1. (1) 用数学归纳法. 假设 $n=k$ 时成立. 证对 $n=k+1$ 也成立.

$r_k \in span\{r_0, \dots, A^k r_0\}$ $p_k \in span\{r_0, \dots, A^k r_0\}$.

$\therefore r_{k+1} = r_k + \alpha_k A p_k \in span\{r_0, Ar_0, \dots, A^{k+1} r_0\}$

$\therefore span\{r_0, \dots, r_{k+1}\} \subseteq span\{r_0, Ar_0, \dots, A^{k+1} r_0\}$.

而 $A^{k+1} r_0 = A \cdot A^k r_0 \in span\{Ap_0, \dots, Ap_k\}$. $Ap_i = (r_{i+1} - r_i)/\alpha_i$

$\therefore A^{k+1} r_0 \in span\{r_0, \dots, r_{k+1}\}$

$\therefore span\{r_0, \dots, r_{k+1}\} \supseteq span\{r_0, Ar_0, \dots, A^{k+1} r_0\}$.

(2) $span\{p_0, \dots, p_k, p_{k+1}\} = span\{p_0, \dots, p_k, r_{k+1}\}$

$= span\{r_0, \dots, A^k r_0, r_{k+1}\}$

$= span\{r_0, \dots, A^k r_0, A^{k+1} r_0\}$ (由(1)可得)

假设 $n=k$ 时, (3) (4) (5) 都成立.

当 $n=k+1$ 时

$\forall i<k$. $r_{k+1}^T p_i = r_k^T p_i + \alpha_k p_k^T A^T p_i = 0$

$i=k$. $r_{k+1}^T p_k = r_k^T p_k + \alpha_k p_k^T A^T p_k = 0$ (由α_k定义) $\Rightarrow$ (3) 成立

$\because r_{k+1} \perp span\{p_0, \dots, p_k\} = span\{r_0, \dots, r_k\}$

$\therefore r_{k+1}^T r_i = 0$, $\forall i<k+1$ $\Rightarrow$ (5) 成立

$\forall i<k$. $p_{k+1}^T A p_i = -r_{k+1}^T A p_i + \beta_{k+1} p_k^T A p_i$

$= -r_{k+1}^T A p_i$

$= -\frac{1}{\alpha_i} r_{k+1}^T (r_{i+1} - r_i) = 0$

$i=k$. $p_{k+1}^T A p_k = 0$ (由β_k定义) $\Rightarrow$ (4) 成立

2. $\phi'(\lambda) = \|d(\lambda)\|^{-3} \cdot d(\lambda)^T \cdot d'(\lambda)$

$\because (B+\lambda I)\, d(\lambda) = -g$

两边对 λ 求导, $(B+\lambda I)\cdot d'(\lambda) + d(\lambda) = 0$

$\therefore d'(\lambda) = -(B+\lambda I)^{-1}\cdot d(\lambda)$

$\therefore \phi'(\lambda) = -\dfrac{1}{\|d(\lambda)\|^3}\cdot d(\lambda)^T (B+\lambda I)^{-1} d(\lambda)$

令 $B+\lambda I = R^T R$, 则 $d(\lambda)^T (B+\lambda I)^{-1} d(\lambda)$

$= d(\lambda)^T R^{-1} R^T d(\lambda)$

$= -\|R^{-T} d(\lambda)\|^2$

$$\therefore \lambda^{(k+1)} = \lambda^{(k)} - \frac{\phi(\lambda^{(k)})}{\phi'(\lambda^{(k)})}$$

$$= \lambda^{(k)} + \frac{\dfrac{\|d(\lambda^{(k)})\| - \Delta}{\Delta \|d(\lambda^{(k)})\|}}{\dfrac{1}{\|d(\lambda^{(k)})\|^3}\|R^{-T} d(\lambda^{(k)})\|^2}$$

$$= \lambda^{(k)} + \frac{\|d(\lambda^{(k)})\|^2 (\|d(\lambda^{(k)})\| - \Delta)}{\|R^{-T} d(\lambda^{(k)})\|^2 \cdot \Delta}$$

3. (1) 若 $P_E(x^k,\sigma^k) > P_E(x^{k+1},\sigma^{k+1})$

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

$\leq f(x^{k+1}) + \frac{1}{2}\sigma^{k+1} \cdot \sum_{i\in\varepsilon} \|c_i(x^{k+1})\|^2$

$= P_E(x^{k+1},\sigma^{k+1}) < P_E(x^k,\sigma^k)$

与 x^k 是 $P_E(x,\sigma^k)$ 的最小值点矛盾 $\therefore P_E(x^k,\sigma^k) \leq P_E(x^{k+1},\sigma^{k+1})$.

若 $\sum_{i\in\varepsilon}\|c_i(x^k)\|^2 \leq \sum_{i\in\varepsilon}\|c_i(x^{k+1})\|^2$

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

$f(x^k) - f(x^{k+1}) + \frac{1}{2}\sigma^{k+1} \cdot \sum_{i\in\varepsilon}\left(\|c_i(x^k)\|^2 - \|c_i(x^k)\|^2\right) \geq 0$

$\therefore f(x^{k+1}) + \frac{1}{2}\sigma^k \cdot \sum_{i\in\varepsilon}\|c_i(x^{k+1})\|^2$

$\leq f(x^k) + \frac{1}{2}\sigma^{k+1} \cdot \sum_{i\in\varepsilon}\left(\|c_i(x^k)\|^2 - \|c_i(x^{k+1})\|^2\right) + \frac{1}{2}\sigma^k \cdot \sum_{i\in\varepsilon}\|c_i(x^{k+1})\|^2$

$= f(x^k) + \frac{1}{2}\sigma^k \cdot \sum\|c_i(x^k)\|^2 + \frac{1}{2}(\sigma^{k+1}-\sigma^k) \cdot \sum_{i\in\varepsilon}\left(\|c_i(x^k)\|^2 - \|c_i(x^{k+1})\|^2\right)$

$\leq f(x^k) + \frac{1}{2}\sigma^k \cdot \sum\|c_i(x^k)\|^2$

与 x^k 是 $P_E(x,\sigma^k)$ 的最小值点矛盾 $\therefore \sum\|c_i(x^k)\|^2 \geq \sum\|c_i(x^{k+1})\|^2$

若 $f(x^k) > f(x^{k+1})$

则 $f(x^{k+1}) + \sigma^k \sum\|c_i(x^{k+1})\|^2 < f(x^k) + \sigma^k \sum\|c_i(x^k)\|^2$

与 x^k 是 $P_E(x,\sigma^k)$ 的最小值点矛盾 $\therefore f(x^k) \leq f(x^{k+1})$

(2) $\because \sigma^k > 0 \quad \|c_i(x^k)\|^2 \geq 0 \quad \therefore P_E(x^k,\sigma^k) \geq f(x^k)$

若 $\bar{x}$ 为原问题最优解，则 $c_i(\bar{x}) = 0$，$\forall i$ $\therefore P_E(\bar{x},\sigma^k) = f(\bar{x})$

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

(3) 若 x^k 不是最优解. 则 $\exists \bar{x}$, s.t. $f(\bar{x}) < f(x^k)$ 且 $\sum|c_i(\bar{x})\|^2 \leq \delta$

则 $P_E(\bar{x},\sigma_k) < f(x^k) + \sigma_k\delta = P_E(x^k,\sigma^k)$

与 x^k 是 $P_E(x,\sigma^k)$ 的最小值点矛盾 $\therefore x^k$ 为最优解.

4. 标准LP： $\min \ c^T x$

$\text{s.t. } Ax = b, \ x \ge 0$

对偶问题： $\max \ b^T \lambda \quad \text{s.t. } A^T\lambda \le c \quad \Rightarrow \quad \min \ -b^T y \quad \text{s.t. } A^T y - c + s = 0, \ s \ge 0$

增广Lagrange函数

$$L_\sigma(y, s, \lambda) = -b^T y + \lambda^T (A^T y - c + s) + \frac{\sigma}{2} \| A^T y - c + s \|_2^2, \quad s \ge 0$$

迭代公式

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

问题 $\min\limits_{s} \frac{\sigma}{2} \| A^T y + s - c + \frac{\lambda}{\sigma} \|_2^2 \quad \text{s.t. } s \ge 0$ 的解为 $s = P_{s \ge 0}(c - A^T y - \frac{\lambda}{\sigma})$

$$= \max\{c - A^T y - \frac{\lambda}{\sigma}, 0\}.$$

迭代公式更新为

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

5. (1) 令 $h(x)=\|x\|_2$. $h(x)$ 只在 $x=0$ 处不可微.

且 $\partial h(0)=\{g \mid \|g\|_2\le 1\}$.

从而 $\partial f(x)=A^T\partial\|Ax-b\|_2=\begin{cases}\dfrac{A^T(Ax-b)}{\|Ax-b\|_2}, & Ax\ne b\\ \{A^Tg \mid \|g\|_2\le 1\}, & Ax=b\end{cases}$

(2) 令 $h(x,y)=\|Ay-x\|_\infty=\max\limits_{1\le j\le m}|(Ay-x)_j|$

设 $A=\begin{bmatrix}a_1^T\\ \vdots\\ a_m^T\end{bmatrix}$，则 $h(x,y)=\max\limits_{1\le j\le m}|a_j^Ty-e_j^Tx|$

记 $g_j(x,y)=|e_j^Tx-a_j^Ty|=\left|(e_j^T\ \ -a_j^T)\begin{pmatrix}x\\y\end{pmatrix}\right|$

$\because (e_j^T\ \ -a_j^T)\begin{pmatrix}x\\y\end{pmatrix}$ 为线性 $\therefore g_j(x,y)$ 为凸函数.

$\therefore h(x,y)=\max\limits_{1\le j\le m}g_j(x,y)$ 关于 (x,y) 联合凸.

给定 $\hat{x}$, $\because \hat{y}$ 满足 $f(\hat{x})=\inf\limits_y\|Ay-\hat{x}\|_\infty=\|A\hat{y}-\hat{x}\|_\infty$

由最优条件 $0\in\left.\dfrac{\partial h(\hat{x},\hat{y})}{\partial y}\right|_{y=\hat{y}}$

不妨设 i 使得 $h(\hat{x},\hat{y})=\left|(e_i^T\ \ -a_i^T)\begin{pmatrix}\hat{x}\\\hat{y}\end{pmatrix}\right|=\max\limits_{1\le j\le m}|e_j^T\hat{x}-a_j^T\hat{y}|$

则 $\mathrm{sign}((\hat{x}-A\hat{y})_i)e_i\in\left.\dfrac{\partial h(x,\hat{y})}{\partial x}\right|_{x=\hat{x}}$

$\therefore (\mathrm{sign}((\hat{x}-A\hat{y})_i)e_i,\ 0)$ 是 $h(\hat{x},\hat{y})$ 的一个次梯度.

$\therefore \mathrm{sign}((\hat{x}-A\hat{y})_i)e_i$ 是 $f(\hat{x})$ 的一个次梯度.