Question 1

(4 points) Find a global optimal solution to the following problem:

$$\begin{array}{ll} \min_{x\in\mathbb{R}^3} & x^T Hx + c^T x \\ \text{s.t.} & A^T x \le b \\ & x \ge 0 \end{array}$$

where $c = [1\ 2\ 1]^T$, $A = [-2 - 1 - 1]^T$, $b = -2$ and

$$H = \begin{bmatrix} 4 & -1 & 0 \\ -1 & 2 & 2 \\ 0 & 2 & 3 \end{bmatrix}.$$

Justify why your solution is a global minimizer. You can also use software (for example, matlab) for solving linear systems.

(Hint: You don't have to compute all the KKT points, let $\lambda_0, \lambda_1, \lambda_2, \lambda_3$ be the corresponding dual variables for the 4 constraints, you can try $\lambda_1 = \lambda_2 = 0$.

Solution:

Let $x = [x_1, x_2, x_3]^T$.

First, we define Lagrangian function of the problem:

$$L(x, \lambda_0, \lambda_1, \lambda_2, \lambda_3) = x^T Hx + c^T x + \lambda_0(A^T x - b) - \lambda_1 x_1 - \lambda_2 x_2 - \lambda_3 x_3$$

Since $A^T x - b \le 0$ and $-x \le 0$ are inequalities, we know $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \ge 0$. Then, since H is symmetric by observation, we try to get the stationary condition:

$$\nabla_x L(x, \lambda_0, \lambda_1, \lambda_2, \lambda_3) = 2Hx + c + \lambda_0 A - \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} = 0$$

$$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \lambda_0 + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} - 2Hx = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

Overall, a part of KKT conditions for the problem are:

$$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \lambda_0 + \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} - 2Hx = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{bmatrix} \ge \mathbf{0}$$

$$\lambda_0 \ge 0$$

$$\lambda_0([-2 \quad -1 \quad -1]x + 2) = 0$$

$$\lambda_i x_i = 0, i = 1, 2, 3$$

First, we try when $\lambda_1 = \lambda_2 = 0$ and other multipliers are not 0, then

$$\begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix} \lambda_0 + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \lambda_3 - 2Hx = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

$$\lambda_0, \lambda_3 > 0$$

$$2x_1 + x_2 - 2 = 0$$

$$x_3 = 0$$

Then, we solve

$$\begin{bmatrix}2\\1\\1\end{bmatrix}\lambda_0+\begin{bmatrix}0\\0\\1\end{bmatrix}\lambda_3-2Hx=\begin{bmatrix}1\\2\\1\end{bmatrix}$$

$$\begin{bmatrix}2\\1\\1\end{bmatrix}\lambda_0+\begin{bmatrix}0\\0\\1\end{bmatrix}\lambda_3-\begin{bmatrix}8&-2&0\\-2&4&4\\0&4&6\end{bmatrix}\begin{bmatrix}x_1\\2-2x_1\\0\end{bmatrix}=\begin{bmatrix}1\\2\\1\end{bmatrix}(*)$$

Since

$$\begin{bmatrix}8&-2&0\\-2&4&4\\0&4&6\end{bmatrix}\begin{bmatrix}x_1\\2-2x_1\\0\end{bmatrix}=\begin{bmatrix}12x_1-4\\-10x_1+8\\-8x_1+8\end{bmatrix}=\begin{bmatrix}12\\-10\\-8\end{bmatrix}x_1+\begin{bmatrix}-4\\8\\8\end{bmatrix}$$

We substitute it back to (*) to get

$$\begin{bmatrix}2\\1\\1\end{bmatrix}\lambda_0+\begin{bmatrix}0\\0\\1\end{bmatrix}\lambda_3-\begin{bmatrix}12\\-10\\-8\end{bmatrix}x_1-\begin{bmatrix}-4\\8\\8\end{bmatrix}=\begin{bmatrix}1\\2\\1\end{bmatrix}$$

$$\begin{bmatrix}2\\1\\1\end{bmatrix}\lambda_0+\begin{bmatrix}0\\0\\1\end{bmatrix}\lambda_3+\begin{bmatrix}-12\\10\\8\end{bmatrix}x_1=\begin{bmatrix}-3\\10\\9\end{bmatrix}$$

$$\begin{bmatrix}2&0&-12\\1&0&10\\1&1&8\end{bmatrix}\begin{bmatrix}\lambda_0\\\lambda_3\\x_1\end{bmatrix}=\begin{bmatrix}-3\\10\\9\end{bmatrix}$$

$$\begin{bmatrix}\lambda_0\\\lambda_3\\x_1\end{bmatrix}=\left(\begin{bmatrix}2&0&-12\\1&0&10\\1&1&8\end{bmatrix}\right)^{-1}\begin{bmatrix}-3\\10\\9\end{bmatrix}=\begin{bmatrix}\frac{45}{16}\\\frac{7}{16}\\\frac{23}{32}\end{bmatrix}>0$$

Then, $x_2=2-2x_1=2-\frac{23}{16}=\frac{9}{16}$. So far, we obtain a KKT point

$(x_1,x_2,x_3,\lambda_0,\lambda_1,\lambda_2,\lambda_3)=\left(\frac{23}{32},\frac{9}{16},0,\frac{45}{16},0,0,\frac{7}{16}\right)$. We again check it:

$$\begin{bmatrix}2\\1\\1\end{bmatrix}\lambda_0+\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\end{bmatrix}-2Hx=\frac{45}{16}\begin{bmatrix}2\\1\\1\end{bmatrix}+\begin{bmatrix}0\\0\\\frac{7}{16}\end{bmatrix}-2\begin{bmatrix}4&-1&0\\-1&2&2\\0&2&3\end{bmatrix}\begin{bmatrix}\frac{23}{32}\\\frac{9}{16}\\0\end{bmatrix}=\begin{bmatrix}1\\2\\1\end{bmatrix}$$

$$\begin{bmatrix}\lambda_1\\\lambda_2\\\lambda_3\end{bmatrix}=\begin{bmatrix}0\\0\\\frac{7}{16}\end{bmatrix}\geq\mathbf{0}$$

$$\lambda_0=\frac{45}{16}\geq 0$$

$$A^T x - b = \begin{bmatrix} -2 & -1 & -1 \end{bmatrix} \begin{bmatrix} \frac{23}{32} \\ \frac{9}{16} \\ 0 \end{bmatrix} + 2 = 0 \leq 0$$

$$x = \begin{bmatrix} \frac{23}{32} \\ \frac{9}{16} \\ 0 \end{bmatrix} \geq \mathbf{0}$$

Therefore, the KKT condition is satisfied.
Also, let $f(x) = x^T H x + c^T x$ be the objective function of the problem. Then, since H is symmetric by observation,

$$\nabla f(x) = 2Hx + c$$

$$\nabla^2 f(x) = 2H = \begin{bmatrix} 8 & -2 & 0 \\ -2 & 4 & 4 \\ 0 & 4 & 6 \end{bmatrix}$$

Let $u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \in R^3$ be any vector, then

$$u^T \nabla^2 f(x) u = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} 8 & -2 & 0 \\ -2 & 4 & 4 \\ 0 & 4 & 6 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

$$= \begin{bmatrix} 8u_1 - 2u_2 & -2u_1 + 4u_2 + 4u_3 & 4u_2 + 6u_3 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix}$$

$$= 8u_1^2 - 2u_2u_1 - 2u_1u_2 + 4u_2^2 + 4u_3u_2 + 4u_2u_3 + 6u_3^2$$
$$= 8u_1^2 - 4u_2u_1 + 4u_2^2 + 8u_3u_2 + 6u_3^2$$

$$= \left(2\sqrt{2}u_1 - \frac{\sqrt{2}}{2}u_2\right)^2 + \left(\frac{\sqrt{14}}{2}u_2 + \frac{4\sqrt{14}}{7}u_3\right)^2 + \frac{10}{7}u_3^2 \geq 0$$

Therefore, $\nabla^2 f(x)$ is PSD by definition, so the objective function of the NLP $f(x) = x^T H x + c^T x$ is convex by Proposition 4.9.1. Then, since the inequality constraints $c_0(x) = A^T x - b$ and $c_i(x) = -x_i, i = 1,2,3$ are affine, and affine functions are convex by Proposition 3.15, the inequality constraints are all convex. Therefore, the NLP is a convex optimization problem by definition. Also, since $(x_1, x_2, x_3, \lambda_0, \lambda_1, \lambda_2, \lambda_3) = \left(\frac{23}{32}, \frac{9}{16}, 0, \frac{45}{16}, 0, 0, \frac{7}{16}\right)$ is a KKT point of the problem,

$x = \begin{bmatrix} \frac{23}{32} \\ \frac{9}{16} \\ 0 \end{bmatrix}$ is a global minimizer by Theorem 10.18.

Question 2

(4 points)

Given the following constrained least square problem:

$$\min_{x\in\mathbb{R}^n} \quad ||Ax-b||_2^2$$
$$\text{s.t.} \quad Gx = h$$

where $A \in \mathbb{R}^{m\times n}$ and $\text{rank}(A) = n$, $G \in \mathbb{R}^{p\times n}$ and $\text{rank}(G) = p$.

1. Derive the dual problem of this problem.
2. Find an expression of the primal and dual optimal solution.

Solution:

Let $G = \begin{bmatrix} G_1 \\ \vdots \\ G_p \end{bmatrix}$, where $G_i^T \in R^n, i = 1,2,\dots,p$. Let $\lambda = \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_p \end{bmatrix}$ be the lagrangian multipliers, where $\lambda_i \in R, i = 1,2,\dots,p$. Let $h = \begin{bmatrix} h_1 \\ \vdots \\ h_p \end{bmatrix}$, where $h_i \in R, i = 1,2,\dots,p$.

Then, the Lagrangian function of the problem is:

$$L(x,\lambda) = \|Ax-b\|_2^2 + \sum_{i=1}^{p} \lambda_i (G_i x - h_i) = \|Ax-b\|_2^2 + \lambda^T (Gx - h)$$

Then, we try to get the infimum of $L(x,\lambda)$:

$$\nabla_x L(x,\lambda) = 2A^T(Ax-b) + G^T\lambda$$

Let $\nabla_x L(x,\lambda) = 0$ to get the critical point:

$$2A^T(Ax-b) + G^T\lambda = 0$$
$$2A^TAx - 2A^Tb + G^T\lambda = 0$$
$$A^TAx = A^Tb - \frac{1}{2}G^T\lambda$$

Then, we check the invertibility of $A^TA \in R^{n\times n}$:

We first get the SVD of A:

$$A = U\Sigma V^T$$
$$A^TA = (U\Sigma V^T)^T U\Sigma V^T = V\Sigma^T U^T U\Sigma V^T$$

Since U and V are orthogonal, we have

$$A^TA = V\Sigma^T U^T U\Sigma V^T = V\Sigma^T \Sigma V^T = V\Sigma^T \Sigma V^T = V\Sigma^T \Sigma V^{-1}$$

Note that since $rank(A) = n = rank(U\Sigma V^T) \le \min\{rank(V), rank(\Sigma), rank(U)\}$ and $rank(\Sigma) \le n \le \min\{m,n\}$ since $\Sigma \in R^{m\times n}$, we get $rank(\Sigma) = n$. Then, let $k_i \neq 0, i = 1,2,\dots,n$ be the singular values of A, we have $\Sigma = \begin{bmatrix} k_1 & & \\ & \ddots & \\ & & k_n \\ & \mathbf{0} & \end{bmatrix}$. Thus,

$\Sigma^T\Sigma = \begin{bmatrix} k_1 & & & \\ & \ddots & & \mathbf{0} \\ & & k_n & \end{bmatrix}\begin{bmatrix} k_1 & & \\ & \ddots & \\ & & k_n \\ & \mathbf{0} & \end{bmatrix} = \begin{bmatrix} k_1^2 & & \\ & \ddots & \\ & & k_n^2 \end{bmatrix}$ is a diagonal matrix with no 0-diagonal entry either. Also, since $A^TA = V\Sigma^T\Sigma V^{-1}$ is also the eigen-decomposition of A^TA, $\Sigma^T\Sigma$ is the eigenvalue diagonal matrix of A^TA, so A^TA doesn't have 0 as its eigenvalue since $\Sigma^T\Sigma$ has no 0-diagonal entry. That said, $A^TAz = 0$ has the only trivial solution $z = 0$, so $nullity(A^TA) = 0$. Hence, by rank-nullity theorem, $rank(A^TA) = n - nullity(A^TA) = n$, A^TA is invertible.

Thus, $x = (A^TA)^{-1}\left(A^Tb - \frac{1}{2}G^T\lambda\right)$ is the only critical point.

Then, we check the convexity of $L(x,\lambda)$:

$$\nabla_x^2 L(x,\lambda) = 2A^T(Ax - b) + G^T\lambda = 2A^TA$$

Since for any $v \in R^n$, $v^T\nabla_x^2 L(x,\lambda)v = 2v^TA^TAv = 2\|Av\|_2^2 \geq 0$, $\nabla_x^2 L(x,\lambda)$ is PSD, which means $q(x) = L(x,\lambda)$ is convex by Proposition 4.9.1.

Hence, the objective function of the dual problem is

$$g(\lambda) = \inf_{x\in dom(f)} L(x,\lambda)$$

$$= \left\|A(A^TA)^{-1}\left(A^Tb - \frac{1}{2}G^T\lambda\right) - b\right\|_2^2$$

$$+ \lambda^T\left[G(A^TA)^{-1}\left(A^Tb - \frac{1}{2}G^T\lambda\right) - h\right]$$

Hence, the dual problem is

$$\max g(\lambda) = \left\|A(A^TA)^{-1}\left(A^Tb - \frac{1}{2}G^T\lambda\right) - b\right\|_2^2$$

$$+ \lambda^T\left[G(A^TA)^{-1}\left(A^Tb - \frac{1}{2}G^T\lambda\right) - h\right]$$

$$s.t. \lambda \in R^p$$

To get the optimal solutions for the primal problem and the dual problem respectively:

Let $f(x) = \|Ax - b\|_2^2$ be the objective function of the primal problem, then

$$\nabla f(x) = 2A^T(Ax - b) = 2A^TAx - 2A^Tb$$

$$\nabla^2 f(x) = 2A^TA$$

Note that since we have proved that $2A^TA$ is PSD, so f is a convex function by Proposition 4.9.1. Then, since the equality constraints of the primal problem are all affine functions, the primal problem is a convex optimization problem by definition. Then, since the dual problem is a convex unconstrained NLP, which is easier to solve, we try to find the critical point of the dual objective function. First, we rearrange the dual problem into an equivalent form:

$$\begin{aligned}\min s(\lambda) &= -g(\lambda)\\ &= -\left\|A(A^TA)^{-1}\left(A^Tb-\frac{1}{2}G^T\lambda\right)-b\right\|_2^2\\ &-\lambda^T\left[G(A^TA)^{-1}\left(A^Tb-\frac{1}{2}G^T\lambda\right)-h\right]\\ &\quad s.t.\,\lambda\in R^p\end{aligned}$$

Then,

$$\begin{aligned}\nabla s(\lambda) &= G(A^TA)^{-1}\left(A^Tb-\frac{1}{2}G^T\lambda\right)-G(A^TA)^{-1}A^Tb-G(A^TA)^{-1}\left(A^Tb-\frac{1}{2}G^T\lambda\right)\\ &+h+\frac{1}{2}G(A^TA)^{-1}G^T\lambda\\ &= G(A^TA)^{-1}A^Tb-\frac{1}{2}G(A^TA)^{-1}G^T\lambda-G(A^TA)^{-1}A^Tb\\ &-G(A^TA)^{-1}A^Tb+\frac{1}{2}G(A^TA)^{-1}G^T\lambda+h+\frac{1}{2}G(A^TA)^{-1}G^T\lambda\\ &= -G(A^TA)^{-1}A^Tb+\frac{1}{2}G(A^TA)^{-1}G^T\lambda+h\end{aligned}$$

Let $\nabla s(\lambda)=0$,

$$-G(A^TA)^{-1}A^Tb+\frac{1}{2}G(A^TA)^{-1}G^T\lambda+h=0$$

$$\frac{1}{2}G(A^TA)^{-1}G^T\lambda=G(A^TA)^{-1}A^Tb-h$$

$$G(A^TA)^{-1}G^T\lambda=2G(A^TA)^{-1}A^Tb-2h$$

Then, we check the invertibility of $G(A^TA)^{-1}G^T$:

Since $G\in R^{p\times n}, (A^TA)^{-1}\in R^{n\times n}, G^T\in R^{n\times p}$, we know $G(A^TA)^{-1}G^T\in R^{p\times p}$. From the previous work, since $(A^TA)^{-1}=(V\Sigma^T\Sigma V^{-1})^{-1}$ and V and $\Sigma^T\Sigma$ are invertible because V is orthogonal and $rank(\Sigma^T\Sigma)=n$, we know $(A^TA)^{-1}=(V\Sigma^T\Sigma V^{-1})^{-1}=V(\Sigma^T\Sigma)^{-1}V^{-1}=V(\Sigma^T\Sigma)^{-1}V^T$. Let $V=[v_1\ \cdots\ v_n]$, where $v_i, i=1,2,\dots,n$ are eigenvectors of A^TA, then

$$(A^TA)^{-1}=V(\Sigma^T\Sigma)^{-1}V^T=V\left(\begin{bmatrix}k_1^2 & & \\ & \ddots & \\ & & k_n^2\end{bmatrix}\right)^{-1}V^T=V\begin{bmatrix}\frac{1}{k_1^2} & & \\ & \ddots & \\ & & \frac{1}{k_n^2}\end{bmatrix}V^T$$

$$=[v_1\ \cdots\ v_n]\begin{bmatrix}\frac{1}{k_1^2} & & \\ & \ddots & \\ & & \frac{1}{k_n^2}\end{bmatrix}\begin{bmatrix}v_1^T\\ \vdots\\ v_n^T\end{bmatrix}=\sum_{i=1}^{n}\frac{1}{k_i^2}v_iv_i^T$$

Thus,

$$G(A^TA)^{-1}G^T=G\left(\sum_{i=1}^{n}\frac{1}{k_i^2}v_iv_i^T\right)G^T=\sum_{i=1}^{n}\frac{1}{k_i^2}Gv_iv_i^TG^T=\sum_{i=1}^{n}\frac{1}{k_i^2}\|Gv_i\|_2^2$$

For any $z \in null(G(A^TA)^{-1}G^T)$, we have

$$G(A^TA)^{-1}G^Tz = 0$$

$$\sum_{i=1}^{n} \frac{1}{k_i^2} \|Gv_i\|_2^2 z = 0$$

Since $rank(G) = p$, we know $nullity(G) = n - p < n$. Suppose $\|Gv_i\|_2^2 = 0$ for all $i = 1,2,\dots,n$ for a contradiction, then $Gv_i = 0$ for all $i = 1,2,\dots,n$ by the definition of vector norms. Thai said, $v_1, \dots, v_n \in null(G)$. However, note that $v_1, \dots, v_n$ are linearly independent since they are orthogonal to each other, $nullity(G) \geq n$, which contradicts the fact that $nullity(G) < n$. Hence, $\|Gv_i\|_2^2 > 0$ for some $i = 1,2,\dots,n$. Also, since $k_i > 0, i = 1,2,\dots,n$, we know $\frac{1}{k_i} > 0, i = 1,2,\dots,n$, which means $\frac{1}{k_i^2}\|Gv_i\|_2^2 > 0$ for all $i = 1,2,\dots,n$. Therefore, $z = 0$ is the only solution to $G(A^TA)^{-1}G^Tz = \sum_{i=1}^{n}\frac{1}{k_i^2}\|Gv_i\|_2^2 z = 0$. Thus, $nullity(G(A^TA)^{-1}G^T) = 0$, so $rank(G(A^TA)^{-1}G^T) = p - 0 = p$, $G(A^TA)^{-1}G^T$ is invertible.

Therefore,

$$G(A^TA)^{-1}G^T\lambda = 2G(A^TA)^{-1}A^Tb - 2h$$

$$\lambda = 2(G(A^TA)^{-1}G^T)^{-1}G(A^TA)^{-1}A^Tb - 2(G(A^TA)^{-1}G^T)^{-1}h$$

Finally, we substitute it back to the KKT stationary condition:

$$x = (A^TA)^{-1}\Big(A^Tb - \frac{1}{2}G^T[2(G(A^TA)^{-1}G^T)^{-1}G(A^TA)^{-1}A^Tb - 2(G(A^TA)^{-1}G^T)^{-1}h]\Big)$$

Therefore, the optimal solution for the primal problem is $x^* = (A^TA)^{-1}\Big(A^Tb - \frac{1}{2}G^T[2(G(A^TA)^{-1}G^T)^{-1}G(A^TA)^{-1}A^Tb - 2(G(A^TA)^{-1}G^T)^{-1}h]\Big)$ and the optimal solution for the dual problem is $\lambda^* = 2(G(A^TA)^{-1}G^T)^{-1}G(A^TA)^{-1}A^Tb - 2(G(A^TA)^{-1}G^T)^{-1}h$.