HW4. $\min f(x) = \frac{1}{2} x^T A x - b^T x$

1. 线性共扼梯度法:

证明: $\text{span}(r_0, r_1, \cdots, r_k) = \text{span}(r_0, Ar_0, \cdots, A^k r_0)$ (1)

$\text{span}(p_0, p_1, \cdots, p_k) = \text{span}(r_0, Ar_0, \cdots, A^k r_0)$ (2)

$r_k^T p_i = 0, \quad \forall i<k$ (3)

$p_k^T A p_i = 0, \quad \forall i<k$ (4)

$r_k^T r_i = 0, \quad \forall i<k.$ (5)

已知: $\alpha_k = \frac{r_k^T r_k}{p_k^T A p_k}$ (6)

$r_{k+1} = r_k + \alpha_k A p_k$ (7)

$\beta_{k+1} = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k}$ (8)

$p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$ (9)

$r_{k+1}^T p_k = 0.$

证明:(3) $r_k^T p_k = r_k^T [-r_k + \beta_k p_{k-1}] = -r_k^T r_k + \beta_k p_{k-1}^T r_k$

$= -r_k^T r_k$

$r_k^T p_{k-1} = r_k^T \frac{p_k + r_k}{\beta_k} = \frac{r_k^T [p_k + r_k]}{\beta_k} = 0.$

由数学归纳法,若已知 $r_k^T p_j = 0$, 则由 $r_k^T p_{j-1} = r_k^T [\frac{p_j + r_j}{\beta_j}]$ 知,

要证明 $r_k^T p_j = 0$. 即可先证明 $r_k^T r_j = 0$.

下面来先证明(4)、(5)式: 当 $k=1$ 时, 显然有 $r_k^T r_i = 0$, $p_k^T A p_i = 0$, $\forall i<k$

假设(4)、(5)式同时对 $k$ 成立., 对(7)式, 等式两侧转置后同乘 $r_i$

$r_{k+1}^T r_i = r_k^T r_i + \alpha_k p_k^T A r_i = r_k^T r_i - \alpha_k p_k^T A [r_i - \beta_i p_{i-1}]$

由(4)、(5)对 $k$ 成立, 有 $r_k^T r_i = 0$, $\alpha_k p_k^T A r_i = 0$.

又 $\because \alpha_k p_k^T A \beta_i p_{i-1} = 0$

$\therefore r_{k+1}^T r_i = 0.$ 证得(5)式对 $(k+1)$ 成立, 由归纳假设得(5)式成立.

由上面的假设，$p_k^T A p_i = 0$ 已知成立，下面证明 $p_{k+1}^T A p_i = 0$.

$$p_{k+1}^T A p_i = (-r_{k+1} + \beta_{k+1} p_k)^T A \quad p_i$$
$$= -r_{k+1}^T A p_i + \beta_{k+1} p_k A p_i$$
$$= r_{k+1}^T \left( \frac{r_{i+1} - r_i}{a_i} \right) + \beta_k p_k A p_i$$

由 $p_k A p_i = 0$ 以及 $r_{k+1}^T r_i = 0 \ \forall i<k$ 都成立.

可知 $p_{k+1}^T A p_i = 0$.

回到对(3)式即 $r_k^T p_i = 0$ 的证明.

由于已证明 $r_k^T r_j = 0$. 即有 $r_k^T p_{j-1} = r_k^T \left[ \frac{p_i + r_j}{\beta_j} \right]$

$$= 0 + 0 = 0.$$

故由数学归纳法可证得 $r_k^T p_i = 0 \quad \forall i<k$.

下面证明：$span(r_0, r_1, \cdots r_k) = span(r_0, A r_0, \cdots, A^k r_0)$

$span(p_0, p_1, \cdots p_k) = span(r_0, A r_0, \cdots, A^k r_0)$

由于 $p_{k+1} = -r_{k+1} + \beta_{k+1} p_k$

∴ 存在可逆方阵 $Q$，使得 $(r_0, \cdots r_k) = Q(p_0, \cdots p_k)$

其中 $Q = \begin{bmatrix} -1 & & & & \\ 0 & -1 & \beta_1 & & \\ & & \ddots & & \\ & & & & \beta_{k-1} \\ & & & & -1 \end{bmatrix}$

故 $span(r_0, r_1, \cdots, r_k) = span(p_0, \cdots, p_k)$

对于 $k=0$，显然 $span(r_0, r_1, \cdots, r_k) = span(r_0, Ar_0, \cdots, A^k r_0)$

假设此式对 K 成立，则对 $(k+1)$

由于 $r_{k+1} = r_k + \alpha_k A p_k$，可知 $r_k \in span\{r_0, \cdots, A^k r_0\}$

且 $\alpha_k A p_k \in span\{Ar_0, \cdots, A^{k+1} r_0\}$

即有 $r_{k+1} \in span\{r_0, Ar_0, \cdots, A^{k+1} r_0\}$

$\because$ $p_0, p_1, \cdots, p_k$ 为一组共轭方向，

$\therefore$ $r_{k+1} \perp span\{p_0, \cdots, p_k\}$

即 $r_{k+1} \in span\{p_0, \cdots, p_k\} = span\{r_0, \cdots, A^k r_0\}$ 所以 $r_{k+1} \in span\{A^{k+1} r_0\}$

综上 $span(r_0, r_1, \cdots, r_k) = span(r_0, Ar_0, \cdots, A^{k+1} r_0)$

(1)式得证

又 $\because$ $span(r_0, \cdots, r_k) = span(p_0, p_1, \cdots, p_k)$

$\therefore$ $span(p_0, \cdots, p_k) = span(r_0, \cdots, r_k)$

$= span(r_0, Ar_0, \cdots, A^k r_0)$

(2)式得证

2. $\phi(\lambda)=\frac{1}{\Delta}-\frac{1}{\|d(\lambda)\|}=0.$

$\lambda_{k+1}=\lambda_k-\frac{\phi(\lambda_k)}{\phi'(\lambda_k)}$ $d(\lambda)=-\sum_{i=1}^{n}\frac{q_i^Tg}{\lambda_i+\lambda}q_i$ $\|d(\lambda)\|=\sum_{i=1}^{n}\left|\frac{q_i^Tg}{\lambda_i+\lambda}\right|$

$$\frac{d}{d\lambda}\left(\frac{1}{\|d(\lambda)\|}\right)=-\frac{1}{\|d(\lambda)\|^2}\frac{d}{d\lambda}\|d(\lambda)\|$$

$$=-\frac{1}{\|d(\lambda)\|^3}\sum_{i=1}^{n}\frac{d_i(\lambda)}{|\lambda_i+\lambda|}$$

$$\phi'(\lambda_k)=\frac{d}{d\lambda}\left(\frac{1}{\|d(\lambda)\|}\right)=-\frac{1}{\|d(\lambda_k)\|^3}\sum_{i=1}^{n}\frac{d_i(\lambda_k)}{|\lambda_i+\lambda_k|}$$

$$\lambda_{k+1}=\lambda_k-\frac{\frac{1}{\Delta}-\frac{1}{\|d(\lambda_k)\|}}{-\frac{1}{\|d(\lambda_k)\|^3}\sum_{i=1}^{n}\frac{d_i(\lambda_k)}{|\lambda_i+\lambda_k|}}$$

3. $x_k$是$P_E(x^k,\sigma^k)$最小值点.

结论1: $\sigma_{k+1}>\sigma_k>0$, 则有$P_E(x^k,\sigma^k)\le P_E(x^{k+1},\sigma^{k+1})$

$\sum_{i\in\varepsilon}\|c_i(x^k)\|^2\ge\sum_{i\in\varepsilon}\|c_i(x^{k+1})\|^2$, $f(x^k)\le f(x^{k+1})$

结论2: 对$\forall\sigma^k>0$, 有$f(\bar{x})\ge P_E(x^k,\sigma^k)\ge f(x^k)$成立, $\bar{x}$为最优解.

结论3: $\delta=\sum_{i\in\varepsilon}\|c_i(x^k)\|^2$, 则$x^k$为 $\min f(x)$

$\text{s.t.}\ \sum_{i\in\varepsilon}\|c_i(x)\|^2<\delta$的最优解.

$P_E(x,\sigma)=f(x)+\frac{1}{2}\sigma\sum_{i\in\varepsilon}c_i^2(x)$

结论1:

$x^k$为$P_E(x^k,\sigma^k)$最小值点，同理有$x^{k+1}$为$P_E(x^{k+1},\sigma^{k+1})$最小值点.

由$x^k$、$x^{k+1}$定义有.

$$f(x^k)+\frac{1}{2}\sigma_k\sum_{i\in\varepsilon}c_i^2(x^k)\le f(x^{k+1})+\frac{1}{2}\sigma_k\sum_{i\in\varepsilon}c_i^2(x^{k+1}) \quad ①$$

$$f(x^{k+1})+\frac{1}{2}\sigma_{k+1}\sum_{i\in\varepsilon}c_i^2(x^{k+1})\le f(x^k)+\frac{1}{2}\sigma_{k+1}\sum_{i\in\varepsilon}c_i^2(x^k) \quad ②$$

①、②式相加，得：$(\sigma_k-\sigma_{k+1})\sum_{i\in\varepsilon}\|c_i(x^k)\|^2\le(\sigma_k-\sigma_{k+1})\sum_{i\in\varepsilon}\|c_i(x^{k+1})\|^2$

又∵ $\sigma_{k+1}>\sigma_k>0$

∴ $\sum_{i\in\varepsilon}\|c_i(x^k)\|^2\ge\sum_{i\in\varepsilon}\|c_i(x^{k+1})\|^2$.

代入 $f(x^k)+\frac{1}{2}\sigma_k\sum_{i\in\varepsilon}c_i^2(x^k)\le f(x^{k+1})+\frac{1}{2}\sigma_k\sum_{i\in\varepsilon}c_i^2(x^{k+1})$

可知 $f(x^k)\le f(x^{k+1})$

$$\begin{aligned}P_E(x^k,\sigma^k)&=f(x^k)+\frac{1}{2}\sigma_k\sum_{i\in\varepsilon}\|c_i(x^k)\|^2\le f(x^{k+1})+\frac{1}{2}\sigma_k\sum_{i\in\varepsilon}c_i^2(x^{k+1})\\&\le f(x^{k+1})+\frac{1}{2}\sigma_{k+1}\sum_{i\in\varepsilon}\|c_i(x^{k+1})\|^2\\&=P_E(x^{k+1},\sigma^{k+1})\end{aligned}$$

结论2: $\bar{x}$为 $\min_x f(x)$ s.t. $c_i(x)=0$, $i\in\varepsilon$ 的最优解.

由于$x_k$为$\min P_E(x_k,\sigma_k)$最优解，对$\forall x$有 $f(x)+\frac{1}{2}\sigma_k\|c_i(x)\|^2\ge f(x^k)+\frac{1}{2}\sigma_k\|c_i(x)\|^2$

当 $x=\bar{x}$ 时 $f(\bar{x})+\frac{1}{2}\sigma_k\sum_{i\in\varepsilon}\|c_i(x)\|^2 \geq P_E(x^k,\sigma^k)$

又 $\because c_i(\bar{x})=0 \quad i\in\varepsilon$

$\therefore f(\bar{x}) \geq P_E(x^k,\sigma^k) \geq f(x^k)$

得证.

结论3: $\delta=\sum_{i\in\varepsilon}\|c_i(x^k)\|^2$.

对 $\forall$ $x$ 满足 $\sum_{i\in\varepsilon}\|c_i(x)\|^2 \leq \sum_{i\in\varepsilon}\|c_i(x^k)\|^2$.

由于 $x^k$ 是 $P_E(x^k,\sigma^k)$ 最小值点有: $f(x)+\frac{1}{2}\sigma_k\sum_{i\in\varepsilon}\|c_i(x)\|^2 \geq f(x^k)+\frac{1}{2}\sigma_k\sum_{i\in\varepsilon}\|c_i(x^k)\|^2$.

又 $\because \frac{1}{2}\sigma_k\sum_{i\in\varepsilon}\|c_i(x)\|^2 \leq \frac{1}{2}\sigma_k\sum_{i\in\varepsilon}\|c_i(x^k)\|^2$.

$\therefore f(x) \geq f(x^k)$ 对任意满足 $\sum_{i\in\varepsilon}\|c_i(x)\|^2 \leq \sum_{i\in\varepsilon}\|c_i(x^k)\|^2$ 的 $x$ 恒成立

即有 $x^k$ 为此问题最优解.

4. 标准线性规划问题: $\min\ c^T x$.

$$\text{s.t.}\ A^T x = b,\ x \geq 0$$

对偶问题 $\max\ b^T y$

$$\text{s.t.}\ A^T y + s = c.\quad s \geq 0.$$

引入λ和σ.

$$L_\sigma(y, s, \lambda) = b^T y + \lambda^T(A^T y + s - c) + \frac{\sigma}{2}\|A^T y - s - c\|_2^2.$$

迭代格式为:

$$\begin{cases} (y^{k+1}, s^{k+1}) = \underset{y,\, s>0}{\arg\min} \left\{ b^T y + \frac{\sigma_k}{2}\left\| A^T y + s - c + \frac{\lambda}{\sigma_k}\right\|_2^2 \right\} \\ \lambda^{k+1} = \lambda^k + \sigma_k (A^T y^{k+1} - s^{k+1} - c) \\ \sigma_{k+1} = \min\{\rho\sigma_k, \bar{\sigma}\} \end{cases}$$

对于 $x_k \to x_{k+1}$. $L_\sigma(x, \lambda) = c^T x + \lambda^T(Ax - b) + \frac{\sigma}{2}\|Ax - b\|_2^2$

$$x_{k+1} = \underset{x \in R^n}{\arg\min} \left\{ c^T x_k + \frac{w}{2}\left\| Ax - b - \frac{\lambda^k}{\sigma}\right\|_2^2 \right\}.$$

其中σ为罚因子, λ为乘子.

5. (1) $f(x) = \|Ax-b\|_2$

(2) $f(x) = \min_y \|Ay-x\|_\infty$ $\qquad \|x\|_\infty = \max_i |x_i|$

存在$\hat{y}$使得 $f(\hat{x}) = \min_y \|Ay-x\|_\infty$.

计算 $f(\hat{x})$ 次梯度.

(1) 由于 $\|Ax-b\|_2$ 为凸函数.

若 $Ax-b\neq 0$，则$f(x)$可微.

故 $\partial f(x) = 2A^TAx - 2A^Tb$

令 $h(x) = \|x\|_2$, $\quad f(x) = h(Ax-b)$

由链式法则 $\quad \partial f(x) = A^T \partial h(Ax-b)$

又∵ $\begin{cases} \partial h(x) = \frac{1}{\|x\|_2} x & x\neq 0 \\ \partial h(0) \subset \{g \mid \|g\|_2 \leq 1\} \end{cases}$

∴ $\partial f(0) \subset A^Tg$ $\quad$ 其中 $\|g\|_2 \leq 1$

(2) 令 $h(x,y) = \|Ay-x\|_\infty$

∵ $f(\hat{x}) = \min_y \|Ay-x\|_\infty$.

∴ 对∀ $x\in R^n$, $y\in R^n$

$$h(x,y) \geq h(\hat{x},\hat{y}) + g^T(x-\hat{x}) + 0^T(y-\hat{y})$$
$$= f(\hat{x}) + g^T(x-\hat{x})$$
$$f(x) = \min_y h(x,y) \geq f(\hat{x}) + g^T(x-\hat{x})$$

$\hat{x}=0$ 时，$f(\hat{x})$ 的一个次梯度为 $\vec{0}$.

$\hat{x}\neq 0$ 时，令 $h(t)=\|t\|_\infty=\max\limits_{s} s^T t$

$s=(0,0,\cdots,1,\cdots 0,\cdots,0)^T$.

$$\partial h(t)=\begin{cases}\vec{s}, & \vec{t}\neq\vec{0}\\ \vec{0}, & \vec{t}=\vec{0}\end{cases}$$

$\therefore \partial f(\hat{x})=A^T(\partial h(t)|_{t=Ay-x})$.