Problem 4.4.8

Given $\inf \left\{ \sum_{i=1}^{n} \frac{c_i}{x_i} \mid \sum_{i=1}^{n} a_i x_i \leq b \ , x > 0 \right\}$

$\Rightarrow \mathcal{L}(x, \lambda, \mu) = \sum_{i=1}^{n} \frac{c_i}{x_i} + \sum_{i=1}^{n} \lambda_i a_i x_i - b \sum_{i=1}^{n} \lambda_i - \mu^T x$

for $\inf_x \mathcal{L}(x, \lambda)$ to be finite, we require

$\nabla \mathcal{L}(x, \lambda, \mu) = \cancel{\sum_{i=1}^{n}} \ 0$

looking at the partial derivative of x_i, we have

$$-\frac{c_i}{x_i^2} + \lambda_i a_i - \mu_i = 0 \Rightarrow x_i = \frac{\sqrt{c_i}}{\sqrt{\lambda_i a_i - \mu_i}}$$

We observe that since $a_i, c_i > 0$, ~~x_i must be~~ $x_i > 0 \Rightarrow \mu_i = 0$

$\Rightarrow x_i = \frac{\sqrt{c_i}}{\sqrt{\lambda_i a_i}}$ and the lagrangian dual is

$$\sum_{i=1}^{n} \sqrt{c_i \lambda_i a_i} + \sum_{i=1}^{n} \sqrt{c_i \lambda_i a_i} - \sum_{i=1}^{n} b \lambda_i$$

$$= 2 \sum_{i=1}^{n} \sqrt{c_i \lambda_i a_i} - \sum_{i=1}^{n} b \lambda_i$$

$$= \min_{\lambda \geq 0} \left[2 \sum_{i=1}^{n} \sqrt{c_i \lambda_i a_i} - \sum_{i=1}^{n} b \lambda_i \right]$$

now taking the partial derivative λ_i, we get

$$\frac{c_i a_i}{\sqrt{c_i \lambda_i a_i}} - b = 0 \Rightarrow \sqrt{c_i a_i} \cdot \sqrt{\lambda_i} = \frac{c_i a_i}{b} \Rightarrow \sqrt{\lambda_i} = \frac{\sqrt{c_i a_i}}{b}$$

$$\Rightarrow \lambda_i = \frac{c_i a_i}{b^2}$$

Substitute back in

$$2 \sum_{i=1}^{n} \frac{c_i a_i}{b} - \sum_{i=1}^{n} \frac{c_i a_i}{b} = \sum_{i=1}^{n} \frac{c_i a_i}{b}$$

Problem 4.4.9 we have a positive-definite matrix A, $b>0$

$$\inf\{-\log\det X \mid \langle A, X\rangle \le b,\ X>0\}$$

we have the following lagrangian

$$\mathcal{L}(X,\lambda,\nu) = -\log\det X + \lambda\langle A, X\rangle - \lambda^T b - \nu^T X$$

$\Rightarrow \nabla\mathcal{L}(X,\lambda,\nu) = -X^{-1} + \lambda^T A - \nu = 0$ (by example 3.2.1 & lemma 4.4.7)

$\Rightarrow X = (\lambda^T A - \nu)^{-1}$, give A is a positive-definite matrix, we observe that $X>0 \Rightarrow \nu=0$, $\nu^T X = 0$

$\Rightarrow X = (\lambda^T A)^{-1} = \cancel{\ }$

$$\Rightarrow \inf_X \mathcal{L}(X,\lambda) = -\log\det((\lambda^T A)^{-1}) + \lambda\langle A, (\lambda^T A)^{-1}\rangle - \lambda^T b$$

$$= -\log\det((\lambda^T A)^{-1}) + \langle \lambda^T A, (\lambda^T A)^{-1}\rangle - \lambda^T b$$

$$= -\log\det((\lambda^T A)^{-1}) + I - \lambda^T b$$

Differentiating withrespect to λ we have

$$\lambda^T A - b = 0 \Rightarrow \lambda A = b \Rightarrow \lambda = b^T A^{-1}$$

$$\Rightarrow -\log\det((b^T A^{-1} A)^{-1}) + I - b^T A^{-1} b$$

$$= -\log\det(b^{-1}) + I - b^T A^{-1} b$$

Problem 4.4.10

1. We have $\mathcal{L}(x,\lambda) = f(x) + \lambda^T(Ax - b)$

$\Rightarrow \mathcal{L}(x^*, \lambda^*) = f(x^*) + \lambda^{*T}(Ax^* - b)$, $\lambda^* > 0$

Given (x^*, λ^*) are the optimal primal dual pair, this implies

$$\mathcal{L}(x^*, \lambda^*) = f(x^*) + \lambda^* g(x^*)$$
$$= f(x^*)$$

$\nabla_x \mathcal{L}(x^*, \lambda^*) = f'(x^*) = 0$ as f is convex, it's gradient at $x^* = 0$

$\nabla_\lambda \mathcal{L}(x^*, \lambda^*) = \nabla_\lambda f(x^*) = 0$

Hence, by the KKT conditions, we have that (x^*, λ^*) is a saddle point and therefore

$$\mathcal{L}(x^*, \lambda^*) \geq \mathcal{L}(x^*, \lambda)$$

Problem 4.4.10

2.
$$\max_{(x,\lambda)} \mathcal{L}(x,\lambda)$$
$$\text{s.t. } f'(x) + \lambda^T A = 0$$
$$\lambda \geq 0$$

We first point out that any feasible (x,λ) is a saddle point as it meets the KKT conditions.

Form 1. we know that

$$\mathcal{L}(x^*,\lambda^*) \geq \mathcal{L}(x^*,\lambda)$$

Now fixing λ, given $\mathcal{L}(x,\lambda)$ is convex. ~~from the subgradient~~ and given any feasible point (x,λ) meets the condition $\nabla_x \mathcal{L}(x,\lambda) = 0$, from the subgradient inequality, we have

$$\mathcal{L}(x^*,\lambda) - \mathcal{L}(x,\lambda) \geq \langle \nabla_x \mathcal{L}(x,\lambda), (x^*-x) \rangle$$
$$\Rightarrow \mathcal{L}(x^*,\lambda) - \mathcal{L}(x,\lambda) \geq \langle 0, (x^*-x) \rangle$$
$$\Rightarrow \mathcal{L}(x^*,\lambda) - \mathcal{L}(x,\lambda) \geq 0$$
$$\Rightarrow \mathcal{L}(x^*,\lambda) \geq \mathcal{L}(x,\lambda)$$

$$\Rightarrow \mathcal{L}(x^*,\lambda^*) \geq \mathcal{L}(x^*,\lambda) \geq \mathcal{L}(x,\lambda)$$
$$\Rightarrow \mathcal{L}(x^*,\lambda^*) \geq \mathcal{L}(x,\lambda)$$

Hence, we deduce that (x^*,λ^*) solves the Wolfe dual.