SOLUTIONS HW 4

1 Problem 1

The $\mathcal{S}$ is a closed convex set. The minimizer x^* is the projection of 0 in $\mathcal{S}$. Thus, in order to show that $x^* = A^T(AA^T)^{-1}b$ is the projection of 0 in $\mathcal{S}$ it suffices to show that

$$(x^* - 0)^T(x - x^*) \geq 0, \quad \forall x \in \mathcal{S} \tag{1}$$

Indeed,

$$\begin{aligned}
(x^*)^T(x - x^*) &= (b^T((AA^T)^{-1})^T A)(x - A^T(AA^T)^{-1}b) \\
&= b^T((AA^T)^{-1})^T Ax - b^T((AA^T)^{-1})^T AA^T(AA^T)^{-1}b \\
&= b^T((AA^T)^{-1})^T b - b^T((AA^T)^{-1})^T b, \quad \textit{we used Ax=b} \\
&= 0
\end{aligned} \tag{2}$$

2 Problem 2

1. Let us consider a vector x such that $x^T AA^T = 0$. Multiplying by x on the right, we have

$$x^T AA^T x = 0 \Rightarrow \|x^T A\|^2 = 0 \tag{3}$$

   Since the rows of A are linearly independent, we must have $x = 0$. Hence,

$$x^T AA^T = 0 \Rightarrow x = 0 \tag{4}$$

   which implies that AA^T is invertible.

2. In order to verify that $z^* = x - A^T(AA^T)^{-1}(Ax - b)$ is the project of x on $\mathcal{S}$ it suffices to show that $(z^* - x)^T(z - z^*) \geq 0$ for all $z \in \mathcal{S}$. Indeed,

$$\begin{aligned}
&(z^* - x)^T(z - z^*) \\
=&(x^T - (Ax - b)^T((AA^T)^{-1})^T A - x^T)(z - x + A^T(AA^T)^{-1}(Ax - b)) \\
=&(b - Ax)^T((AA^T)^{-1})^T Az + (Ax - b)^T((AA^T)^{-1})^T Ax - (Ax - b)^T((AA^T)^{-1})^T AA^T(AA^T)^{-1}(Ax - b) \\
=&(b - Ax)^T((AA^T)^{-1})^T b + (Ax - b)^T((AA^T)^{-1})^T Ax - (Ax - b)^T((AA^T)^{-1})^T Ax + (Ax - b)^T((AA^T)^{-1})^T b \\
=&0
\end{aligned} \tag{5}$$

3 Problem 3

1. The derivative of the Lagrangian is

$$\nabla f + \lambda \nabla h = 0 \Rightarrow 2x + \lambda \mathbf{1} = 0 \tag{6}$$

   This implies that $x_1 = \ldots = x_n = -\lambda/2$ and $\sum_{i=1}^n x_i = 2$. Hence, $x^* = [2/n, \ldots, 2/n]^T$.

   We can also check $\nabla^2_{xx} L(x^*, \lambda^*) = 2I \succ 0$. Hence x^* is a local min. Since f is coercive, we know the global min exists and the only local min x^* is also the global min.

2. The derivative of the Lagrangian is

$$\nabla f + \lambda \nabla h = 0 \Rightarrow \mathbf{1} + \lambda 2x = 0 \tag{7}$$

This implies that $x_1 = \ldots = x_n = -1/(2\lambda)$ and $\|x\|^2 = 1$. Hence, we have two stationary points $x^* = [1/\sqrt{n}, \ldots, 1/\sqrt{n}]^T$ or $x^* = -[1/\sqrt{n}, \ldots, 1/\sqrt{n}]^T$. However, for $x^* = -[1/\sqrt{n}, \ldots, 1/\sqrt{n}]^T$, we have

$$\nabla^2 f(x^*) + \lambda^* \nabla^2 h(x^*) = -\sqrt{n} I \prec 0$$

This is not a local min. For $x^* = -[1/\sqrt{n}, \ldots, 1/\sqrt{n}]^T$, we have

$$\nabla^2 f(x^*) + \lambda^* \nabla^2 h(x^*) = \sqrt{n} I \succ 0$$

This is a local min. Since the feasible set is compact, we know the global min exists and this point will also be the local min.

3. The derivative of the Lagrangian is

$$\begin{aligned} \nabla f + \lambda \nabla h = 0 &\Rightarrow 2x + 2\lambda Q x = 0 \\ &\Rightarrow (Q - \mu I)x = 0, \quad \text{where } \mu = -1/\lambda \end{aligned} \tag{8}$$

We observe that μ stands for an eigenvalue and thus x^* is the corresponding eigenvector. Also, if we multiply (8) on the left by x^T we have

$$x^T(x + \lambda Q x) = 0 \Rightarrow \|x\|^2 + \lambda x^T Q x = 0 \Rightarrow \|x\|^2 = -\lambda = 1/\mu \tag{9}$$

Hence $\|x^*\|^2 = 1/\mu$ in order to minimize $\|x\|^2$ the x^* has to be the eigenvector which corresponds to the maximum eigenvalue of Q say μ^* such that $(x^*)^T Q x^* = 1$. For a normalized eigenvector u which corresponds to μ^*, we have $u^T Q u = \mu^*$. Thus, $x^* = \pm u/\sqrt{\mu^*}$. If the multiplicity of the largest eigenvalue of Q is 1, then it is straightforward to use the second-order sufficient condition to show that for $y^T \nabla h(x^*) = c y^T u = 0$, we have

$$y^T(\nabla^2 f(x^*) + \lambda \nabla^2 h(x^*))y = y^T(2I - \frac{2}{\mu^*}Q)y$$

In general, we know $Q \preceq \mu^* I$. More importantly, if y is orthogonal to the eigenvector u, then we have

$$y^T(2I - \frac{2}{\mu^*}Q)y > 0.$$

The second-order sufficient condition holds. Hence x^* is the local min. Since the feasible set is compact, the global min exists and we can verify it is achieved by the above solution.

More comments: For this problem, it is OK to assume the multiplicity of the largest eigenvalue of Q is 1. In general, the multiplicity of the largest eigenvalue of Q is larger than 1. However, if y is orthogonal to the eigenspace for μ^*, we still have

$$y^T(2I - \frac{2}{\mu^*}Q)y > 0.$$

Then a generalized version of the second-order sufficient condition can be applied to guarantee local optimality. The only difference is that now x^* forms a set, and the y vector in the sufficient condition should be taken to be orthogonal to the eigenspace of μ^*.

4 Problem 4

First, we note that $\nabla h = [2, 1] \neq 0$, thus the regularity conditions are satisfied. The derivative of the Lagrangian is

$$\nabla f + \lambda \nabla h = 0 \Rightarrow [2 - x_2,\, 1 - x_1]^T + \lambda [2,\, 1]^T = 0 \tag{10}$$

which implies that $2 - x_2 = 2(1 - x_1) \Rightarrow 2x_1 = x_2$. The $2x_1 = x_2$ together with the constraint $2x_1 + x_2 = 2$ implies that $x^* = [0.5, 1]^T$.

To show that this is the global min of the original problem, we can substitute $x_2 = 2 - 2x_1$ into f and obtain $f = 2x_1 + (1 - x_1)(2 - 2x_1) = 2 - 2x_1 + 2x_1^2 = 2(x_1 - 0.5)^2 + 1.5 \geq 1.5$. Therefore, x^* does lead to the global minimum value.

It is also OK to use the second-order sufficient condition and the corollary to Weierstrass' Theorem to show x^* is the global min. We have

$$\nabla^2_{xx} L(x^*, \lambda) = \nabla^2 f(x^*) = \frac{1}{2}\begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$$

Suppose $y^\top \nabla h(x^*) = 2y_1 + y_2 = 0$, i.e. $y_2 = -2y_1$. Then we have

$$y^\top \nabla^2_{xx} L(x^*, \lambda) y = 2y_1^2 > 0$$

This guarantees x^* is a local min. On the feasible set, we have $f = 2x_1^2 - 2x_1 + 2$ which is coercive. Hence by the corollary to Weierstrass' Theorem, the global min exists and x^* is the global min. Notice that f is not coercive on $\mathbf{R}^2$. It is only coercive when we enforce the feasibility condition $2x_1 + x_2 = 2$.

5 Problem 5

1. $$\nabla f = [4(x_1 - 1) + \cos(x_1),\ 8(x_2 - 2) + \sin(x_2)]^T, \quad \nabla^2 f = \begin{pmatrix} 4 - \sin(x_1) & 0 \\ 0 & 8 + \cos(x_2) \end{pmatrix} \tag{11}$$

   Since $4 - \sin(x_1) > 0$ and $8 + \cos(x_2) > 0$ the $\nabla^2 f$ is positive definite and thus f is convex.

2. The projected Newton iteration is

   $$\mathbf{x}_{k+1} = [\mathbf{x}_k - \alpha(\nabla^2 f(\mathbf{x}_k))^{-1} \nabla f(\mathbf{x}_k)]^{\mathcal{S}} \tag{12}$$

   If $(x_1, x_2) \notin \mathcal{S}$ then we find the closest number in $\mathcal{S}$, i.e.

   $$\begin{cases} x_i^{\mathcal{S}} = -1, & x_i < -1 \\ x_i^{\mathcal{S}} = 1, & x_i > 1 \\ x_i^{\mathcal{S}} = x_i, & \text{otherwise} \end{cases} \tag{13}$$

   for $i \in \{1, 2\}$.

3. The algorithm converges