HW3.

无约束优化

6.1 由于 $p^{(1)}(t)$、$p^{(2)}(t)$ 均为2次多项式

不妨设 $p^{(1)}(t)=A_1t^2+B_1t+C_1$　　$p^{(2)}(t)=A_2t^2+B_2t+C_2$

$p^{(1)}(t)$ 已知 φ_1、φ_1'、φ、

则 $\begin{cases} A_1\alpha_1^2+B_1\alpha_1+C_1=\varphi_1 \\ 2A_1\alpha_1+B_1=\varphi_1' \\ A_1\alpha^2+B_1\alpha+C_1=\varphi. \end{cases}$ 解得: $\begin{cases} A_1=\dfrac{\varphi_1-\varphi-\varphi_1'(\alpha_1-\alpha)}{(\alpha_1-\alpha)^2} \\ B_1=\dfrac{\varphi_1'(\alpha^2-\alpha_1)^2+2\alpha_1(\varphi_1-\varphi)}{(\alpha_1-\alpha)^2} \\ C_1=\dfrac{\varphi_1\alpha(\alpha-2\alpha_1)+\varphi_1'2\alpha_1(\alpha_1-\alpha)+\varphi\alpha_1^2}{(\alpha_1-\alpha)^2} \end{cases}$

$p^{(2)}(t)$ 已知 φ、φ'、φ_1'

则 $\begin{cases} A_2\alpha^2+B_2\alpha+C_1=\varphi \\ 2A_2\alpha+B_2=\varphi' \\ 2A_2\alpha_1+B_2=\varphi_1' \end{cases}$ 解得: $\begin{cases} A_2=\dfrac{\varphi'-\varphi_1'}{2\alpha-2\alpha_1} \\ B_2=\dfrac{\varphi_1'\alpha-\varphi'\alpha_1}{\alpha-\alpha_1} \\ C_2=\dfrac{\varphi(2\alpha-2\alpha_1)+\varphi'(2\alpha\alpha_1-\alpha^2)-\varphi_1'\alpha.}{2\alpha-2\alpha_1} \end{cases}$

6.2 要证明基于backtracking linesearch的非精确一维搜索算法

记所有的k，$g_k=\nabla f(x^k)$，$f_k=f(x^{(k)})$

由梯度李氏连续性可知. $\|g_{k+1}-g_k\|\le L\|x_{k+1}-x_k\|=L\alpha_k\|d^{(k)}\|$

考虑Armijo-Goldstein条件　$f(x^k+\gamma^t d^{(k)})>f(x^k)+c\gamma^t\nabla f(x^k)d^{(k)}$

$d^{(k)}=-\nabla f(x^k)$

$$\|-d^{(k+1)}+d^{(k)}\|\le L\alpha_k\|d^{(k)}\|$$

$$\text{由}\ \frac{\|d^{(k)}-d^{(k+1)}\|}{\|d^{(k)}\|}<L$$

有 $\lim\limits_{k\to+\infty}\|\nabla f(x^{(k)})\|=\lim\limits_{k\to\infty}\|-d^{(k)}\|=0.$

6.3 (1) 证明:

$$f(x) = -\left(\prod_{k=1}^{n} x_k\right)^{\frac{1}{n}} \quad (\text{for } x \in R_+^n) \text{ 是凸函数.}$$

$$\frac{\partial f(x)}{\partial x} = -\frac{1}{n}\left(\prod_{k=1}^{n} x_k\right)^{\frac{1-n}{n}} \left(\prod_{k\neq i} x_k\right)$$

$$\frac{\partial f(x)}{\partial x_i \partial x_j} = \begin{cases} -\frac{1-n}{n^2}\left(\prod_{k=1}^{n} x_k\right)^{\frac{1-2n}{n}} \cdot \left(\prod_{k\neq i} x_k\right), & i=j\text{时} \\ -\frac{1-n}{n^2}\left(\prod_{k=1}^{n} x_k\right)^{\frac{1-2n}{n}} \left(\prod_{k\neq i}^{n} x_k\right)\left(\prod_{k\neq j}^{n} x_k\right) - \frac{1}{n}\left(\prod_{k=1}^{n} x_k\right)^{\frac{1-n}{n}} \left(\prod_{k\neq i,\, k\neq j} x_k\right), & i\neq j\text{时} \end{cases}$$

其中 $i\neq j$ 时 $\frac{\partial f(x)}{\partial x_i \partial x_j}$ 可化为:

$$-\frac{1}{n^2}\left(\prod_{k=1}^{n} x_k\right)^{\frac{1-2n}{n}} \cdot \left[(1-n)\left(\prod_{k\neq i}^{n} x_k\right)\left(\prod_{k\neq i}^{n} x_k\right) + n\left(\prod_{k=1}^{n} x_k\right)\left(\prod_{\substack{k\neq i\\ k\neq j}}^{n} x_k\right)\right]$$

$$= -\frac{1}{n^2}\left(\prod_{k=1}^{n} x_k\right)^{\frac{1-2n}{n}} \left[\left(\prod_{k\neq i} x_k\right) \left(\prod_{k\neq j}^{n} x_k\right)\right].$$

Hessian矩阵为:

$$H = \frac{\left(\prod_{k=1}^{n} x_k\right)^{\frac{1-2n}{n}}}{n^2} \begin{bmatrix} (n-1)\left(\prod_{k\neq 1} x_k\right)^2 & (-1)\left(\prod_{k\neq 1} x_k\right)\left(\prod_{k\neq 2} x_k\right) & \cdots & \vdots \\ (-1)\left(\prod_{k\neq 1} x_k\right)\left(\prod_{k\neq 2} x_k\right) & (n-1)\left(\prod_{k\neq 2} x_k\right)^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ (-1)\left(\prod_{k\neq 1} x_k\right)\left(\prod_{k\neq n} x_k\right) & & & (n-1)\left(\prod_{k\neq n} x_k\right)^2 \end{bmatrix}$$

H的系数为正，令 T为H除系数部分外的矩阵，则有：

令 $z = [(\prod_{k\neq 1} x_k), (\prod_{k\neq 2} x_k), \cdots, (\prod_{k\neq n} x_k)]^T$.

$T = \mathrm{diag}(z)\begin{pmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & & \vdots \\ \vdots & & \ddots & -1 \\ -1 & \cdots & -1 & n-1 \end{pmatrix}\mathrm{diag}(z)$

对 $\forall$ 向量 $v \in R^n$，$v \neq 0$，有.

$v^T T v = v^T \mathrm{diag}(z) \begin{pmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & & \vdots \\ \vdots & & \ddots & -1 \\ -1 & \cdots & -1 & n-1 \end{pmatrix} \mathrm{diag}(z) v$

其中 $\begin{pmatrix} n-1 & -1 & \cdots & -1 \\ -1 & & & \vdots \\ \vdots & & \ddots & \\ -1 & \cdots & & n-1 \end{pmatrix} = \begin{pmatrix} n & & \\ & \ddots & \\ & & n \end{pmatrix} - \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \cdots & & 1 \end{pmatrix}$

令 $x = \mathrm{diag}(z) v$，则 $v^T T v = x^T \left[\begin{pmatrix} n & & \\ & \ddots & \\ & & n \end{pmatrix} - \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \right] x$

$= n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2$

由柯西不等式有：$n\sum_{i=1}^{n} x_i^2 \geq (\sum_{i=1}^{n} x_i)^2$. $v^T T v \geq 0$.

即证得 $f(x) = -(\sum_{k=1}^{n} x_k)^{\frac{1}{n}}$ (for $x \in R_+^n$) 为凸函数.

(2) $f(x,y) = \frac{x^2}{y}$. $\{(x,y) \mid y > 0\}$

$\frac{\partial f(x)}{\partial x} = \frac{2x}{y}$ $\quad \frac{\partial f(x)}{\partial y} = -\frac{x^2}{y^2}$ $\quad \frac{\partial f(x)}{\partial x^2} = \frac{2}{y}$ $\quad \frac{\partial f(x)}{\partial y^2} = 2\frac{x^2}{y^3}$

$\frac{\partial f(x)}{\partial x \partial y} = -\frac{2x}{y^2}$ $\quad \frac{\partial f(x)}{\partial y \partial x} = -\frac{2x}{y^2}$.

Hessian 矩阵为：

$H = \begin{bmatrix} \frac{2}{y} & -\frac{2x}{y^2} \\ -\frac{2x}{y} & \frac{2x^2}{y^3} \end{bmatrix}$ 对于 $\forall v \in R^2$，$v \neq 0$ 有

$v^T H v = \frac{2}{y} v_1^2 - \frac{4x}{y} v_1 v_2 + \frac{2x^2}{y^3} v_2^2$

$= \left(\sqrt{\frac{2}{y}} v_1 - \sqrt{\frac{2x^2}{y^3}} v_2\right)^2 \geq 0$

由凸函数的二阶定理有：$f(x,y)$ 为凸函数

在定义域 $\{(x,y) \mid y > 0\}$ 上是凸函数

6.4 对于 $\min f(x)=\frac{1}{N}\sum_{i=1}^{N}\log(1+e^{-y_i a_i^T x})$ $y_i \geq 0$，a_i已知。

估计 ∇f 李氏常数L.

$$\nabla f(x) = -\frac{1}{N}\sum_{i=1}^{N}\frac{y_i a_i}{1+e^{y_i a_i^T x}}$$

$$\nabla^2 f(x) = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i a_i a_i^T e^{y_i a_i^T x}}{(1+e^{y_i a_i^T x})^2} = (-1)\cdot\frac{1}{N}\sum_{i=1}^{N}\frac{y_i a_i}{1+e^{y_i a_i^T x}}\cdot\frac{a_i^T e^{y_i a_i^T x}}{1+e^{y_i a_i^T x}}$$

要求 ∇f 的L即求满足 $\|\nabla^2 f(x)-\nabla^2 f(x_1)\| \leq L\|\nabla f(x)-\nabla f(x_1)\|$ 的L

$$\frac{a_i^T e^{y_i a_i^T x}}{1+e^{y_i a_i^T x}} < a_i^T$$

$$\nabla f(x)-\nabla f(x_1) = -\frac{1}{N}\sum_{i=1}^{N}\frac{y_i a_i(e^{y_i a_i^T x_1}-e^{y_i a_i^T x})}{(1+e^{y_i a_i^T x})(1+e^{y_i a_i^T x_1})}$$

$$\nabla^2 f(x)-\nabla^2 f(x_1) = -\frac{1}{N}\sum_{i=1}^{N}\frac{y_i a_i a_i^T\cdot[(e^{y_i a_i^T x}-e^{y_i a_i^T x_1})(1+e^{y_i a_i^T(x_1+x)})]}{(1+e^{y_i a_i^T x})^2(1+e^{y_i a_i^T x_1})^2}$$

$$= -\frac{1}{N}\sum_{i=1}^{N}\frac{y_i a_i(e^{y_i a_i^T x_1}-e^{y_i a_i^T x})}{(1+e^{y_i a_i^T x})(1+e^{y_i a_i^T x_1})}\left[-\frac{a_i^T(1+e^{y_i a_i^T(x_1+x)})}{(1+e^{y_i a_i^T x})(1+e^{y_i a_i^T x_1})}\right]$$

其中 $\frac{a_i^T(1+e^{y_i a_i^T(x_1+x)})}{(1+e^{y_i a_i^T x_1})(1+e^{y_i a_i^T x})} < a_i^T$

即有 $\|\nabla^2 f(x)-\nabla^2 f(x_1)\| \leq \max_{i\in[1,N]}|a_i^T|\,\|\nabla f(x)-\nabla f(x_1)\|$

$\max|a_i^T|$ 为L的可取值.

6.5 $A \in R^{n\times n}$ $f(x) \in R^n$. $\phi(y) = f(Ay)$

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

$$y_{k+1} = y_k - \nabla^2 \phi(y_k)^{-1} \nabla \phi(y_k)$$

证明：若 $y_0 = A^{-1}x_0$，则对任意 $k \geq 1$，$y_k = A^{-1}x_k$.

$$\nabla \phi(y_k) = \frac{\partial \phi(y_k)}{\partial Ay} \cdot \frac{\partial Ay}{\partial y} = A^T \nabla f(Ay)$$

$$\nabla^2 \phi(y_k) = \nabla(A^T \nabla f(Ay_k)) = A^T \nabla^2 f(Ay_k) A.$$

$$y_{k+1} = y_k - \nabla^2 \phi(y_k)^{-1} \nabla \phi(y_k) = y_k - (A^T \nabla^2 f(Ay_k) A)^{-1} A^T \nabla f(Ay_k)$$

由于 $y_0 = A^{-1}x_0$ 故 $y_1 = A^{-1}x_0 - (A^T \nabla^2 f(AA^{-1}x_0) A)^{-1} A^T \nabla f(AA^{-1}x_0)$

$$= A^{-1}x_0 - (A^T \nabla^2 f(x_0) A)^{-1} A^T \nabla f(x_0)$$

$$x_1 = x_0 - \nabla^2 f(x_0)^{-1} \nabla f(x_0)$$

$$(A^T \nabla^2 f(x_0) A) A^{-1} x_1 = A^T \nabla^2 f(x_0) A A^{-1} x_1 = A^T \nabla^2 f(x_0) x_1$$

$$= A^T \nabla^2 f(x_0) x_0 - A^T \nabla^2 f(x_0) \nabla^2 f(x_0)^{-1} \nabla f(x_0)$$

$$= A^T \nabla^2 f(x_0) x_0 - A^T \nabla f(x_0)$$

$$(A^T \nabla^2 f(x_0) A) y_1 = A^T \nabla^2 f(x_0) x_0 - A^T \nabla f(x_0)$$

$\because$ $(A^T \nabla^2 f(x_0) A)(A^{-1}x_1) = (A^T \nabla^2 f(x_0) A) y_1$

$\therefore$ $y_1 = A^{-1}x_1$

下面用数学归纳法证明：若已知 $y_k = A^{-1}x_k$，则 $y_{k+1} = A^{-1}x_{k+1}$

$$y_{k+1} = y_k - (A^T \nabla^2 f(Ay_k) A)^{-1} A^T \nabla f(Ay_k)$$

$$x_{k+1} = x_k - \nabla^2 f(x_k)^{-1} \nabla f(x_k) = Ay_k - A(A^T \nabla^2 f(Ay_k) A)^{-1} A^T \nabla f(Ay_k)$$

$y_{k+1} = A^{-1}x_{k+1}$ 又由 $y_1 = A^{-1}x_1$，由数学归纳法证得结论成立

6.6 利用秩一校正的求逆公式，由 $H_{k+1}^{(DFP)}$ 推导 $B_{k+1}^{(DFP)}$

首先写出在DFP形式下的 $H_{k+1}^{(DFP)}$

$$H_{k+1}^{(DFP)} = H_k^{(DFP)} + \frac{s^k s^{kT}}{s^{kT} y^k} - \frac{H_k^{(DFP)} y^k y^{kT} H_k}{y^{kT} H_k y_k}$$

$$B_{k+1}^{(DFP)} = \left(H_{k+1}^{(DFP)}\right)^{-1}$$

令 $M = H + \frac{ss^T}{s^Ty}$ ①

则把①代入 sherman-Morrison 定理中有.

$$M^{-1} = \left(H + \frac{ss^T}{s^Ty}\right)^{-1} = H^{-1} + \frac{H^{-1}ss^TH^{-1}}{1+s^TH^{-1}u} = B + \frac{Bss^TB}{1+s^TBu}$$

由 $H_{k+1}^{(DFP)} = H_k^{(DFP)} + \frac{s^k s^{kT}}{s^{kT} y^k} - \frac{H_k^{(DFP)} y^k y^{kT} H_k}{y^{kT} H_k y_k}$

$$= M_k - \frac{H_k^{(DFP)} y^k y^k H_k}{y^{kT} H_k y^k} \quad ①$$

代入 sherman-Morrison 公式有:

$$(B_{k+1}^{(DFP)}) = (H_{k+1}^{DFP})^{-1} = M_k^{-1} + \frac{M_k^{-1} H_k^{(DFP)} y^k y^{kT} H_k M_k^{-1}}{y^T H y - y^T H M^{-1} H y} \quad ②$$

①代入②得

$$\frac{M^{-1}Hyy^THM^{-1}}{y^THy - y^THM^{-1}Hy} = \frac{yy^T(s^Ty+s^TBs)}{y^Tss^Ty} - \frac{ys^TB}{s^Ty} - \frac{Bsy^T}{y^Ts} + \frac{Bss^TB}{s^Ty+s^TBs}$$

$$= \left(1+\frac{s^TBs}{y^Ts}\right)\frac{yy^T}{y^Ts} - \frac{ys^TB+Bsy^T}{y^Ts} + \frac{Bss^TB}{s^Ty+s^TBs} \quad ③$$

①、③代入②得

$$B_{k+1}^{(DFP)} = B_k + \left(1+\frac{s^{(k)}B_k s^{(k)}}{y^{(k)T}s^{(k)}}\right)\frac{y^{(k)}y^{(k)T}}{y^{(k)T}s^{(k)}} - \frac{B_k s^{(k)} y^{(k)T} + y^{(k)} s^{(k)T} B_k}{y^{(k)T} s^{(k)}}$$

非线性优化.

习题1. ① 对于 $\forall d \in T(\bar{x}|S)$ 存在 $\bar{x}$ 的一个邻域 $(\bar{x}-\varepsilon, \bar{x}+\varepsilon)$ $\varepsilon>0$

使得对 $\forall x \in (\bar{x}-\varepsilon, \bar{x}+\varepsilon)$ 有 $f(x) \geq f(\bar{x})$

定义一个路径 $\gamma(t)$，其中 $\gamma'(0)=d$，$\gamma(0)=\bar{x}$

对于使 $\bar{x}-\varepsilon<\gamma(t)<\bar{x}+\varepsilon$ 的足够小的 t，

有 $f(\gamma(t)) \geq f(\bar{x})$

$$\frac{d}{dt} f(\gamma(t))\Big|_{t=0} = \nabla f(\bar{x})^T d$$

$\because$ $f(\gamma(t)) \geq f(\bar{x})$ 对于足够小的 t 成立

$\therefore$ 对于足够小的 t，有 $\frac{d}{dt} f(\gamma(t))\Big|_{t=0} \geq 0$

$\nabla f(\bar{x})^T d \geq 0$.

② $\max \ (-\nabla f(\bar{x}))^T d$

$$\begin{aligned} \text{s.t.} \quad & \nabla g_j(\bar{x})^T d \geq 0, \quad j \in I(\bar{x}) \\ & \nabla h_i(\bar{x})^T d = 0, \quad i=1,2,\cdots,l. \end{aligned}$$

$$L(\bar{x},\lambda,\mu) = -\nabla f(\bar{x})^T d + \sum_{j\in I(\bar{x})} \lambda_i (\nabla g_j(\bar{x})^T d) + \sum_{i=1}^{l} \mu_i (\nabla h_i(\bar{x})^T d)$$

对偶问题：

对偶函数 $g(\lambda,\mu) = \inf_{\bar{x}} L(\bar{x},\lambda,\mu)$

对偶问题 $\text{maximize} \ g(\lambda,\mu)$

$\text{s.t.} \quad \lambda_j \geq 0, \ j \in I(\bar{x})$

证明在MFCQ条件下，KKT条件成立

由强对偶定理可知：若原始问题和对偶问题都有解，并满足slater条件，则存在一对原始问题和对偶问题的最优解，使得对偶间隙为零，即 $s^T z = 0$。

假设满足MFCQ条件，根据KKT条件，互补性松弛条件为：

$$x_i s_i = 0, \quad i = 1, 2, \cdots, n.$$

若 $x_i > 0$，则必有 $s_i = 0$，反之亦然。这与互补松弛的定义一致。

故有互补松弛性条件成立。

在MFCQ条件下，平稳性条件确保了在最优点点梯度为0，对偶性条件确保了对偶问题的可行性，互补松弛性条件则表达了对原始问题和对偶问题变量的互补性。如果这些条件在最优点满足，则根据强对偶定理，原始问题和对偶问题的最优值相等。因此，KKT条件在MFCQ下确保了原始问题和对偶问题的最优性。

习题2. ∵ $f(w) = c^T w$、$g(w) = \sum_{i=1}^{n} w_i \log(w_i)$，$h(w) = \sum_{i=1}^{n} w_i - 1$

均为凸函数

∴ 这是一个凸优化问题

构造拉格朗日函数：

$$L(w, \lambda, \mu) = c^T w + \rho \sum_{i=1}^{n} (w_i \log(w_i) - \mu_i w_i) + \lambda \left(\sum_{i=1}^{n} w_i - 1 \right)$$

令 $\frac{\partial L}{\partial w_i} = c_i + \rho(1 + \log(w_i)) - \mu_i + \lambda = 0 \quad i = 1, 2, \cdots, n.$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^{n} w_i - 1 = 0.$$

$$w_i = e^{-1 - \frac{c_i - \lambda}{\rho}} \qquad \sum_{i=1}^{n} e^{-1 - \frac{c_i - \lambda}{\rho}} - 1 = 0.$$

最优解：$w_i = \dfrac{e^{-1-\rho c_i - \lambda}}{\sum_{j=1}^{n} e^{-1-\rho c_j - \lambda}}$