

Assignment 3

Songtuan Lin u6162630

April 5, 2019

Q1

We first define the set that contain all the nodes within graph:

$$\mathcal{N} = \{1, 2, \dots, n\} \quad (1)$$

(a)

A path is start at node 1 and end at node n if there is only one edge flow out of node 1 and one edge flow into node n in addition to no edge flow into node 1 and no edge flow out of edge n . Hence, according to this property, the equation can be written as:

$$\begin{cases} \sum_{i \in \mathcal{N}} x_{1,i} = 1 \\ \sum_{j \in \mathcal{N}} x_{j,n} = 1 \\ \sum_{j \in \mathcal{N}} x_{i,1} = 0 \\ \sum_{j \in \mathcal{N}} x_{n,j} = 0 \end{cases} \quad (2)$$

(b)

The path pass through any node at most once indicate the number of edge that flow out of any node is small or equal to one. Hence, the constrain can be written as:

$$\begin{cases} \sum_{j \in \mathcal{N}} x_{1,j} \leq 1 \\ \sum_{j \in \mathcal{N}} x_{2,j} \leq 1 \\ \vdots \\ \sum_{j \in \mathcal{N}} x_{n,j} \leq 1 \end{cases} \quad (3)$$

We can define function $f_i(\mathbf{x})$ as:

$$f_i(\mathbf{x}) = \sum_{j \in \mathcal{N}} x_{i,j} - 1 \quad (4)$$

Where $i = 1, 2, \dots, n$.

(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:

$$\begin{cases} \sum_{i \in \mathcal{N}} x_{i,2} = \sum_{j \in \mathcal{N}} x_{2,j} \\ \sum_{i \in \mathcal{N}} x_{i,3} = \sum_{j \in \mathcal{N}} x_{3,j} \\ \vdots \\ \sum_{i \in \mathcal{N}} x_{i,n-1} = \sum_{j \in \mathcal{N}} x_{n-1,j} \end{cases} \quad (5)$$

We can also define function $h_k(\mathbf{x})$ as:

$$h_k(\mathbf{x}) = \sum_{i \in \mathcal{N}} x_{i,k} - \sum_{j \in \mathcal{N}} x_{k,j} \quad (6)$$

Where $k = 2, 3, \dots, n-1$.

(d)

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

$$\sum_{i,j} c_{i,j} x_{i,j} \quad (7)$$

(e)

Following equation 2 to 7, the *Shortest Path Problem* can be formalized into optimization problem as:

$$\begin{aligned} \min_{x_{i,j}} \quad & \sum_{i,j} c_{i,j} x_{i,j} \\ \text{s.t.} \quad & f_i(\mathbf{x}) \leq 0 \quad i = 1, 2, \dots, n \\ & \sum_{i \in \mathcal{N}} x_{1,i} = 1 \\ & \sum_{j \in \mathcal{N}} x_{j,n} = 1 \\ & \sum_{j \in \mathcal{N}} x_{i,1} = 0 \\ & \sum_{j \in \mathcal{N}} x_{n,j} = 0 \\ & h_k(\mathbf{x}) = 0 \quad k = 2, 3, \dots, n-1 \\ & x_{i,j} \in \{0, 1\} \end{aligned} \quad (8)$$

(f)

This is a convex optimization problem as all the equality and inequality constrain are linear function.

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:

$$\nabla f_0(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \geq 0 \quad (9)$$

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

$$\nabla f_0(\mathbf{x}^*) = \mathcal{P} \mathbf{x}^* + \mathbf{q}^T = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} \quad (10)$$

By substituting equation 10 to 9, we can rewrite the condition that need to be verified as:

$$\begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix} (\mathbf{x} - \begin{bmatrix} 1 \\ 0.5 \\ -1 \end{bmatrix}) = -1(x_1 - 1) + 2(x_3 + 1) \geq 0 \quad (11)$$

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:

$$\begin{aligned} \min_{x_1, x_3} \quad & -x_1 + 2x_3 + 3 \\ \text{s.t.} \quad & -1 \leq x_1 \leq 1 \\ & -1 \leq x_3 \leq 1 \end{aligned} \quad (12)$$

The optimal point of this problem can be obtained by simple inspection, which is $(1, -1)$. Hence, the minimum value of $-x_1 + 2x_3 + 3$ 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:

$$\begin{aligned} \mathcal{A}\mathbf{x} &= \mathcal{A}(\mathbf{y} + \mathbf{x}_0) \\ &= \mathcal{A}\mathbf{y} + \mathcal{A}\mathbf{x}_0 \\ &= \mathbf{0} + \mathbf{b} \end{aligned} \quad (13)$$

Based on this decomposition, there is:

$$\nabla f_0(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0) = \mathbf{c}^T \mathbf{y} \quad (14)$$

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:

$$\begin{cases} \mathbf{c}^T \mathbf{y} \geq 0 \\ -\mathbf{c}^T \mathbf{y} \geq 0 \end{cases} \quad (15)$$

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:

$$\mathbf{c} = \mathcal{A}^T \lambda \quad (\lambda \in \mathcal{R}^n) \quad (16)$$

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:

$$\begin{aligned} \mathbf{c}^T \mathbf{x}_0 &= \lambda^T \mathcal{A}\mathbf{x}_0 \\ &= \lambda^T \mathbf{b} \end{aligned} \quad (17)$$

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:

$$p^* = \begin{cases} -\infty & \mathbf{c} \notin \mathcal{R}(\mathcal{A}) \\ \lambda^T \mathbf{b} & \mathbf{c} \in \mathcal{R}(\mathcal{A}) \\ \infty & \mathbf{x} \text{ is infeasible} \end{cases} \quad (18)$$

(b)

The objective function can be written as:

$$f_0(\mathbf{x}) = \sum_{i=1}^n c_i x_i \quad (19)$$

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

$$x_i^* = \begin{cases} u & c_i \leq 0 \\ l & \text{otherwise} \end{cases} \quad (20)$$

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

$$\nabla f_0(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = \sum_i c_i (x_i - x_i^*) \quad (21)$$

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:

$$c_1^* + c_2^* + \dots + c_\alpha^* \quad (22)$$

Where c_i^* is the i th 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 $0 \preceq \mathbf{x} \preceq \mathbf{1}$ 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 i th 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 $0 \preceq \mathbf{x} \preceq \mathbf{1}$ 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 22, the optimal value under this situation is:

$$\sum_{i=1}^{\lfloor \alpha \rfloor} c_i^* + q c_{\lfloor \alpha \rfloor + 1}^* \quad (23)$$

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:

$$\begin{aligned} \nabla f_0(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) &= \mathbf{c}^T (\mathbf{x} - \mathbf{x}^*) \\ &= \mathbf{c}^T \mathcal{A}^{-1} \mathcal{A} (\mathbf{x} - \mathbf{x}^*) \\ &= (\mathcal{A}^{-T} \mathbf{c})^T \mathcal{A} (\mathbf{x} - \mathbf{x}^*) \end{aligned} \quad (24)$$

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

$$\mathcal{A} (\mathbf{x} - \mathbf{x}^*) = \mathcal{A} \mathbf{x} - \mathcal{A} \mathbf{x}^* \preceq 0 \quad (25)$$

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

$$(\mathcal{A}^{-T}\mathbf{c})^T \mathcal{A}(\mathbf{x} - \mathbf{x}^*) \geq 0 \quad (26)$$

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:

$$\mathbf{x}^* = \mathcal{A}^{-1}\mathbf{b} \quad (27)$$

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

$$\begin{aligned} p^* &= \mathbf{c}^T \mathbf{x}^* \\ &= \mathbf{c}^T \mathcal{A}^{-1}\mathbf{b} \end{aligned} \quad (28)$$

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:

$$\mathcal{A}\mathbf{x}' \prec \mathbf{b} \prec 0 \quad (29)$$

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

$$\mathcal{A}\mathbf{x}' \prec 0 \prec \mathbf{b} \quad (30)$$

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 29 and 30 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:

$$\mathcal{A}(t\mathbf{x}') \prec 0 \quad (31)$$

Additionally, there is:

$$\begin{aligned} \mathbf{c}^T t\mathbf{x}' &= \mathbf{c}^T \mathcal{A}^{-1} \mathcal{A} t\mathbf{x}' \\ &= (\mathcal{A}^{-T}\mathbf{c})^T \mathcal{A} t\mathbf{x}' \end{aligned} \quad (32)$$

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

$$\mathbf{c}^T t\mathbf{x}' = (\mathcal{A}^{-T}\mathbf{c})^T \mathcal{A} t\mathbf{x}' < 0 \quad (33)$$

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

$$p^* = \begin{cases} \mathbf{c}^T \mathcal{A}^{-1}\mathbf{b} & \mathcal{A}^{-T}\mathbf{c} \preceq 0 \\ -\infty & \text{otherwise} \end{cases} \quad (34)$$

Q5

We first introduce a set of new variables that satisfied:

$$y_i = b_i - \mathbf{a}_i^T \mathbf{x} \quad (35)$$

Therefore, the original problem can be transformed to:

$$\begin{aligned} \min \quad & -\sum_{i=1}^m \log y_i \\ \text{s.t.} \quad & y_i - b_i + \mathbf{a}_i^T \mathbf{x} = 0 \quad i = 1, 2, \dots, m \end{aligned} \quad (36)$$

The Lagrangian of problem 36 can be written as:

$$\mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{v}) = -\sum_{i=1}^m \log y_i + \sum_{i=1}^m v_i (y_i - b_i + \mathbf{a}_i^T \mathbf{x}) \quad (37)$$

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

$$g(\lambda, \mathbf{v}) = \inf_{\mathbf{x}, \mathbf{y}} \mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{v}) \quad (38)$$

Indeed, the Lagrangian can be further expressed as:

$$\begin{aligned}\mathcal{L}(\mathbf{x}, \mathbf{y}, \mathbf{v}) &= \sum_{i=1}^m (v_i y_i - \log y_i) - \sum_{i=1}^m v_i (b_i - \mathbf{a}_i^T \mathbf{x}) \\ &= \sum_{i=1}^m (v_i y_i - \log y_i) - \mathbf{b}^T \mathbf{v} + \mathbf{v}^T \mathcal{A} \mathbf{x}\end{aligned}\tag{39}$$

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:

$$\begin{aligned}\inf_{\mathbf{y}} \sum_{i=1}^m (v_i y_i - \log y_i) &= \sum_{i=1}^m (1 - \log \frac{1}{v_i}) \\ &= m + \sum_{i=1}^m \log v_i\end{aligned}\tag{40}$$

As a result, we can conclude that:

$$g(\lambda, \mathbf{v}) = \begin{cases} m + \sum_{i=1}^m \log v_i - \mathbf{b}^T \mathbf{v} & \mathbf{v}^T \mathcal{A} = \mathbf{0} \text{ and } \mathbf{v} \succ 0 \\ -\infty & \text{otherwise} \end{cases}\tag{41}$$

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

$$\begin{aligned}\max_{\lambda, \mathbf{v}} \quad & m + \sum_{i=1}^m \log v_i - \mathbf{b}^T \mathbf{v} \\ \text{s.t.} \quad & \mathbf{v} \succ 0 \\ & \mathbf{v}^T \mathcal{A} = \mathbf{0}\end{aligned}\tag{42}$$

Q6

From the last equation of *KKT* condition, there is:

$$\begin{aligned}\left(\nabla f_0(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(\mathbf{x}^*) \right)^T &= \nabla f_0^T(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i^T(\mathbf{x}^*) \\ &= \mathbf{0}\end{aligned}\tag{43}$$

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

$$\nabla f_0^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = - \sum_{i=1}^m \left(\lambda_i^* \nabla f_i^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \right)\tag{44}$$

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:

$$\begin{cases} f_i(\mathbf{x}^*) \leq 0 \\ \lambda_i \geq 0 \\ \lambda_i f_i(\mathbf{x}^*) = 0 \end{cases}\tag{45}$$

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

$$\lambda_i^* \nabla f_i^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) = 0\tag{46}$$

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:

$$f_i(\mathbf{x}) \geq f_i(\mathbf{x}^*) + \nabla f_i^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \quad (47)$$

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:

$$0 \geq \nabla f_i^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \quad (48)$$

Combine the inequality 46 and 48, we can conclude:

$$\sum_{i=1}^m \left(\lambda_i^* \nabla f_i^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \right) \leq 0 \quad (49)$$

Hence:

$$\begin{aligned} \nabla f_0^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) &= - \sum_{i=1}^m \left(\lambda_i^* \nabla f_i^T(\mathbf{x}^*)(\mathbf{x} - \mathbf{x}^*) \right) \\ &\geq 0 \end{aligned} \quad (50)$$