最优化作业 2

龚舒凯 2022202790

1. 考虑采用精确线搜索的最速下降法，用于强凸二次函数。证明当初始点 $\mathbf{x}_0$ 满足 $\mathbf{x}_0-\mathbf{x}^*$ 平行于 Hessian 矩阵 $\mathbf{Q}$ 的特征向量时，最速下降法只需要一步就可以找到最小点。(20 分)

证明. 设强凸二次函数为 $f(\mathbf{x})=\frac{1}{2}\mathbf{x}^\top\mathbf{G}\mathbf{x}+\mathbf{b}^\top\mathbf{x}+\frac{m}{2}\mathbf{x}^\top\mathbf{x}$，其中 $\mathbf{G}$ 为正定矩阵，$\mathbf{b}$ 为常向量，m 为强凸参数。先重写强凸二次函数为

$$f(\mathbf{x})=\frac{1}{2}\mathbf{x}^\top\mathbf{G}\mathbf{x}+\mathbf{b}^\top\mathbf{x}+\frac{m}{2}\mathbf{x}^\top\mathbf{x}=\frac{1}{2}\mathbf{x}^\top(\mathbf{G}+m\mathbf{I})\mathbf{x}+\mathbf{b}^\top\mathbf{x}\triangleq\frac{1}{2}\mathbf{x}^\top\mathbf{Q}\mathbf{x}+\mathbf{b}^\top\mathbf{x} \tag{1}$$

显然 $\mathbf{Q}=\mathbf{G}+m\mathbf{I}$ 是正定矩阵。由题意，$\mathbf{x}_0-\mathbf{x}^*$ 是 $\mathbf{Q}$ 的特征向量，那么 $\exists\lambda\in\mathbb{R}$，

$$\mathbf{Q}(\mathbf{x}_0-\mathbf{x}^*)=\lambda(\mathbf{x}_0-\mathbf{x}^*) \tag{2}$$

在最小点 $\mathbf{x}^*$ 处，$\mathbf{g}(\mathbf{x})=\nabla f(\mathbf{x}^*)=\mathbf{Q}\mathbf{x}^*+\mathbf{b}=\mathbf{0}$，那么初始点 $\mathbf{x}_0$ 处的梯度可以表示为

$$\mathbf{g}_0=\mathbf{Q}\mathbf{x}_0+\mathbf{b}=\mathbf{Q}\mathbf{x}_0-\mathbf{Q}\mathbf{x}^*=\lambda(\mathbf{x}_0-\mathbf{x}^*) \tag{3}$$

最速下降法的下降方向为 $\mathbf{d}_0=-\mathbf{g}_0$，根据精确线搜索可确定搜索步长为

$$\alpha_0=\arg\min_\alpha f(\mathbf{x}_0-\alpha\mathbf{g}_0)=\frac{\mathbf{g}_0^\top\mathbf{g}_0}{\mathbf{g}_0^\top\mathbf{Q}\mathbf{g}_0}=\frac{\lambda^2(\mathbf{x}_0-\mathbf{x}^*)^\top(\mathbf{x}_0-\mathbf{x}^*)}{\lambda^2(\mathbf{x}_0-\mathbf{x}^*)^\top\mathbf{Q}(\mathbf{x}_0-\mathbf{x}^*)} \tag{4}$$

$$=\frac{\lambda^2(\mathbf{x}_0-\mathbf{x}^*)^\top(\mathbf{x}_0-\mathbf{x}^*)}{\lambda^3(\mathbf{x}_0-\mathbf{x}^*)^\top(\mathbf{x}_0-\mathbf{x}^*)}=\frac{1}{\lambda} \tag{5}$$

因此最速下降法走过一步后，

$$\mathbf{x}_1=\mathbf{x}_0-\alpha_0\mathbf{g}_0=\mathbf{x}_0-\frac{1}{\lambda}\lambda(\mathbf{x}_0-\mathbf{x}^*)=\mathbf{x}^* \tag{6}$$

因此在题设条件下，最速下降法只需要一步就可以找到最小点。 □

2. 假设 $\mathbf{Q}$ 是一个正定对称矩阵，证明如下 Kantorovich 不等式

$$\frac{(\mathbf{x}^\top\mathbf{x})^2}{(\mathbf{x}^\top\mathbf{Q}\mathbf{x})(\mathbf{x}^\top\mathbf{Q}^{-1}\mathbf{x})}\geq\frac{4\lambda_n\lambda_1}{(\lambda_n+\lambda_1)^2} \tag{7}$$

其中，λ_n,λ_1 分别为 $\mathbf{Q}$ 的最大、最小特征值，$\mathbf{x}$ 为任意非 $\mathbf{0}$ 向量。(该题目有一定的难度)(20 分)

证明. 由于 $\mathbf{Q}$ 为正定对称阵，可对其作特征值分解 $\mathbf{Q}=\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$，其中 $\mathbf{U}=\begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n\end{bmatrix}$ 为正交矩阵，$\mathbf{\Lambda}$ 为对角矩阵，对角元素为 $\mathbf{Q}$ 的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$。于是要证的不等式可转化为

$$\frac{(\mathbf{x}^\top\mathbf{x})^2}{((\mathbf{U}^\top\mathbf{x})^\top\mathbf{\Lambda}\mathbf{U}^\top\mathbf{x})((\mathbf{U}^\top\mathbf{x})^\top\mathbf{\Lambda}^{-1}\mathbf{U}^\top\mathbf{x})}\xlongequal{\mathbf{y}=\mathbf{U}^\top\mathbf{x}}\frac{(\mathbf{y}^\top(\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{y})^2}{(\mathbf{y}^\top\mathbf{\Lambda}\mathbf{y})(\mathbf{y}^\top\mathbf{\Lambda}^{-1}\mathbf{y})} \tag{8}$$

$$=\frac{\left(\sum\limits_{i=1}^n y_i^2\right)^2}{\left(\sum\limits_{i=1}^n\lambda_i y_i^2\right)\left(\sum\limits_{i=1}^n\lambda_i^{-1}y_i^2\right)}\geq\frac{4\lambda_n\lambda_1}{(\lambda_n+\lambda_1)^2} \tag{9}$$

记 $p_i = \dfrac{y_i^2}{\sum\limits_{i=1}^n y_i^2} \geq 0$，显然 $\sum\limits_{i=1}^n p_i = 1$，要证明的不等式可进一步转化为

$$\frac{1}{\left(\sum\limits_{i=1}^n \lambda_i p_i\right)\left(\sum\limits_{i=1}^n \lambda_i^{-1} p_i\right)} \geq \frac{4\lambda_n\lambda_1}{(\lambda_n+\lambda_1)^2} \tag{10}$$

记 $f(\mathbf{p}) = \left(\sum\limits_{i=1}^n \lambda_i p_i\right)\left(\sum\limits_{i=1}^n \lambda_i^{-1} p_i\right)$，我们先证明一个引理：

$$\left(\sum_{i=1}^n \lambda_i p_i\right)\left(\sum_{i=1}^n \lambda_i^{-1} p_i\right) = 1 - \sum_{i=1}^n p_i(\lambda_i - \sum_{i=1}^n \lambda_i p_i)(\lambda_i^{-1} - \sum_{i=1}^n \lambda_i^{-1} p_i) \tag{11}$$

记 $A = \sum\limits_{i=1}^n \lambda_i p_i$，$B = \sum\limits_{i=1}^n \lambda_i^{-1} p_i$，那么

$$\begin{aligned}
(11) \iff AB &= 1 - \sum_{i=1}^n p_i(\lambda_i - A)(\lambda_i^{-1} - B) && (12)\\
\iff AB &= 1 - \sum_{i=1}^n \left(\lambda_i\lambda_i^{-1}p_i - \lambda_i B p_i - A\lambda_i^{-1}p_i - ABp_i\right) && (13)\\
\iff AB &= 1 - \sum_{i=1}^n p_i + B\sum_{i=1}^n \lambda_i p_i + A\sum_{i=1}^n \lambda_i^{-1}p_i + AB\sum_{i=1}^n p_i && (14)\\
\iff AB &= 1 - 1 + AB + AB - AB && (15)\\
\iff AB &= AB \qquad \text{于是引理显然得证} && (16)
\end{aligned}$$

再对引理使用 Cauchy 不等式：

$$\begin{aligned}
f(\mathbf{p}) &= 1 - \sum_{i=1}^n p_i(\lambda_i - \sum_{i=1}^n \lambda_i p_i)(\lambda_i^{-1} - \sum_{i=1}^n \lambda_i^{-1} p_i) && (17)\\
&\leq 1 + \sqrt{\sum_{i=1}^n p_i(\lambda_i - \sum_{i=1}^n \lambda_i p_i)^2}\sqrt{\sum_{i=1}^n p_i(\lambda_i^{-1} - \sum_{i=1}^n \lambda_i^{-1} p_i)^2} && (18)
\end{aligned}$$

又注意到

$$\begin{aligned}
\sum_{i=1}^n p_i(\lambda_i - \sum_{i=1}^n \lambda_i p_i)^2 = \min_{a\in\mathbb{R}} \sum_{i=1}^n p_i(\lambda_i - a)^2 &\underbrace{\leq}_{\text{取}a=\frac{\lambda_1+\lambda_n}{2}} \sum_{i=1}^n p_i(\lambda_i - \frac{\lambda_1+\lambda_n}{2})^2 && (19)\\
&\leq \sum_{i=1}^n p_i(\lambda_n - \frac{\lambda_1+\lambda_n}{2})^2 = (\frac{\lambda_n-\lambda_1}{2})^2 && (20)
\end{aligned}$$

因此

$$f(\mathbf{p}) \leq 1 + \frac{\lambda_n - \lambda_1}{2}\cdot\frac{\lambda_1^{-1}-\lambda_n^{-1}}{2} = 1 + \frac{(\lambda_n-\lambda_1)^2}{4\lambda_n\lambda_1} = \frac{(\lambda_n+\lambda_1)^2}{4\lambda_1\lambda_n} \tag{21}$$

从而不等式 (9)

$$\frac{1}{f(\mathbf{p})} \geq \frac{4\lambda_1\lambda_n}{(\lambda_n+\lambda_1)^2} \tag{22}$$

这证明了 Kantorovich 不等式。 □

3. 拟牛顿条件为 $\mathbf{B}_{k+1}\mathbf{s}_k = \mathbf{y}_k$ ，为了保证 $\mathbf{B}_{k+1}$ 正定，要求 $\mathbf{s}_k^\top \mathbf{y}_k > 0$, 证明当目标函数 f 强凸时，该条件一定满足。(20 分)

证明. 规定 $\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k$，$\mathbf{y}_k = \nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)$。先指出一个引理：当 f 为凸函数时，成立 $\mathbf{s}_k^\top \mathbf{y}_k \geq 0$。证明如下：由 f 的凸性，有

$$\begin{cases} f(\mathbf{x}_{k+1}) \geq f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^\top (\mathbf{x}_{k+1} - \mathbf{x}_k) \\ f(\mathbf{x}_k) \geq f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})^\top (\mathbf{x}_k - \mathbf{x}_{k+1}) \end{cases} \Rightarrow 0 \geq (\nabla f(\mathbf{x}_k) - \nabla f(\mathbf{x}_{k+1}))^\top (\mathbf{x}_{k+1} - \mathbf{x}_k) \tag{23}$$

$$\Rightarrow -\mathbf{y}_k^\top \mathbf{s}_k \leq 0 \Rightarrow \mathbf{s}_k^\top \mathbf{y}_k \geq 0 \tag{24}$$

当 f 为强凸函数时，存在强凸参数 $m > 0$，使得 $g(\mathbf{x}) = f(\mathbf{x}) - \frac{m}{2}\mathbf{x}^\top \mathbf{x}$ 为凸函数。对 $g(\mathbf{x})$ 求梯度得 $\nabla g(\mathbf{x}) = \nabla f(\mathbf{x}) - m\mathbf{x}$。由于 g 为凸函数，根据引理

$$(\mathbf{x}_{k+1} - \mathbf{x}_k)^\top (\nabla g(\mathbf{x}_{k+1}) - \nabla g(\mathbf{x}_k)) \tag{25}$$
$$=(\mathbf{x}_{k+1} - \mathbf{x}_k)^\top (\nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k) - m(\mathbf{x}_{k+1} - \mathbf{x}_k)) \tag{26}$$
$$=(\mathbf{x}_{k+1} - \mathbf{x}_k)^\top (\nabla f(\mathbf{x}_{k+1}) - \nabla f(\mathbf{x}_k)) - m\|\mathbf{x}_{k+1} - \mathbf{x}_k\|^2 \geq 0 \tag{27}$$

从而 $\mathbf{s}_k^\top \mathbf{y}_k \geq m\|\mathbf{s}_k\|^2 > 0$，即当目标函数 f 强凸时，拟牛顿条件一定满足。 □

4. 证明采用强 Wolfe 准则，可以保证 $\mathbf{s}_k^\top \mathbf{y}_k > 0$，即可以保证 $\mathbf{B}_{k+1}$ 的正定性。(上课证明了精确线搜索和 Wolfe，这里要求证明强 Wolfe）(10 分)

证明. 由强 Wolfe 准则的第二个条件知

$$|g(\mathbf{x}_k + \alpha_k \mathbf{d}_k)^\top \mathbf{d}_k| = |\mathbf{g}_{k+1}| \leq -\sigma \mathbf{g}_k^\top \mathbf{d}_k, \quad \sigma \in (0,1) \tag{28}$$

于是

$$\mathbf{s}_k^\top \mathbf{y}_k = \alpha_k \mathbf{d}_k^\top (\mathbf{g}_{k+1} - \mathbf{g}_k) = \alpha_k (\mathbf{g}_{k+1}^\top \mathbf{d}_k - \mathbf{g}_k^\top \mathbf{d}_k) > \alpha_k (\sigma - 1)\mathbf{g}_k^\top \mathbf{d}_k > 0 \tag{29}$$

□

5. $f(\mathbf{x})$ 为正定二次函数，采用 SR1 方法进行优化，假设在迭代过程中一直满足 $(\mathbf{s}_k - \mathbf{H}_k \mathbf{y}_k)^\top \mathbf{y}_k > 0$，证明：

$$\mathbf{H}_k \mathbf{y}_j = \mathbf{s}_j,\ j = 0, 1, \cdots, k-1 \tag{30}$$

（该题目试图证明 SR1 方法更新出来的 $\mathbf{H}_k$ 不但和前一步的 $\mathbf{y}_{k-1}, \mathbf{s}_{k-1}$ 满足拟牛顿条件，而且还和之前所有步的 $\mathbf{y}_j, \mathbf{s}_j$ 都满足拟牛顿条件。事实上在共轭梯度章节的 Broyden 族方法的共轭特性介绍中，我们也得到了相似的结论 (20 分)

证明. 使用数学归纳法证明：

1. 当 $k = 1$ 时，$\mathbf{H}_1 \mathbf{y}_0 = \mathbf{s}_0$ 显然成立 (拟牛顿条件)。

2. 设当 $k = n$ 时，$\mathbf{H}_n \mathbf{y}_j = \mathbf{s}_j$，$j = 0, 1, \cdots, n-1$ 成立，那么当 $k = n+1$ 时

(1) 若 $i=n$，$\mathbf{H}_{n+1}\mathbf{y}_n=\mathbf{s}_n$(拟牛顿条件) 显然成立。
(2) 若 $i\in\{0,\cdots,n-1\}$，则由 SR1 法的更新公式，

$$\mathbf{H}_{n+1}\mathbf{y}_i=\mathbf{H}_n\mathbf{y}_i+\frac{(\mathbf{s}_n-\mathbf{H}_n\mathbf{y}_n)(\mathbf{s}_n-\mathbf{H}_n\mathbf{y}_n)^\top\mathbf{y}_i}{(\mathbf{s}_n-\mathbf{H}_n\mathbf{y}_n)^\top\mathbf{y}_n} \tag{31}$$

$$=\mathbf{H}_n\mathbf{y}_i+\frac{(\mathbf{s}_n-\mathbf{H}_n\mathbf{y}_n)(\mathbf{s}_n^\top\mathbf{y}_i-\mathbf{y}_n^\top\mathbf{H}_n^\top\mathbf{y}_i)}{(\mathbf{s}_n-\mathbf{H}_n\mathbf{y}_n)^\top\mathbf{y}_n} \tag{32}$$

其中

$$\mathbf{s}_n^\top\mathbf{y}_i-\mathbf{y}_n^\top\mathbf{H}_n^\top\mathbf{y}_i=(\mathbf{G}^{-1}\mathbf{y}_n)^\top\mathbf{y}_i-\mathbf{y}_n^\top\mathbf{s}_i=\mathbf{y}_n^\top\mathbf{G}^{-1}\mathbf{y}_i-\mathbf{y}_n^\top\mathbf{s}_i \tag{33}$$

$$=\mathbf{y}_n^\top\mathbf{s}_i-\mathbf{y}_n^\top\mathbf{s}_i=0 \tag{34}$$

因此 $\mathbf{H}_{n+1}\mathbf{y}_i=\mathbf{H}_n\mathbf{y}_i=\mathbf{s}_i$
3. 即当 $k=n+1$ 时，$\mathbf{H}_{n+1}\mathbf{y}_j=\mathbf{s}_j$，$j=0,1,\cdots,n$ 成立。由数学归纳法，结论得证。 □

6. 对于一个正定二次函数 $\phi(\mathbf{x})=\frac{1}{2}\mathbf{x}^\top\mathbf{A}\mathbf{x}-\mathbf{b}^\top\mathbf{x}$, 假设我们已经得到了一套共轭方向 $\mathbf{p}_0,\cdots,\mathbf{p}_l$, 满足 $\mathbf{p}_i^\top\mathbf{A}\mathbf{p}_j=0$，$i\neq j$，我们按照如下的方式进行迭代：

$$\mathbf{x}_{k+1}=\mathbf{x}_k+\alpha_k\mathbf{p}_k \tag{35}$$

请证明：如果步长采用精确线搜索，那么 $\alpha_k=-\dfrac{\mathbf{g}_k^\top\mathbf{p}_k}{\mathbf{p}_k^\top\mathbf{A}\mathbf{p}_k}$，其中 $\mathbf{g}_k=\mathbf{A}\mathbf{x}_k-\mathbf{b}$。如果我们进一步假设这套共轭方法是按照线性共轭梯度法得到的，这个步长的公式是否可以进一步简化？(10 分)

证明. 首先

$$\phi(\mathbf{x}_k+\alpha_k\mathbf{p}_k)=\frac{1}{2}(\mathbf{x}_k+\alpha_k\mathbf{p}_k)^\top\mathbf{A}(\mathbf{x}_k+\alpha_k\mathbf{p}_k)-\mathbf{b}^\top(\mathbf{x}_k+\alpha_k\mathbf{p}_k) \tag{36}$$

$$=\frac{1}{2}\mathbf{x}_k^\top\mathbf{A}\mathbf{x}_k+\alpha_k\mathbf{p}_k^\top\mathbf{A}\mathbf{x}_k+\frac{1}{2}\alpha_k^2\mathbf{p}_k^\top\mathbf{A}\mathbf{p}_k-\mathbf{b}^\top(\mathbf{x}_k+\alpha_k\mathbf{p}_k) \tag{37}$$

$$\tag{38}$$

关于 α_k 求导得

$$\frac{\partial\phi(\mathbf{x}_k+\alpha_k\mathbf{p}_k)}{\partial\alpha_k}=\mathbf{p}_k^\top\mathbf{A}\mathbf{x}_k+\alpha_k\mathbf{p}_k^\top\mathbf{A}\mathbf{p}_k-\mathbf{b}^\top\mathbf{p}_k=0 \tag{39}$$

$$\Rightarrow\alpha_k=\frac{\mathbf{b}^\top\mathbf{p}_k-\mathbf{x}_k^\top\mathbf{A}\mathbf{p}_k}{\mathbf{p}_k^\top\mathbf{A}\mathbf{p}_k}=-\frac{(\mathbf{A}\mathbf{x}_k-\mathbf{b})^\top\mathbf{p}_k}{\mathbf{p}_k^\top\mathbf{A}\mathbf{p}_k}=-\frac{\mathbf{g}_k^\top\mathbf{p}_k}{\mathbf{p}_k^\top\mathbf{A}\mathbf{p}_k} \tag{40}$$

当共轭方向由线性共轭梯度法得到时，第 k 次迭代方向 $\mathbf{p}_k$ 满足

$$\mathbf{p}_k=-\mathbf{g}_k+\frac{\mathbf{g}_k^\top\mathbf{g}_k}{\mathbf{g}_{k-1}^\top\mathbf{g}_{k-1}}\mathbf{p}_{k-1} \tag{41}$$

且由子空间拓展定理知 $\mathbf{g}_k^\top\mathbf{p}_i=0$，$i=0,1,\cdots,k-1$。两边同时左乘 $\mathbf{g}_k^\top$，得

$$\mathbf{g}_k^\top\mathbf{p}_k=-\mathbf{g}_k^\top\mathbf{g}_k+\frac{\mathbf{g}_k^\top\mathbf{g}_k}{\mathbf{g}_{k-1}^\top\mathbf{g}_{k-1}}\mathbf{g}_k^\top\mathbf{p}_{k-1}=-\mathbf{g}_k^\top\mathbf{g}_k \tag{42}$$

那么步长可以化简为 $\alpha_k=\dfrac{\mathbf{g}_k^\top\mathbf{g}_k}{\mathbf{p}_k^\top\mathbf{A}\mathbf{p}_k}$ □