\| \le \rho\|x - y\|$$

解

(a). $\forall x, y \in \mathrm{dom}\, f, g_x \in \partial f(x), g_y \in \partial f(y)$，由次微分定义，有
$f(y) \ge f(x) + \langle g_x, y - x\rangle, f(x) \ge f(y) + \langle g_y, x - y\rangle$，
相加，$0 \ge \langle g_x, y-x\rangle + \langle g_y, x-y\rangle = \langle g_x - g_y, y - x\rangle$，两侧乘 -1 即为单调算子定义。

(b). 此时 $f(y) \ge f(x) + \langle g_x, y - x\rangle + \frac{\mu}{2}\|y - x\|^2, f(x) \ge f(y) + \langle g_y, x - y\rangle + \frac{\mu}{2}\|x - y\|^2$
仍然相加，$0 \ge \langle g_x, y-x\rangle + \langle g_y, x-y\rangle + \mu\|x-y\|^2 = -\langle g_y - g_x, y - x\rangle + \mu\|x-y\|^2$，
$\Rightarrow \langle g_y - g_x, y - x\rangle \ge \mu\|x-y\|^2$，即为强单调算子定义。

(c). $\Longrightarrow$: 记 $T = \frac{1}{2}(I+A), u = Tx, v = Ty, A = 2T-I, \|(2T-I)x-(2T-I)y\|^2 \le \|x-y\|^2$，
即 $\|2(u-v)-(x-y)\|^2 \le \|x-y\|^2 \Rightarrow 4\|u-v\|^2 - 4\langle u-v, x-y\rangle + \|x-y\|^2 \le \|x-y\|^2$，
$\Rightarrow \|u-v\|^2 \le \langle u-v, x-y\rangle$，即 $\|Tx - Ty\|^2 \le \langle Tx - Ty, x - y\rangle$，故 T 固定非扩张。

$\Longleftarrow$: 由 $\|u-v\|^2 \le \langle u-v, x-y\rangle$ 有 $4\|u-v\|^2 \le 4\langle u-v, x-y\rangle$，此时
$\Rightarrow 4\|u-v\|^2 - 4\langle u-v, x-y\rangle + \|x-y\|^2 \le \|x-y\|^2 \Rightarrow \|2(u-v)-(x-y)\|^2 \le \|x-y\|^2$
$\Rightarrow \|(2T-I)x - (2T-I)y\|^2 = \|Ax - Ay\|^2 \le \|x-y\|^2$，即 A 非扩张。

(d).
- 由 ∇f 是 L-Lipchitz 连续的，有 $\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2$。
记 $H = 2G - I = I - 2t\nabla f$，则有

$$\begin{aligned}
\|H(x) - H(y)\|^2 =& \|x - y - 2t(\nabla f(x) - \nabla f(y))\|^2 \\
=& \|x-y\|^2 - 4t\langle \nabla f(x) - \nabla f(y), x - y\rangle + 4t^2\|\nabla f(x) - \nabla f(y)\|^2 \\
\le& \|x-y\|^2 - \frac{4t}{L}\|\nabla f(x) - \nabla f(y)\|^2 + 4t^2\|\nabla f(x) - \nabla f(y)\|^2 \\
=& \|x-y\|^2 + 4(t^2 - \frac{t}{L})\|\nabla f(x) - \nabla f(y)\|^2 \le \|x-y\|^2
\end{aligned}$$

故 H 是非扩张的，进而 G 是固定非扩张的。

- 由 f 是 μ-强凸的，有 $\langle \nabla f(x) - \nabla f(y), x - y\rangle \ge \mu\|x-y\|^2$。

$$\begin{aligned}
\|G(x) - G(y)\| =& \|x - y - t(\nabla f(x) - \nabla f(y))\|^2 \\
=& \|x-y\|^2 - 2t\langle \nabla f(x) - \nabla f(y), x - y\rangle + t^2\|\nabla f(x) - \nabla f(y)\|^2 \\
\le& \|x-y\|^2 - 2t\mu\|x-y\|^2 + t^2L^2\|x-y\|^2 \\
=& (L^2t^2 - 2\mu t + 1)\|x-y\|^2 \le \|x-y\|^2
\end{aligned}$$

故 G 是压缩算子。