P1 $\min_{y\in\mathbb{R}^n} \frac{1}{2}\|x-y\|_2^2$ subject to $Ay \preceq b$

(a) Lagrangian $\mathcal{L}(y,\lambda) = \frac{1}{2}\|x-y\|_2^2 + \lambda^T(Ay-b)$

Dual function $g(\lambda) = \inf_{y\in\mathbb{R}^n} \mathcal{L}(y,\lambda)$

$= \inf_{y\in\mathbb{R}^n}\left(\frac{1}{2}\|x-y\|_2^2 + \lambda^T(Ay-b)\right)$ → convex over y.

$= \frac{1}{2}\|x-(x-A^T\lambda)\|_2^2 + \lambda^T(A(x-A^T\lambda)-b)$

$= \frac{1}{2}\lambda^TAA^T\lambda + \lambda^TAx - \lambda^TAA^T\lambda - \lambda^Tb$

$= -\frac{1}{2}\lambda^TAA^T\lambda + (Ax-b)^T\lambda$.

(b) As the primal problem is feasible by assumption, the strong duality holds. And, the problem is convex, thus KKT condition is necessary & sufficient condition for optimality. Let y^*, λ^* be the optimal points for primal, dual problem, resp.

KKT conditions:

$$\begin{cases} Ay^*-b \preceq 0 & \text{(PF)} \\ \lambda^* \succeq 0 & \text{(DF)} \\ (\lambda^*)^T(Ay^*-b) = 0 & \text{(CS)} \\ \nabla_y \mathcal{L}(y,\lambda^*)\big|_{y=y^*} = 0 & \text{(LS)} \end{cases}$$

(LS) → $-x+y^*+A^T\lambda^* = 0$ ⇒ $y^* = x - A^T\lambda^* = x - A^T(AA^T)^{\dagger}(Ax-b)$.

$d^* = \max_{\lambda\succeq 0} g(\lambda) = \min_{\lambda\succeq 0} -g(\lambda)$

$-\nabla_\lambda g(\lambda) = AA^T\lambda + b - Ax = 0$ ⇒ $\lambda^* = (AA^T)^{\dagger}(Ax-b)$.

P2-(c).

$-g(\lambda) = \frac{1}{2}\lambda^TAA^T\lambda + (b-Ax)^T\lambda$

$-\nabla_\lambda g(\lambda) = AA^T\lambda + b - Ax$

Let $\lambda_1, \lambda_2 \succeq 0$. $\|(-\nabla_\lambda g(\lambda_1)) - (-\nabla_\lambda g(\lambda_2))\|_2$

$= \|(AA^T\lambda_1 + b - Ax) - (AA^T\lambda_2 + b - Ax)\|_2$

$= \|AA^T(\lambda_1-\lambda_2)\|_2 \leq \|AA^T\|_F\|\lambda_1-\lambda_2\|_2$

Hence $-g(\lambda)$ has a Lipschitz continuous gradient with the Lipschitz constant $L = \|AA^T\|_F$. Thus I used $s_k = 1/L = \|AA^T\|_F^{-1}$ for the stepsize.

P3-(c)

$f_0(x) = \frac{1}{2}x^THx + c^Tx$. $\nabla f_0(x) = Hx + c$

$\|\nabla f_0(x) - \nabla f_0(y)\| = \|H(x-y)\|_2 \leq \|H\|_F\|x-y\|_2 \quad \forall x, y$.

Hence f_0 has a Lipschitz continuous gradient with the Lipschitz constant $L = \|H\|_F$. Thus I used $s_k = 1/L = \|H\|_F^{-1}$ for the step size.