SUN Yilin
520030910361

Homework 9

1

(a).

By substituting x_1 with $1-2x_2$ and simplification we get $f = x_2^2 - 2x_2 - \frac{1}{2}$. We can easily see $x_2^* = 1$, then we calculate that $x_1^* = -1$. So optimal variable $(x_1^*, x_2^*) = (-1, 1)$ and optima $f^* = -\frac{3}{2}$.

(b).

$$\mathcal{L} = \frac{1}{2}x_1^2 + x_1x_2 + x_2^2 - x_1 - 3x_2 + \lambda(x_1 + 2x_2 - 1)$$

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x_1} = x_1 + x_2 - 1 + \lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial x_2} = x_1 + 2x_2 - 3 + 2\lambda = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = x_1 + 2x_2 - 1 = 0 \end{cases}$$

Solve that equation and we get

$$\begin{cases} x_1^* = -1 \\ x_2^* = 1 \\ \lambda^* = 1 \end{cases}$$

Also, $-f^* = \frac{3}{2}$.

2

(a).

The Lagrange function is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \frac{1}{2}\boldsymbol{x}^T\boldsymbol{Q}\boldsymbol{x} + \boldsymbol{g}^T\boldsymbol{x} + c + \boldsymbol{\lambda}^T(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

So the Lagrange condition is

$$\begin{cases} \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{Q}\boldsymbol{x}^* + \boldsymbol{g} + \boldsymbol{A}^T\boldsymbol{\lambda}^* = \boldsymbol{0} \\ \nabla_{\boldsymbol{\lambda}}\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*) = \boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b} = \boldsymbol{0} \end{cases}$$

(b).

$\boldsymbol{Q} \succ \boldsymbol{O}$ so $\boldsymbol{Q}$ is invertible. So we get

$$\boldsymbol{x}^* = -\boldsymbol{Q}^{-1}\boldsymbol{g} - \boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}^*$$

Then from $\boldsymbol{A}\boldsymbol{x}^* - \boldsymbol{b} = \boldsymbol{0}$ we get

$$\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{\lambda}^* = -\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{g} - \boldsymbol{b}$$

Now we show $\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T \succ \boldsymbol{O}$.
To show that, we only need to show $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{x} > 0, \forall \boldsymbol{x} \in \mathbb{R}^n, \boldsymbol{x} \neq \boldsymbol{0}$. Let $\boldsymbol{y}^T = \boldsymbol{x}^T\boldsymbol{A}$, then $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T\boldsymbol{x} = \boldsymbol{y}^T\boldsymbol{Q}^{-1}\boldsymbol{y}$. Now we need to show $\boldsymbol{y}^T\boldsymbol{Q}^{-1}\boldsymbol{y} > 0$.
Firstly note that $\boldsymbol{A}^T\boldsymbol{x} = \boldsymbol{y}$ and rank$\boldsymbol{A}^T = k$, meaning that when $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{y} \neq \boldsymbol{0}$.
Secondly note that $\boldsymbol{Q}^{-1} \succ \boldsymbol{O}$ because $\boldsymbol{Q} \succ \boldsymbol{O}$. So $\boldsymbol{y}^T\boldsymbol{Q}^{-1}\boldsymbol{y} > 0, \forall \boldsymbol{y} > \boldsymbol{0}$.
So $\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T \succ \boldsymbol{O}$, hence invertible.
So we get

$$\begin{cases} \boldsymbol{x}^* = -\boldsymbol{Q}^{-1}\boldsymbol{g} + \boldsymbol{Q}^{-1}\boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{g} + \boldsymbol{b}) \\ \boldsymbol{\lambda}^* = -(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{A}^T)^{-1}(\boldsymbol{A}\boldsymbol{Q}^{-1}\boldsymbol{g} + \boldsymbol{b}) \end{cases}$$

(c).

Here $\boldsymbol{Q} = \boldsymbol{I}, \boldsymbol{g} = -\boldsymbol{x}_0, c = \frac{1}{2}\boldsymbol{x}_0^T\boldsymbol{x}_0$. By using the formulas in part (b) we get

$$\boldsymbol{x}^* = \boldsymbol{x}_0 + \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}(-\boldsymbol{A}\boldsymbol{x}_0 + \boldsymbol{b})$$

When $\boldsymbol{x}_0 = \boldsymbol{0}$ we get

$$\boldsymbol{x}^* = \boldsymbol{A}^T(\boldsymbol{A}\boldsymbol{A}^T)^{-1}\boldsymbol{b}$$

which is exactly the result on slides.

(d).

Here $\boldsymbol{A} = \boldsymbol{w}^T$. By using the formula in part (c) we get

$$\boldsymbol{x}^* - \boldsymbol{x}_0 = \frac{-\boldsymbol{w}^T\boldsymbol{x}_0 + b}{||\boldsymbol{w}||^2}\boldsymbol{w}$$

So dist$(\boldsymbol{x}_0, P)$=

$$||\boldsymbol{x}^* - \boldsymbol{x}_0||$$

which is

$$\frac{|\boldsymbol{w}^T\boldsymbol{x}_0 - b|}{||\boldsymbol{w}||}$$

which is the result on slides.

3

The Lagrange function is

$$\mathcal{L} = x_1x_2 + \lambda(x_1^2 + 4x_2^2 - 1)$$

And the Lagrange condition is

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial x_1} = x_2 + 2\lambda x_1 = 0 \\ \frac{\partial \mathcal{L}}{\partial x_2} = x_1 + 8\lambda x_2 = 0 \\ \frac{\partial \mathcal{L}}{\partial \lambda} = x_1^2 + 4x_2^2 - 1 = 0 \end{cases}$$

We solve the equations and note that not all solutions are global minimum. So we must check each solutions. Finally we get the optima variables $(x_1^*, x_2^*) = (\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{4})$ or $(x_1^*, x_2^*) = (\frac{-\sqrt{2}}{2}, \frac{\sqrt{2}}{4})$, and optima value $f^* = -\frac{1}{4}$.

4

(a).

The Lagrange function is

$$\mathcal{L} = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} + \lambda(\boldsymbol{x}^T \boldsymbol{x} - 1)$$

And the Lagrange condition is

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \boldsymbol{x}} = 2(\boldsymbol{A} + \lambda^* \boldsymbol{I})\boldsymbol{x^*} = \boldsymbol{0} \\ \frac{\partial \mathcal{L}}{\partial \lambda} = \boldsymbol{x}^{*T} \boldsymbol{x}^* - 1 = 0 \end{cases}$$

The first equation can be transformed into

$$\boldsymbol{A}\boldsymbol{x}^* = -\lambda^* \boldsymbol{x}^*$$

Also note that we want to minimize

$$\boldsymbol{x}^{*T} \boldsymbol{A} \boldsymbol{x}^*$$

which is

$$-\lambda^*$$

To minimize it, note that $-\lambda^*$ is an eigenvalue of $\boldsymbol{A}$, so the optimal value is λ_1. So $\boldsymbol{A}\boldsymbol{x}^* = \lambda_1 \boldsymbol{x}^*$, so $\boldsymbol{x}^*$ is an eigenvector of $\boldsymbol{A}$ associated to λ_1.

(b).

i)

The Lagrange function is

$$\mathcal{L} = \boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x} + \alpha(\boldsymbol{x}^T \boldsymbol{x} - 1) + \beta \boldsymbol{v}_1^T \boldsymbol{x}$$

And the Lagrange condition is

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \boldsymbol{x}} = 2(\boldsymbol{A} + \alpha \boldsymbol{I})\boldsymbol{x^*} + \beta \boldsymbol{v}_1 = \boldsymbol{0} \\ \frac{\partial \mathcal{L}}{\partial \alpha} = \boldsymbol{x}^{*T} \boldsymbol{x}^* - 1 = 0 \\ \frac{\partial \mathcal{L}}{\partial \beta} = \boldsymbol{v}_1^T \boldsymbol{x}^* = 0 \end{cases}$$

The first condition can be transformed into

$$\boldsymbol{A}\boldsymbol{x}^* = -\alpha \boldsymbol{x}^* - \frac{\beta}{2} \boldsymbol{v}_1$$

So there exists $c_0 = -\alpha, c_1 = -\frac{\beta}{2}$ s.t.

$$\boldsymbol{A}\boldsymbol{x}^* = c_0 \boldsymbol{x}^* + c_1 \boldsymbol{v}_1$$

ii)

We left multiply $\boldsymbol{A}\boldsymbol{x}^* = c_0 \boldsymbol{x}^* + c_1 \boldsymbol{v}_1$ by $\boldsymbol{v}_1^T$ and get $\boldsymbol{v}_1^T \boldsymbol{A}\boldsymbol{x}^* = c_0 \boldsymbol{v}_1^T \boldsymbol{x}^* + c_1 \boldsymbol{v}_1^T \boldsymbol{v}_1$. Note that $\boldsymbol{v}_1^T \boldsymbol{x}^* = 0$, so $\boldsymbol{v}_1^T \boldsymbol{A}\boldsymbol{x}^* = c_1 \boldsymbol{v}_1^T \boldsymbol{v}_1$. Then we take the transpose and get $\boldsymbol{x}^{*T} \boldsymbol{A} \boldsymbol{v}_1 = c_1 \boldsymbol{v}_1^T \boldsymbol{v}_1$. Note that $\boldsymbol{v}_1$ is an eigenvector associated to λ_1, so $c_1 \boldsymbol{v}_1^T \boldsymbol{v}_1 = \lambda_1 \boldsymbol{x}^{*T} \boldsymbol{v}_1 = 0$. Since $\boldsymbol{v}_1$ is nonzero, c_1 must be 0.

iii)

Since $c_1 = 0$, $\boldsymbol{A}\boldsymbol{x}^* = c_0 \boldsymbol{x}^*$. So $\boldsymbol{x}^*$ is an eigenvector associated with c_0 Similarly as part (a), we want to minimize $\boldsymbol{x}^{*T} \boldsymbol{A} \boldsymbol{x}^*$, which is c_0. But here to we cannot set c_0 as λ_1 because $\boldsymbol{x}^*$ and $\boldsymbol{v}_1$ are orthogonal to each other hence they cannot both be associated with λ_1. So the optimal value is λ_2 and $\boldsymbol{x}^*$ is an eigenvector associated with λ_2