Homework 8

陈远洋

Problem 1

Assume C is a convex set, show that $x^* \in \text{argmin}_{y\in C} \frac{1}{2}\|x-y\|_2^2$ if and only if for any $y \in C$, $\langle x-x^*, y-x^*\rangle \le 0$.

证明. 先证明充分性：$x^* \in \text{argmin}_{y\in C} \frac{1}{2}\|x-y\|_2^2 \Rightarrow \forall y \in C\,, \langle x-x^*, y-x^*\rangle \le 0$

若 $x^* \in \text{argmin}_{y\in C} \frac{1}{2}\|x-y\|_2^2$, 则 $\forall y \in C$, 有：

$$\begin{aligned}\frac{1}{2}(\|x-y\|_2^2 - \|x-x^*\|_2^2) &= \frac{1}{2}(\|x-x^*+x^*-y\|_2^2 - \|x-x^*\|_2^2)\\ &= \frac{1}{2}(\langle x^*-y, x^*-y\rangle - 2\langle x-x^*, y-x^*\rangle)\\ &\ge 0\end{aligned}$$

而若 $\exists y_0 \in C\text{s.t.}\langle x-x^*, y_0-x^*\rangle > 0$, 取 $\alpha \in (0,1)$. 由 C 是凸集且 $x^*, y_0 \in C$ 故 $\forall \alpha \in (0,1)$: $x^*+\alpha(y_0-x^*) \in C$ 令 $g(\alpha) = \frac{1}{2}\|x-x^*+\alpha(y_0-x^*)\|_2^2 - \frac{1}{2}\|x-x^*\|_2^2$, 而进一步展开:

$$\begin{aligned}g(\alpha) &= \frac{\alpha^2}{2}\langle x^*-y_0, x^*-y_0\rangle - \alpha\langle x-x^*, y_0-x^*\rangle\\ &= \frac{\alpha}{2}(\alpha\|x^*-y_0\|_2^2 - 2\langle x-x^*, y_0-x^*\rangle)\end{aligned}$$

则 $\forall \alpha \in (0,1)\bigcap(0, \frac{\langle x-x^*, y_0-x^*\rangle}{\|x^*-y_0\|_2^2})$: $g(\alpha) < \frac{\alpha}{2}(\frac{\langle x-x^*, y_0-x^*\rangle}{\|x^*-y_0\|_2^2}\|x^*-y_0\|_2^2 - 2\langle x-x^*, y_0-x^*\rangle) = -\frac{\alpha}{2}\langle x-x^*, y_0-x^*\rangle < 0$ 这与 x^* 是最优解的条件矛盾，故不存在这样的 y_0。也即 $\forall y \in C\,, \langle x-x^*, y-x^*\rangle \le 0$

再来看必要性：$\forall y \in C\,, \langle x-x^*, y-x^*\rangle \le 0 \Rightarrow x^* \in \text{argmin}_{y\in C} \frac{1}{2}\|x-y\|_2^2$

若 $\forall y \in C$: $\langle x-x*, y-x^*\rangle \le 0$ 且 $\exists \bar{x} \ne x^* \in C$: $\|x-\bar{x}\|_2^2 < \|x-x^*\|_2^2$（即 x^* 并非最优点）那么有：

$$\begin{aligned}\|x-\bar{x}\|_2^2 - \|x-x^*\|_2^2 &= \langle x-x^*+x^*-\bar{x}, x-x^*+x^*-\bar{x}\rangle - \langle x-x^*, x-x^*\rangle\\ &= 2\langle x-x^*, x^*-\bar{x}\rangle + \|\bar{x}-x^*\|_2^2\\ &< 0\end{aligned}$$

可得 $\langle x-x^*, \bar{x}-x^*\rangle > \frac{1}{2}\|\bar{x}-x^*\|_2^2 > 0$ 矛盾！故不存在这样的 $\bar{x}$ 也即 $x^* \in \mathrm{argmin}_{y\in C} \frac{1}{2}\|x-y\|_2^2$. $\square$

Problem 2

Use the optimality conditions to solve the following problem:

$$\min_{x\in\mathbb{R}^n} f(x)$$
$$s.t.\ h(x) = 0,$$

where

(a) $f(x) = \|x\|_2^2$, $h(x) = \sum_{i=1}^n x_i - 1$.

(b) $f(x) = \sum_{i=1}^n x_i$, $h(x) = \|x\|_2^2 - 1$.

(c) $f(x) = \|x\|_2^2, h(x) = x^TQx - 1$, where Q is positive definite.

- (a) 由一阶必要条件 $\exists\lambda$, 对 $L(x;\lambda) = f(x) - \lambda h(x)$ 有：$\begin{cases} h(x) = \sum_{i=1}^n x_i - 1 = 0 \\ \nabla f(x) - \lambda\nabla h(x) = 2x - \lambda\vec{1} = 0 \end{cases}$
  联合两式解得 $x_i = \frac{1}{n}, i = 1, 2, \cdots n$. 而 $\nabla_{xx}^2 L(x) = 2I$,$\nabla h(x) = \vec{1}$, 代入二阶充分条件 $d^T\nabla_{xx}^2 L(x)d > 0, \forall d \neq 0 : d^T\nabla h(\frac{\vec{1}}{n}) = 0$ 成立，所以 $\frac{\vec{1}}{n}$ 是最优解。对应函数值为 $f(x) = \frac{1}{n}$

- (b) 由一阶必要条件 $\exists\lambda$, 对 $L(x;\lambda) = f(x) - \lambda h(x)$ 有：$\begin{cases} h(x) = \langle x, x\rangle - 1 = 0 \\ \nabla f(x) - \lambda\nabla h(x) = \vec{1} - 2\lambda x = 0 \end{cases}$
  联合两式解得 $x_i^2 = \frac{1}{n}, i = 1, 2, \cdots n$, 且 $x_1 = \cdots = x_n$. 而 $\nabla_{xx}^2 L(x) = -2\lambda I$,$\nabla h(x) = 2x$, 代入二阶充分条件 $d^T\nabla_{xx}^2 L(x)d > 0, \forall d \neq 0 : d^T\nabla h(x) = 0$. 解得 $\lambda < 0$，则可得 $x_i = \frac{-1}{\sqrt{n}}$ 为最优点。对应函数值为 $f(\frac{-\vec{1}}{\sqrt{n}}) = -\sqrt{n}$

- (c) 由一阶必要条件 $\exists\lambda$, 对 $L(x;\lambda) = f(x) - \lambda h(x)$ 有：$\begin{cases} h(x) = xQ^Tx - 1 = 0 \\ \nabla f(x) - \lambda\nabla h(x) = 2x - 2\lambda Qx = 0 \end{cases}$
  如果 $\lambda = 0$ 那么由第二个等式可推出 $x = 0$ 与第一式矛盾，所以 $\lambda \neq 0$ 所以原来的等式可以改写为: $\begin{cases} x^TQx = 1 \\ (Q - \frac{1}{\lambda}I)x = 0 \end{cases}$ 可以看出，这是一个特征值问题。代入 $f(x)$ 可得 $f(x) = x^Tx = x^T\lambda Qx = \lambda$. 而容易看出 $\frac{1}{\lambda}$ 为 Q 的特征值。要想求出 $f(x)$ 的最小值，我们需要 $\frac{1}{\lambda}$ 尽可能大。于是取 $\frac{1}{\lambda} = \|Q\|_2 = \rho(Q)$,$x^*$ 为对应的模长为 $\|Q\|_2^{-\frac{1}{2}}$ 特征向量。而 $\nabla_{xx}^2 L(x) = 2(I - \lambda Q)$,$\nabla h(x) = 2Qx$。而由于 $d \perp span\{x^*\}$，$d^TQd < d^Td\|Q\|_2$，代入二阶充分条件 $d^T\nabla_{xx}^2 L(x^*)d > 0, \forall d \neq 0 : d^T\nabla h(x^*)$ 成立. 进一步推出该点为最优点。对应函数值为 $f(x^*) = \frac{1}{\|Q\|_2}$

Problem 3

Consider a symmetric $n \times n$ matrix Q. Define

$$\lambda_1 = \min_{\|x\|_2^2=1} x^T Q x, \ e_1 \in \operatorname*{argmin}_{\|x\|_2^2=1} x^T Q x,$$

and for $k = 0, \cdots, n-1$,

$$\lambda_{k+1} = \min_{\substack{\|x\|_2^2=1 \\ e_i^T x=0, i=1,\dots,k}} x^T Q x, \quad e_{k+1} \in \arg \min_{\substack{\|x\|_2^2=1 \\ e_i^T x=0, i=1,\dots,k}} x^T Q x$$

(a) Show that

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

(b) Show that the vectors $e_1, \cdots, e_n$ are linearly independent.

(c) Interpret $\lambda_1, \lambda_2, \cdots, \lambda_n$ as Lagrange multipliers, and show that $\lambda_1, \lambda_2, \cdots, \lambda_n$ are the eigenvalues of Q, while $e_1, \cdots, e_n$ are corresponding eigenvectors.

证明. (a) 从直观上来看，由于越往后约束更多。所以最优点函数值一定越来越大。作以下严格证明：如果 $\exists k, \text{s.t. } \lambda_k > \lambda_{k+1}$ 那么 $e_{k+1}^T Q e_{k+1} = \lambda_{k+1} < \lambda_k$ 且: $\begin{cases} \|e_{k+1}\|_2 = 1 \\ e_j^T e_{k+1} = 0 \quad \forall j < k (k > 1) \end{cases}$ 即 e_{k+1} 是可行的。那么这与 λ_k 是子问题的最优解函数值矛盾！所以 $\forall k, \lambda_k \leq \lambda_{k+1}$ 即 $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$.

(b) 由第 k 步条件（$k > 1$）：$e_i^T x = 0, \forall i < k$ 并且 e_k 是可行的所以 $e_i^T e_k = 0, \forall i < k$。如果 $e_1, \cdots, e_n$ 不是线性无关。那么就 $\exists k$，s.t.e_k 可以由前 $k-1$ 个向量线性表出（i.e.k 可以取线性组合为 0 的最大不为 0 系数对应下标）。而这与正交性条件矛盾：$0 < e_k^T e_k = e_k^T (\sum_{i=1}^{k-1} \alpha_i e_i) = 0$. 所以 $e_1, \cdots, e_n$ 线性无关。

(c) 令 $f(x) = x^T Q x$, $h(x) = \|x\|_2^2 - 1 = \langle x, x \rangle - 1$ 定义问题 1：

$$\min_{x \in \mathbb{R}^n} f(x)$$
$$s.t. \ h(x) = 0$$

应用拉格朗日乘子 λ, 得到拉格朗日函数 $L(x; \lambda) = f(x) - \lambda h(x)$. 由一阶必要条件 $\nabla f(x) - \lambda \nabla h(x) = 2(Q - \lambda I)x = 0$ 化简得到 $(Q - \lambda I)x = 0$ 左乘 x 得到 $f(x) = \lambda$ 而这个等式是一个特征值问题，蕴含 λ, x 分别是 Q 的特征值与特征向量。此时 $\nabla_{xx}^2 L = 2(Q - \lambda I)$, 代入二阶充分条件 $d^T \nabla_{xx}^2 L d > 0, \forall d \neq 0 : d^T \nabla h(x) = 0$. 显然此时 λ 取 Q 最小的特征值可以使得函数值最优。对应本问题我们得到 λ_1 与 e_1 是 Q 的特征值与特征向量。

假设我们要证明的命题：λ_i, e_i 是 Q 的第 i 个特征对在前 k 对成立，采用数学归纳法，下证 $k+1$ 时的情况：在问题 k 的基础上定义问题 $k+1$

$$\min_{x \perp span\{e_1, \cdots, e_k\}} f(x)$$
$$s.t.\ h(x) = 0$$

引入乘子 $\mu_1, \cdots, \mu_k, \lambda$, 得到拉格朗日函数 $L(x;\lambda;\mu) = f(x) - \lambda h(x) - \sum_{i=1}^{k} \mu_i x^T e_i$. 由一阶必要条件 $\nabla f(x) - \lambda \nabla h(x) - \sum_{i=1}^{k} \mu_i e_i = 0$ 化简得到 $(Q - \lambda I)x = \sum_{i=1}^{k} \frac{\mu_i}{2} e_i$ 分别左乘 $e_i, i = 1, \cdots, k$ 得到 $e_i^T Q x = \lambda_i e_i^T x = 0 = \frac{\mu_i}{2}$. 再左乘 x 得到 $x^T Q x = \lambda x^T x = \lambda$ 由上面两个条件，必要条件进一步化简为: $(Q - \lambda I)x = 0$ 而这同样是一个特征值问题，蕴含 λ, x 分别是 Q 的特征值与特征向量。由于已经求出 $\mu_i = 0$, 此时 $\nabla_{xx}^2 L = 2(Q - \lambda I)$, 同样代入二阶充分条件 $d^T \nabla_{xx}^2 L d > 0, \forall d \neq 0 : d^T \nabla h(x) = 0$, 注意此时方向 d 的选取局限在前 k 个特征向量的补空间。显然此时 λ 取 Q 第 $k+1$ 最小的特征值可以使得函数值最优。(由于前面已经取过前 k 小特征对了) 故对应 λ_{k+1} 与 e_{k+1} 是 Q 的第 $k+1$ 小的特征值与特征向量。

综上所述，将 λ_i 作为 Lagrange 乘子的情况下，我们得到了 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 Q 的特征值, 而 $e_1, \cdots, e_n$ 是对应特征向量。 $\square$