Assignment 3

Songtuan Lin u6162630

April 5, 2019

Q1

(a)

A path is start at node 1 and end at node n if the number of edges that outflow node 1 is equal to the number of edges that flow into node n. Hence, according to this property, the equation can be written as:

$$\sum_{i \neq 1} x_{1,i} = \sum_{j \neq n} x_{j,n} \tag{1}$$

(b)

The path pass through any node i except node 1 and node n only once indicate the number of edge that flow into node i small or equal to one. Hence, the constrain can be written as:

$$\sum_{j \neq i} x_{j,i} \leq 1 \quad \textit{for } i \neq 1 \textit{ and } i \neq n \tag{2}$$

(c)

The number of times a path goes into any node i is the same as the number of times a path goes out of the same node indicate that the number of edges that flow into node i is equal to the number of edges that flow out this node. Hence, the constrain can be expressed as:

$$\sum_{j \neq i} x_{i,j} = \sum_{k \neq i} x_{k,i} \quad \textit{for } i \neq 1 \textit{ and } i \neq n \tag{3}$$

(d)

It is easy to find out that the cost of a path which specified by $x_{i,j}$ is:

$$\sum_{i,j} c_{i,j} x_{i,j} \tag{4}$$

(e)

Following equation 1 to 4, the *Shortest Path Problem* can be formalized into optimization problem as:

$$\begin{aligned} \min_{x_{i,j}} \quad & \sum_{i,j} c_{i,j} x_{i,j} \\ s.t. \quad & \sum_{i \neq 1} x_{1,i} = \sum_{j \neq n} x_{j,n} \\ & \sum_{j \neq i} x_{j,i} \leq 1 \quad \textit{for } i \neq 1 \textit{ and } i \neq n \\ & \sum_{j \neq i} x_{i,j} = \sum_{k \neq i} x_{k,i} \quad \textit{for } i \neq 1 \textit{ and } i \neq n \\ & x_{i,j} \in \{0, 1\} \end{aligned} \tag{5}$$

(f)

Need modification

Q2

In order to prove that $\mathbf{x}^*$ is the optimal points of function $f_0(\mathbf{x})\frac{1}{2}\mathbf{x}^T\mathcal{P}\mathbf{x}+\mathbf{q}^T\mathbf{x}+r$ at point $\mathbf{x}^*$, we need to verify that for any $-1 \preceq \mathbf{x} \preceq 1$, the following equation:

$$\nabla f_0(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) \geq 0 \tag{6}$$

always hold. Indeed, the gradient of function f_0 at $\mathbf{x}^*$ is:

$$\nabla f_0(\mathbf{x}^*) = \mathcal{P}\mathbf{x}^* + \mathbf{q}^T = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \tag{7}$$

By substituting equation 7 to 6, we can rewrite the condition that need to be verified as:

$$\begin{aligned} \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}(\mathbf{x} - \begin{bmatrix} 1 \\ 0.5 \\ -1 \end{bmatrix}) &= -1(x_1-1)+2(x_3+1) \\ &\geq 0 \end{aligned} \tag{8}$$

As a result, we need to verify the minimum value of $-1(x_1-1)+2(x_3+1)$ is greater or equal to zero, which can be formalized as a *Linear Programming Problem* as:

$$\begin{aligned} \min_{x_1,x_3} \quad & -x_1+2x_2+3 \\ s.t. \quad & -1 \leq x_1 \leq 1 \\ & -1 \leq x_3 \leq 1 \end{aligned} \tag{9}$$

The optimal point of this problem can be obtained by simple inspection, which is $(1,-1)$. Hence, the minimum value of $-x_1+2x_2+3$ is 0, which means $\nabla f_0(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) \geq 0$ always hold and $\mathbf{x}^*$ is the optimal point.

Q3

(a)

The variable $\mathbf{x}$ can be decomposed as:

$$\mathbf{x} = \mathbf{y} + \mathbf{x_0}$$

Where $\mathbf{y} \in \mathcal{N}(\mathcal{A})$ can be any vector within the null space of $\mathcal{A}$ and $\mathbf{x_0}$ is the optimal point of original problem. The reason this equation always hold is:

$$\begin{aligned} \mathcal{A}\mathbf{x} &= \mathcal{A}(\mathbf{y}+\mathbf{x_0}) \\ &= \mathcal{A}\mathbf{y} + \mathcal{A}\mathbf{x_0} \\ &= 0 + \mathbf{b} \end{aligned} \tag{10}$$

Based on this decomposition, there is:

$$\nabla f_0(\mathbf{x_0})(\mathbf{x}-\mathbf{x_0}) = \mathbf{c}^T\mathbf{y} \tag{11}$$

If the optimal point $\mathbf{x_0}$ exist, then, $\mathbf{c}^T\mathbf{y} \geq 0$ for any $\mathbf{y} \in \mathcal{N}(\mathcal{A})$ must be true. However, since both $\mathbf{y}$ and $-\mathbf{y}$ are in $\mathcal{N}(\mathcal{A})$, there is:

$$\begin{cases} \mathbf{c}^T\mathbf{y} \geq 0 \\ -\mathbf{c}^T\mathbf{y} \geq 0 \end{cases} \tag{12}$$

Hence, $\mathbf{c}^T\mathbf{y}$ must equal to 0 for any $\mathbf{y} \in \mathcal{N}(\mathcal{A})$, which means: $\mathbf{c}$ is perpendicular to $\mathcal{N}(\mathcal{A})$. Furthermore, since $\mathcal{R}(\mathcal{A}) \perp \mathcal{N}(\mathcal{A})$, where $\mathcal{R}(\mathcal{A})$ is the row space of $\mathcal{A}$, we can conduct that $\mathbf{c} \in \mathcal{R}(\mathcal{A})$. Therefore, $\mathbf{c}$ is some linear combination of $\mathcal{A}$'s rows:

$$\mathbf{c} = \mathcal{A}^T\lambda \quad (\lambda \in \mathcal{R}^n) \tag{13}$$

Additionally, since the optimal point $\mathbf{x_0}$ must be feasible, $\mathbf{x_0}$ should satisfied $\mathcal{A}\mathbf{x} = \mathbf{b}$. As a result, the optimal solution of original problem under this circumstance is:

$$\begin{aligned} \mathbf{c}^T\mathbf{x_0} &= \lambda^T\mathcal{A}\mathbf{x_0} \\ &= \lambda^T\mathbf{b} \end{aligned} \tag{14}$$

Now, we tend to discuss the situation that $\mathbf{c} \notin \mathcal{R}(\mathcal{A})$. We can still decompose variable $\mathbf{x}$ with some $\mathbf{x}'$ which satisfied $\mathcal{A}\mathbf{x}' = \mathbf{b}$ as:

$$\mathbf{x} = \mathbf{y} + \mathbf{x}'$$

Where $\mathbf{y} \in \mathcal{N}(\mathcal{A})$. As a result, the objective function can be written as:

$$f_0 = \mathbf{c}^T(\mathbf{y} + \mathbf{x}')$$

Since $\mathbf{c}^T\mathbf{x}'$ is a constant, we are actually minimize $\mathbf{c}^T\mathbf{y}$. Furthermore, we can always find some $\mathbf{y}^*$ which satisfied $\mathbf{c}^T\mathbf{y}^* < 0$. As a result, by multiple $\mathbf{y}$ with any $t > 0$, $\mathbf{c}^T t\mathbf{y}$ is unbound below, which means, the optimal value is $-\infty$.

In conclusion, the optimal value of this problem is:

$$p^* = \begin{cases} -\infty & \mathbf{c} \notin \mathcal{R}(\mathcal{A}) \\ \lambda^T\mathbf{b} & \mathbf{c} \in \mathcal{R}(\mathcal{A}) \\ \infty & \mathbf{x} \textit{ is infeasible} \end{cases} \tag{15}$$

(b)

The objective function can be written as:

$$f_0(\mathbf{x}) = \sum_{i=1}^{n} c_i x_i \tag{16}$$

It is obvious that for each c_i, f_0 will achieve the minimum at $\mathbf{x}^*$ if:

$$x_i^* = \begin{cases} u & c_i \leq 0 \\ l & \textit{otherwise} \end{cases} \tag{17}$$

This can be proved through verifying $\nabla f_0(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$. Indeed, we have:

$$\nabla f_0(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = \sum_i c_i(x_i - x_i^*) \tag{18}$$

For each $c_i(x_i - x_i^*)$, if $c_i \leq 0$, then, $x_i^* = u$ and $x_i - x_i^* \leq 0$ as $x_i \leq u = x_i^*$. As a result, $c_i(x_i - x_i^*) \geq 0$ when $c_i \leq 0$. Additionally, by using similar method, we can also prove that $c_i(x_i - x_i^*) \geq 0$ when $c_i > 0$. As a result $c_i(x_i - x_i^*) \geq 0$ always hold, hence, $\nabla f_0(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0$.

(c)

Intuitively, the minimum value of $\sum_{i=1}^{n} c_i x_i$ under constrain $\sum_{i=1}^{n} x_i = 1$ and $0 \leq x_i \leq 1$ is:

$$c_1^* + c_2^* + \cdots + c_\alpha^* \tag{19}$$

Where c_i^* is the ith largest element within $\mathbf{c}$. This formulation indicate the minimum value of $\mathbf{c}^T\mathbf{x}$ is the sum of first α smallest element within $\mathbf{c}$. Indeed, in *Question 7(b)* of last assignment, we already proved that the maximum value of $\mathbf{y}^T\mathbf{x}$ for any $\mathbf{y}$ under constrain $\mathbf{1}^T\mathbf{x} = \alpha$ and $0 \preceq \mathbf{x} \preceq 1$ is $\sum_{i=1}^{\alpha} y_i^*$, which is the first α largest element within $\mathbf{y}$. By multiplying each $\mathbf{y}$ with -1, $-y_i^*$ becomes the ith smallest element within $-\mathbf{y}$ and $-\sum_{i=1}^{\alpha} y_i^*$ become the minimum value of $-\mathbf{y}^T\mathbf{x}$. Since $\mathbf{y}$ is arbitrary, we can conduct that the minimum value of $\mathbf{y}^T\mathbf{x}$ under constrain $\mathbf{1}^T\mathbf{x} = \alpha$ and $0 \preceq \mathbf{x} \preceq 1$ is the first α smallest number within $\mathbf{y}$. As a result, the optimal value of $\mathbf{c}^T\mathbf{x}$ is $\sum_{i=1}^{\alpha} c_i^*$

Furthermore, if α is not integer, we can decompose α with some fraction $0 \leq q \leq 1$ as:

$$\alpha = \lfloor \alpha \rfloor + q$$

Extend from equation 19, the optimal value under this situation is:

$$\sum_{i=1}^{\lfloor \alpha \rfloor} c_i^* + q c_{\lfloor \alpha \rfloor + 1}^* \tag{20}$$

Q4

We first discuss the situation that $\mathcal{A}^{-T}\mathbf{c} \preceq 0$. We take a point $\mathbf{x}^*$ satisfied $\mathcal{A}\mathbf{x}^*$. In order to verify $\mathbf{x}^*$ is the optimal point, we need to check the first-order condition always hold. Indeed, for any feasible points of this problem, these is:

$$\begin{aligned} \nabla f_0(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) &= \mathbf{c}^T(\mathbf{x}-\mathbf{x}^*) \\ &= \mathbf{c}^T\mathcal{A}^{-1}\mathcal{A}(\mathbf{x}-\mathbf{x}^*) \\ &= (\mathcal{A}^{-T}\mathbf{c})^T\mathcal{A}(\mathbf{x}-\mathbf{x}^*) \end{aligned} \tag{21}$$

Since $\mathcal{A}\mathbf{x} \preceq \mathbf{b}$ and $\mathcal{A}\mathbf{x}^* = \mathbf{b}$, we have:

$$\mathcal{A}(\mathbf{x}-\mathbf{x}^*) = \mathcal{A}\mathbf{x} - \mathcal{A}\mathbf{x}^* \preceq 0 \tag{22}$$

Combine this inequality with our assumption: $\mathcal{A}^{-T}\mathbf{c} \preceq 0$, we have:

$$(\mathcal{A}^{-T}\mathbf{c})^T\mathcal{A}(\mathbf{x}-\mathbf{x}^*) \geq 0 \tag{23}$$

As a result, we have verified that $\mathbf{x}^*$ is the optimal point. Furthermore, $\mathbf{x}^*$ is unique as $\mathcal{A}$ is square and invert-able:

$$\mathbf{x}^* = \mathcal{A}^{-1}\mathbf{b} \tag{24}$$

Therefore, the optimal value p^* under the situation: $\mathcal{A}^{-T}\mathbf{c} \preceq 0$ is:

$$\begin{aligned} p^* &= \mathbf{c}^T\mathbf{x}^* \\ &= \mathbf{c}^T\mathcal{A}^{-1}\mathbf{b} \end{aligned} \tag{25}$$

Then, we start to prove that $\mathbf{c}^T\mathbf{x}$ is unbound below if $\mathcal{A}^{-T}\mathbf{c} \succ 0$. Indeed, if $\mathbf{b} \prec 0$, then, for any $\mathbf{x}'$ in the feasible set, there is:

$$\mathcal{A}\mathbf{x}' \prec b \prec 0 \tag{26}$$

If $\mathbf{b} \succeq 0$, we can also choose a $\mathbf{x}'$ so that:

$$\mathcal{A}\mathbf{x}' \prec 0 \prec \mathbf{b} \tag{27}$$

This is because $\mathcal{A}$ is a square matrix and is full rank, which means the column space of $\mathcal{A}$ is the whole $\mathcal{R}^n$. As a result, the linear combination of $\mathcal{A}$'s columns: $\mathcal{A}\mathbf{x}$ can be any vector within $\mathcal{R}^n$. Hence, we can choose some specific combination $\mathbf{x}'$ so that $\mathcal{A}\mathbf{x}' \prec 0$.

The equations 26 and 27 indicate that we can always find some $\mathbf{x}'$ that satisfied $\mathcal{A}\mathbf{x}' \prec 0$. Furthermore, we can multiple $\mathbf{x}'$ with any $t > 0$ and we still have:

$$\mathcal{A}(t\mathbf{x}') \prec 0 \tag{28}$$

Additionally, there is:

$$\begin{aligned} \mathbf{c}^T t\mathbf{x}' &= \mathbf{c}^T\mathcal{A}^{-1}\mathcal{A}\, t\mathbf{x}' \\ &= (\mathcal{A}^{-T}\mathbf{c})^T\mathcal{A}\, t\mathbf{x}' \end{aligned} \tag{29}$$

Since $\mathcal{A}^{-T}\mathbf{c} \succ 0$ and $\mathcal{A}(t\mathbf{x}') \prec 0$, we have:

$$\mathbf{c}^T t\mathbf{x}' = (\mathcal{A}^{-T}\mathbf{c})^T\mathcal{A}\, t\mathbf{x}' < 0 \tag{30}$$

Which will goes to $-\infty$ when $t \to \infty$. Therefore, we have proved $\mathbf{c}^T\mathbf{x}$ is unbound below. In conclusion, we have:

$$p^* = \begin{cases} \mathbf{c}^T\mathcal{A}^{-1}\mathbf{b} & \mathcal{A}^{-T}\mathbf{c} \preceq 0 \\ -\infty & \textit{otherwise} \end{cases} \tag{31}$$

Q5

We first introduce a set of new variables that satisfied:

$$y_i = b_i - \mathbf{a}_i^T\mathbf{x} \tag{32}$$

Therefore, the original problem can be transformed to:

$$\begin{aligned} \min \quad & -\sum_{i=1}^{m} \log y_i \\ s.t. \quad & y_i - b_i + \mathbf{a}_i^T\mathbf{x} = 0 \quad i = 1, 2, \cdots, m \end{aligned} \tag{33}$$

The Lagrangian of problem 33 can be written as:

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{v}) = -\sum_{i=1}^{m} \log y_i + \sum_{i}^{m} v_i(y_i - b_i + \mathbf{a}_i \mathbf{x}) \tag{34}$$

with the domain: $y_i > 0$ and $\mathbf{x} \in \mathcal{R}^n$. The dual function:

$$g(\lambda, \mathbf{v}) = \inf_{\mathbf{x}, \mathbf{y}} \mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{v}) \tag{35}$$

Indeed, the Lagrangian can be further expressed as:

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{v}) &= \sum_{i=1}^{m} (v_i y_i - \log y_i) - \sum_{i=1}^{m} v_i(b_i - \mathbf{a}_i^T \mathbf{x}) \\ &= \sum_{i=1}^{m} (v_i y_i - \log y_i) - \mathbf{b}^T \mathbf{v} + \mathbf{v}^T \mathcal{A} \mathbf{x} \end{aligned} \tag{36}$$

Where $\mathcal{A} = \begin{bmatrix} -\mathbf{a}_1^T - \\ -\mathbf{a}_2^T - \\ \vdots \\ -\mathbf{a}_m^T - \end{bmatrix}$ and $\mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$. We can first inspect the part $-\mathbf{b}^T \mathbf{v} + \mathbf{v}^T \mathcal{A} \mathbf{x}$, which is basically an affine function with domain $\mathbf{x} \in \mathcal{R}^n$. This affine function is unbound below unless it satisfied:

$$\mathbf{v}^T \mathcal{A} = \mathbf{0}$$

Under this situation, $\inf_{\mathbf{x}}(-\mathbf{b}^T \mathbf{v} + \mathbf{v}^T \mathcal{A} \mathbf{x}) = -\mathbf{b}^T \mathbf{v}$. Furthermore, for each component $v_i y_i - \log y_i$: If $v_i \leq 0$, this function is obviously unbound below. When $v_i > 0$, it reaches the minimum when $y_i = \frac{1}{v_i}$. This minimum point can be obtained by noticing this is a convex function and setting it's first order derivative to zero. Hence, there is:

$$\begin{aligned} \inf_{\mathbf{y}} \sum_{i=1}^{m} (v_i y_i - \log y_i) &= \sum_{i=1}^{m} (1 - \log \frac{1}{v_i}) \\ &= m + \sum_{i=1}^{m} \log v_i \end{aligned} \tag{37}$$

As a result, we can conclude that:

$$g(\lambda, \mathbf{v}) = \begin{cases} m + \sum_{i=1}^{m} \log v_i - \mathbf{b}^T \mathbf{v} & \mathbf{v}^T \mathcal{A} = \mathbf{0} \textit{ and } \mathbf{v} \succ 0 \\ -\infty & \textit{otherwise} \end{cases} \tag{38}$$

Bases on this dual function, we can write the dual function as:

$$\begin{aligned} \max_{\lambda, \mathbf{v}} \quad & m + \sum_{i=1}^{m} \log v_i - \mathbf{b}^T \mathbf{v} \\ s.t. \quad & \mathbf{v} \succ 0 \\ & \mathbf{v}^T \mathcal{A} = \mathbf{0} \end{aligned} \tag{39}$$

Q6

From the last equation of KKT condition, there is:

$$\begin{aligned} \left(\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) \right)^T &= \nabla f_0^T(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i^T(\mathbf{x}^*) \\ &= \mathbf{0} \end{aligned} \tag{40}$$

By multiplying this equation with $\mathbf{x} - \mathbf{x}^*$, where $\mathbf{x}$ could be any point in the feasible set, we have:

$$\nabla f_0^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = -\sum_{i=1}^{m} \left(\lambda_i^* \nabla f_i^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \right) \tag{41}$$

Hence, we can start our prove by first consider the sign of single component: $\lambda_i^* \nabla f_i^T(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)$. Indeed, the KKT conditions:

$$\begin{cases} f_i(\mathbf{x}^*) \leq 0 \\ \lambda_i \geq 0 \\ \lambda_i f_i(\mathbf{x}^*) = 0 \end{cases} \tag{42}$$

Indicate that either λ_i^* or $f_i(\mathbf{x}^*)$ must equal to zero. For the situation that $\lambda_i^* = 0$, we have:

$$\lambda_i^* \nabla f_i^T(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) = 0 \tag{43}$$

For the situation $f_i(\mathbf{x}^*) = 0$, we first notice that f_i is convex function. Hence, the first-order condition always hold, which indicate for any $\mathbf{x}$ and $\mathbf{x}^*$ in feasible set, there is:

$$f_i(\mathbf{x}) \geq f_i(\mathbf{x}^*) + \nabla f_i^T(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) \tag{44}$$

Based on our assumption that $f_i(\mathbf{x}^*) = 0$ and the fact: for every feasible $\mathbf{x}$, $f_i(\mathbf{x}) \leq 0$, we can conduct:

$$0 \geq \nabla f_i^T(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) \tag{45}$$

Combine the inequality 43 and 45, we can conclude:

$$\sum_{i=1}^{m} \Big(\lambda_i^* \nabla f_i^T(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)\Big) \leq 0 \tag{46}$$

Hence:

$$\begin{aligned} \nabla f_0^T(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*) &= -\sum_{i=1}^{m} \Big(\lambda_i^* \nabla f_i^T(\mathbf{x}^*)(\mathbf{x}-\mathbf{x}^*)\Big) \\ &\geq 0 \end{aligned} \tag{47}$$