最优化练习题

要求：请写出详细的推导过程或证明过程。

1. 考虑优化问题
$$\min_{x} \quad f(x) = x_1^2 + 2x_2^2 + 4x_1 + 4x_2.$$
   (a) 证明如果使用精确线搜索的最速下降梯度法从初始点$x^1 = (0,0)^\top$求解该问题，产生的点列为
$$x^{k+1} = \left(\frac{2}{3^k} - 2, \left(-\frac{1}{3}\right)^k - 1\right)^\top.$$
   (b) 这个迭代点列收敛到哪一个点？收敛速度是什么？

2. 考虑函数$f(x) = x_1^2 + x_2^2$，以及迭代点列$x^k = (1 + \frac{1}{2^k})(\cos k, \sin k)^T$，请说明
   (a) $\{f(x^{k+1})\}$是否收敛？若收敛，给出收敛速度。
   (b) $\{x^{k+1}\}$是否收敛？若收敛，给出收敛速度。

3. 考虑对称矩阵的特征值问题
$$\min_{x \in \mathbb{R}^n} \quad x^T A x, \quad \text{s.t.} \quad \|x\|_2 = 1,$$
   其中$A^T = A$。试分析其所有的局部极小点、鞍点以及全局极小点

4. 设可微函数$f(x)$的定义域$\mathbf{dom} f = \mathbb{R}^n$，如果梯度$\nabla f(x)$是$L$-利普希茨连续，即存在$L > 0$有
$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall\, x, y \in \mathbf{dom} f, \tag{1}$$
   证明函数$f(x)$有二次上界：
$$f(y) \le f(x) + \nabla f(x)^{\mathrm{T}}(y - x) + \frac{L}{2}\|y - x\|^2, \quad \forall\, x, y \in \mathbf{dom} f. \tag{2}$$

5. 设函数$f(x)$是$\mathbf{dom} f = \mathbb{R}^n$上的凸可微函数。假设梯度$\nabla f(x)$是$L$-利普希茨连续，即存在$L > 0$有
$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \quad \forall\, x, y \in \mathbf{dom} f, \tag{3}$$
   证明：$\nabla f(x)$ 有余强制性, 即对任意的$x, y \in \mathbb{R}^n$, 有
$$(\nabla f(x) - \nabla f(y))^{\mathrm{T}}(x - y) \ge \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2.$$

6. 假设可微函数$f(x)$的梯度$\nabla f(x)$是L-利普希茨连续，且f是可微强凸函数，即存在常数$m > 0$使得对任意$x, y \in \mathbb{R}^n$和$\theta \in (0,1)$，有：
$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y) - \frac{m}{2}\theta(1-\theta)\|x - y\|^2.$$
   证明:
$$(\nabla f(x) - \nabla f(y))^{\mathrm{T}}(x - y) \ge \frac{mL}{m+L}\|x - y\|^2 + \frac{1}{m+L}\|\nabla f(x) - \nabla f(y)\|^2.$$

7. 假设函数$f(x)$可微且梯度$\nabla f(x)$是L-利普希茨连续. 从初始步长$t=\hat{t}>0$开始考虑线搜索回退法. 如果不等式

$$f(x-t\nabla f(x)) < f(x)-\frac{t}{2}\|\nabla f(x)\|_2^2 \tag{4}$$

不满足，则缩小步长$t=\beta t$，其中$\beta\in(0,1)$，然后继续验证(4)是否满足. 证明回退法生成的步长满足$t\geq \min\{\hat{t},\beta/L\}$.

8. 假设函数$f(x)$可微且梯度$\nabla f(x)$是L-利普希茨连续. 考虑无约束优化问题：

$$\min_{x\in\mathbb{R}^n} \ f(x).$$

- 写出求解该问题的梯度法。
- 假设$f(x)$为非凸函数，明确给出梯度法步长的取法并写出收敛性结果.
- 假设$f(x)$为凸函数，明确给出梯度法步长的取法并写出的收敛性(复杂度分析)结果.
- 假设$f(x)$为强凸函数，明确给出梯度法步长的取法并写出的收敛性(复杂度分析)结果.

9. （考虑可微函数$f(x):\mathbb{R}^n\to\mathbb{R}$和点列$\{x^i\}_{i=1}^{\infty}$。令$k\geq 2$，$s^{k-1}=x^k-x^{k-1}$ 以及$y^{k-1}=\nabla f(x^k)-\nabla f(x^{k-1})$。

(a) 请分别计算如下两个优化问题的全局最优解:

$$\min_{\alpha\in\mathbb{R}} \quad \|\alpha y^{k-1}-s^{k-1}\|_2^2, \tag{5}$$

$$\min_{\alpha\in\mathbb{R}} \quad \|y^{k-1}-\alpha^{-1}s^{k-1}\|_2^2. \tag{6}$$

(b) 请给出问题(5)和(6)构造合理性的一些直观解释。并根据它们构造求解优化问题$\min_x \ f(x)$的梯度算法。

10. 给定实对称矩阵$A\in\mathcal{S}^n$, $b\in\mathbb{R}^n$, $\Delta>0$。考虑优化问题

$$\min_{x\in\mathbb{R}^n} \quad x^TAx+2b^Tx, \quad \text{s.t.} \quad \|x\|_2\leq\Delta.$$

给出并证明点$z\in\mathbb{R}^n$ 是该问题最优解的充要条件。

11. 考虑BFGS公式：

$$B_{k+1}=B_k-\frac{B_ks_ks_k^\top B_k}{s_k^\top B_ks_k}+\frac{y_ky_k^\top}{y_k^\top s_k}.$$

(a) 证明:

$$\mathrm{trace}(B_{k+1})=\mathrm{trace}(B_k)-\frac{\|B_ks_k\|_2^2}{s_k^\top B_ks_k}+\frac{\|y_k\|_2^2}{y_k^\top s_k}.$$

(b) 证明：

$$\det(B_{k+1})=\det(B_k)\frac{y_k^\top s_k}{s_k^\top B_ks_k}.$$

12. 假设x^*是最优化问题$\min_{x\in\mathbb{R}^n} \ f(x)$的最优解。假设函数$f(x)$二阶可微，海瑟矩阵$\nabla^2 f(x)$在$x^*$的一个领域内利普希茨连续且$\nabla^2 f(x^*)$严格正定。考虑牛顿算法

$$x_{k+1}=x_k+p, \quad p=-(\nabla^2 f(x_k))^{-1}\nabla f(x_k).$$

如果初始点x_0充分靠近x^*，证明迭代序列$\{x_k\}$收敛到x^*且收敛速度为二阶。

13. 设函数$f(x)=\|x\|^\beta$，其中$\beta>0$为给定的常数。考虑步长为1的牛顿法对$f(x)$ 进行极小化，初值$x^0\neq 0$。试证明：

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

(b) 若$0<\beta<1$，则牛顿法发散。

14. 给定向量$g \in \mathbb{R}^n$, 对称矩阵$B \in \mathbb{R}^{n\times n}$ 和信赖域半径Δ，考虑信赖域子问题：

$$\min_{p\in\mathbb{R}^n} \quad m(p) = g^\top p + \frac{1}{2}p^\top Bp, \quad \text{s.t.} \quad \|p\|_2 \le \Delta.$$

(a) 写出下面问题的最优解p^s：

$$p^s = \arg\min_{p\in\mathbb{R}^n} \quad g^\top p, \quad \text{s.t.} \quad \|p\|_2 \le \Delta.$$

(b) 写出下面问题的最优解τ：

$$\tau = \arg\min_{\tau\ge 0} \quad m(\tau p^s), \quad \text{s.t.} \quad \|\tau p^s\|_2 \le \Delta.$$

(c) 令科西步为$p^c = \tau p^s$。证明如下不等式：

$$m(0) - m(p^c) \ge \frac{1}{2}\|g\|_2 \min\left\{\Delta, \frac{\|g\|_2}{\|B\|_2}\right\}.$$

15. 给定连续可微函数$f(x)$ 和列向量$l, u \in \mathbb{R}^n$。考虑优化问题：

$$\min_{x\in\mathbb{R}^n} \ f(x), \ \text{s.t.} \ l \le x \le u. \tag{7}$$

(a) 给出点$x \in \mathbb{R}^n$到$[l, u]$的投影算子$P(x)$的显式表达式：

$$P(x) = \arg\min_{y} \|x - y\|_2^2, \ \text{s.t.} \ x \in [l, u].$$

(b) 对所有的$x \in \mathbb{R}^n$ 和任意$y \in [l, u]$，证明不等式$(P(x) - x)^\top (y - P(x)) \ge 0$成立。

(c) 对所有的$x, y \in \mathbb{R}^n$，证明不等式$(P(x) - P(y))^\top (x - y) \ge \|P(x) - P(y)\|_2^2$成立。

(d) 证明问题(7)的KKT条件等价于

$$x - P(x - \nabla f(x)) = 0.$$

(e) 写出求解问题(7)的投影梯度法，并给出一种全局收敛性结果。要求写清楚步长如何选取，以及保证该收敛性所需要的假设条件。

(f) 写出求解问题(7)的二次罚函数法。该二次罚函数是否可微？如果可微，给出其梯度的表达式。该二次罚函数是否二次可微？

(g) 写出求解问题(7)的增广拉格朗日函数法。要求写出: i) 具体的增广拉格朗日函数的表达式，并且该函数只有一个优化变量x。ii) 具体的乘子更新公式。

16. 给定对称矩阵$C \in \mathbb{R}^{n\times n}$，令$p$为正整数且$p \in [1, n]$，考虑优化问题:

$$\begin{aligned} \max_{Y\in\mathbb{R}^{n\times p}} \quad & f(Y) = \mathrm{Tr}(CYY^\mathrm{T}) \\ \text{s.t.} \quad & \mathrm{diag}(YY^\mathrm{T}) = \mathbf{1}, \end{aligned} \tag{8}$$

其中对于对称矩阵$X \in \mathbb{R}^{n\times n}$，$\mathrm{Tr}(X) = \sum_{i=1}^n X_{ii}, \mathrm{diag}(X) = [X_{11}, X_{22}, \cdots, X_{nn}]^\mathrm{T}$。

注： 如果对矩阵形式相对不熟悉，可以先将变量写成向量形式: $Y = \begin{pmatrix} y_1^\top \\ \cdots \\ y_n^\top \end{pmatrix}$，其中$y_i \in \mathbb{R}^p$是列向量.

(a) 讨论可行点Y处线性无关约束品性(LICQ)是否满足？

(b) 写出目标函数$f(Y)$关于Y的梯度，估计梯度的利普希茨常数。

(c) 优化问题(8)的一阶最优性(KKT)条件。

(d) 二阶充分条件：假设在可行点Y处存在拉格朗日乘子满足KKT条件，如果拉格朗日函数的海瑟矩阵在临界锥上严格正定，则Y是问题(8)的严格局部极小点。请写出拉格朗日函数的海瑟矩阵(可以是矩阵形式或者算子形式)和临界锥具体形式。

(e) 假设Y是局部极小点，写出拉格朗日乘子的显式表达形式(只与C和Y有关)。

(f) 给定矩阵B，计算如下投影算子的最优解：

$$P(B) = \arg\min_{Y\in\mathbb{R}^{n\times p}} \|Y-B\|_F^2, \text{ s.t. } \mathrm{diag}(YY^{\mathrm{T}}) = \mathbf{1}. \tag{9}$$

(g) 请根据(9)写出求解问题(8)的投影梯度法，并参考无约束优化的线搜索条件给出一种步长选取条件。简要讨论该算法的全局收敛性以及保证该收敛性所需要的假设条件。

17. 给定矩阵$A\in\mathbb{R}^{m\times n}$和向量$b\in\mathbb{R}^m$，$\mu>0$，考虑优化问题:

$$\min_x \quad \psi(x) = \mu\|x\|_1 + \frac{1}{2}\|Ax-b\|_2^2. \tag{10}$$

(a) 写出并验证$\psi(x)$的次梯度表达式。给出并证明问题(10)的最优性条件。

(b) 给出求解问题(10)的次梯度算法及其步长的取法，写出收敛性(复杂度分析)结果。

(c) 写出邻近算子的具体表达式:

$$\mathrm{prox}_t(y) = \arg\min_x \; t\|x\|_1 + \frac{1}{2}\|x-y\|_2^2.$$

(d) 给出求解问题(10)的近似点梯度算法及其步长的取法，写出收敛性(复杂度分析)结果。

(e) 给出求解问题(10)的Nesterov加速算法及其步长的取法，写出收敛性(复杂度分析)结果。

(f) 引入合适的拆分写出问题(10)的等价形式，写出求解该等价形式的增广拉格朗日函数法。要求明确写出如何求解增广拉格朗日函数子问题。

(g) 继上一小问: 写出求解该等价形式的交替方向乘子法(ADMM)。要求推导并明确写出每一个子问题的显式解。

(h) 写出问题(10)的对偶问题，写出求解该对偶问题的增广拉格朗日函数法。要求明确写出如何求解增广拉格朗日函数子问题。

(i) 继上一小问: 写出求解该对偶问题的交替方向乘子法(ADMM)。要求推导并明确写出每一个子问题的显式解。