Network flow problems

These are LP problems having a special structure

We will see that the simplex method is simplified in these problems.

Why study them separately?

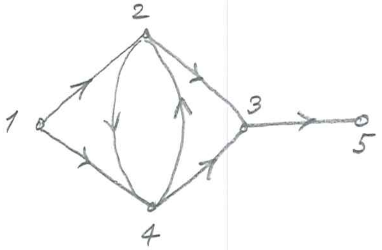
(1) They constitute important examples of LP problems that arise frequently.

(2) They have an associated rich theory, which provides insight.

What is a network?

pair $(\underset{\text{nodes (m)}}{N}, \underset{\text{directed edges between pairs of nodes (n)}}{E})$

Example.



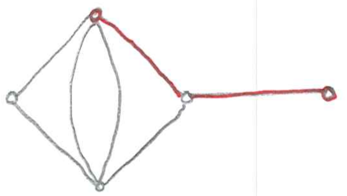
Nodes = $\{1, 2, 3, 4, 5\}$

Edges = $\{(1,2), (1,4), (2,4), (4,2), (2,3), (4,3), (3,5)\}$

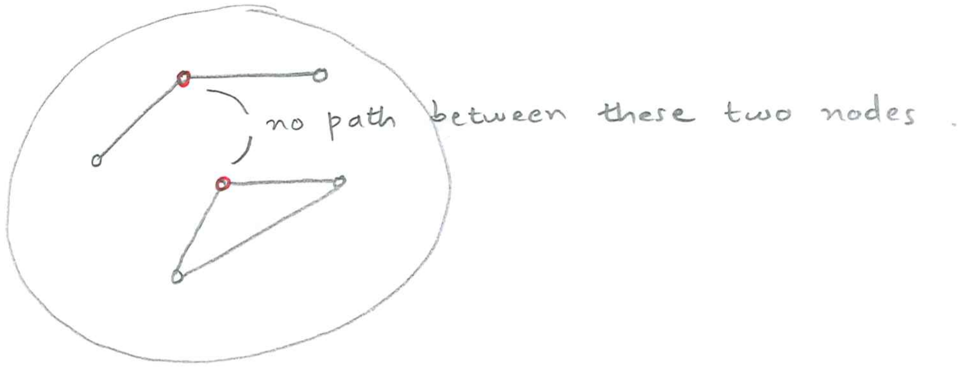
Here: directed edge from i to j is denoted by (i,j).

E.g. $(2,4)$, $(4,2)$ are two different directed edges between nodes 2 and 4.

Network is connected, i.e., there exists a path of edges between any two nodes.



Connected network.



Not connected network.

c_{ij} = cost of unit flow in edge $(i,j) \in E$. (Given)

x_{ij} = flow in edge $(i,j) \in E$. (Variable)

What is the network flow problem?

Roughly, it is the problem of minimizing the total cost of flow, i.e., to find out which flow minimizes the total cost.

The total cost is given by $\sum_{(i,j) \in E} c_{ij} x_{ij}$.

Note that it is a linear function of the variables (x_{ij}'s).

But there are also (linear) constraints given by the fact that nothing is stored at the nodes (flow balance).

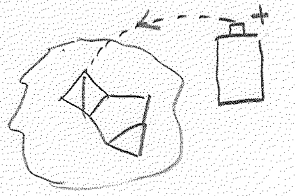
So we are going to get a LP problem.

But first let us see what the constraints are.

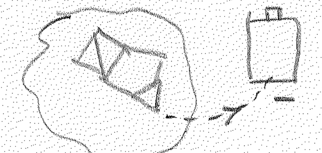
First, some terminology:

A node is called a

(1) <u>source node</u> if there is flow added at this node from the outside of the network. (Imagine the place where the +ve terminal of a battery is connected to an electrical network)



(2) <u>sink node</u> if there is flow taken out at this node and sent outside the network. (Imagine the place where the -ve terminal of a battery is connected to an electrical network.)



(3) <u>intermediate node</u> if it is neither a source node nor a sink node.

At each node, there is <u>flow balance</u>

i.e, inflow = outflow (at each node)

$$\begin{tikzcd} & \text{inflow} \arrow[dl, phantom, "//"] \arrow[r, phantom, "="] & \text{outflow} \arrow[dr, phantom, "\backslash\backslash"] & \\ \begin{array}{l}\text{sum of flows into node}\\ \quad\text{from directed edges}\\ \quad\text{into the node}\\ +\\ \text{flow from outside (if it is}\\ \quad\text{a source node)}\end{array} & & & \begin{array}{l}\text{sum of flows going out of the node}\\ \quad\text{via directed edges starting from that node}\\ +\\ \text{flow to the outside (if it is a sink node)}\end{array} \end{tikzcd}$$

Let us define

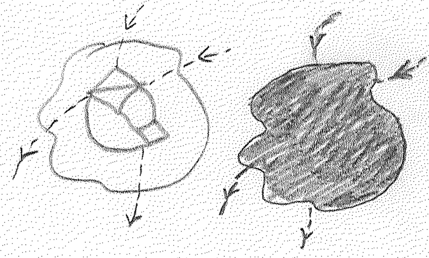
b_i := amount of flow supplied <u>to</u> the network <u>from</u> the outside at node i.

For source nodes $b_i > 0$;

for sink nodes $b_i < 0$;

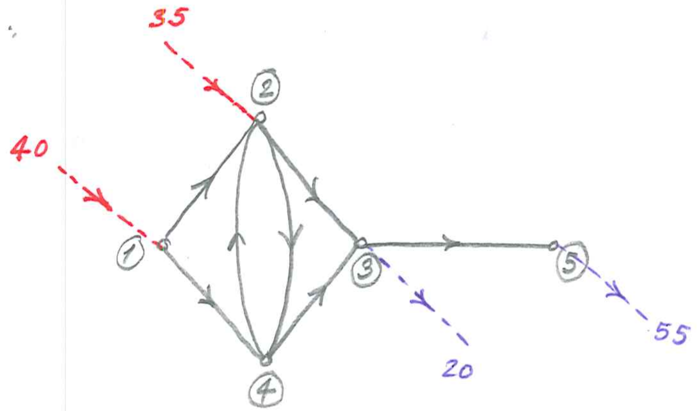
for intermediate nodes $b_i = 0$.

Assumption : $\boxed{\sum_{i=1}^{m} b_i = 0}$. (Network stores nothing)



From now on, we will illustrate the theory by focussing on a concrete example.

Consider the network:



$m = 5$

$n = 7$

Source nodes : ①, ②

Sink nodes : ③, ⑤

Intermediate nodes : ④ .

$$\tilde{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{bmatrix} = \begin{bmatrix} 40 \\ 35 \\ -20 \\ 0 \\ -55 \end{bmatrix} .$$ Note that $\sum_{i \in N} b_i = 40 + 35 + (-20) + 0 + (-55) = 0$.

Edges

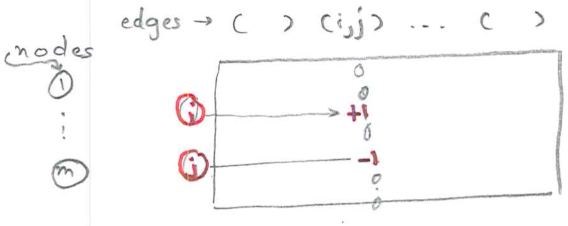
$$E = \{ (1,2), (1,4), (2,4), (4,2), (2,3), (4,3), (3,5) \} .$$

Order of edges	1st edge	2nd edge	3rd edge	4th edge	5th edge	6th edge	7th edge

Having fixed the order of the edges, we can now write down the <u>incidence matrix</u> $\tilde{A}$ of the network. (Using $\tilde{A}$, we can write the constraints succinctly.)

Writing the incidence matrix of a network:

$\tilde{A} =$



$$a_{ik} = \begin{cases} 1 & \text{if the } k^{th} \text{ edge begins at node } \textcircled{i} \\ -1 & \text{if the } k^{th} \text{ edge ends at node } \textcircled{i} \\ 0 & \text{for other } i\text{'s} \end{cases}$$

For our network:

$$\tilde{A} = \begin{array}{c|ccccccc} \text{node} \backslash \text{edge} & (1,2) & (1,4) & (2,4) & (4,2) & (2,3) & (4,3) & (3,5) \\ \hline \textcircled{1} & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ \textcircled{2} & -1 & 0 & 1 & -1 & 1 & 0 & 0 \\ \textcircled{3} & 0 & 0 & 0 & 0 & -1 & -1 & 1 \\ \textcircled{4} & 0 & -1 & -1 & +1 & 0 & 1 & 0 \\ \textcircled{5} & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{array}$$

Q. What is the use of the incidence matrix?

A. Flow balance is now just $\tilde{A} x = \tilde{b}$. Let us verify this in our example.

$$\tilde{b} = \begin{array}{c|c} & \text{node} \\ \left[\begin{array}{c} 40 \\ 30 \\ -20 \\ 0 \\ -55 \end{array}\right] & \begin{array}{c} ① \\ ② \\ ③ \\ ④ \\ ⑤ \end{array} \end{array} \qquad x = \begin{bmatrix} x_{12} \\ x_{14} \\ x_{24} \\ x_{42} \\ x_{23} \\ x_{43} \\ x_{35} \end{bmatrix} \downarrow \text{same ordering as the chosen ordering of edges}$$

$$\tilde{A}x = \begin{array}{c|ccccccc} \text{node} & \text{edge }(1,2) & (1,4) & (2,4) & (4,2) & (2,3) & (4,3) & (3,5) \\ ① & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ ② & -1 & 0 & 1 & -1 & 1 & 0 & 0 \\ ③ & 0 & 0 & 0 & 0 & -1 & -1 & 1 \\ ④ & 0 & -1 & -1 & 1 & 0 & 1 & 0 \\ ⑤ & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{array} \begin{bmatrix} x_{12} \\ x_{14} \\ x_{24} \\ x_{42} \\ x_{23} \\ x_{43} \\ x_{35} \end{bmatrix} = \begin{bmatrix} \overbrace{x_{12}+x_{14}}^{\text{flow going out of node ① via edges touching ①}} \\ -x_{12}+x_{24}-x_{42}+x_{23} \\ -x_{23}-x_{43}+x_{35} \\ -x_{14}-x_{24}+x_{42}+x_{43} \\ -x_{35} \end{bmatrix} = \begin{bmatrix} 40 \\ 35 \\ -20 \\ 0 \\ -55 \end{bmatrix}$$

So the flow balances at all nodes are given by $\tilde{A}x = b$.

Now we can write the network flow problem:

$$(\text{NFP}): \begin{cases} \text{minimize} & c^T x \\ \text{s.t.} & \tilde{A}x = \tilde{b} \\ & x \geq 0 \end{cases}$$

In our example, let us take $c = \begin{bmatrix} c_{12} \\ c_{14} \\ c_{24} \\ c_{42} \\ c_{23} \\ c_{43} \\ c_{35} \end{bmatrix} = \begin{bmatrix} 2 \\ 5 \\ 2 \\ 2 \\ 1 \\ 1 \\ 2 \end{bmatrix}$.

6.9

The problem is in standard form, but rank $\tilde{A} \neq m$, since the sum of all rows of $\tilde{A}$ is 0.

Delete the last row of $\tilde{A}$ to obtain A.

Delete the last entry of $\tilde{b}$ to obtain b.

$$A = \begin{array}{c|ccccccc} & (1,2) & (1,4) & (2,4) & (4,2) & (2,3) & (4,3) & (3,5) \\ \hline ① & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ ② & -1 & 0 & 1 & -1 & 1 & 0 & 0 \\ ③ & 0 & 0 & 0 & 0 & -1 & -1 & 1 \\ ④ & 0 & -1 & -1 & 1 & 0 & 1 & 0 \end{array}$$

$$b = \begin{bmatrix} 40 \\ 35 \\ -20 \\ 0 \end{bmatrix}$$

$$(\text{NFP}): \left\{ \begin{array}{ll} \text{minimize} & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 . \end{array} \right.$$

$$\left\{ x : \underset{\begin{array}{c}\boxed{A}\\ \hline *\end{array}}{\tilde{A}} x = \underset{\begin{array}{c}\boxed{b}\\ *\end{array}}{\tilde{b}} , \ x \geq 0 \right\} \overset{?}{=} \left\{ x : Ax = b , \ x \geq 0 \right\} .$$

$\subseteq$ (obvious)

$\supseteq$ ($\because$ sum of rows of A is 0 and sum of entries of b is zero)

A now has independent rows:

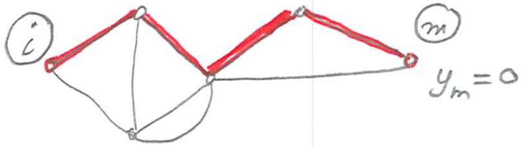
Suppose $y^T A = 0$. Then we will show $y = 0$.

$$0 = y^T A = \begin{bmatrix} y^T & 0 \end{bmatrix} \left.\begin{bmatrix} \boxed{A} \\ * \end{bmatrix}\right\}\tilde{A} = \begin{bmatrix} y_1 \cdots y_{m-1} & \overset{0}{\overset{\|}{y_m}} \end{bmatrix} \begin{bmatrix} & \overset{(i,j)}{} & \\ & +1 & \leftarrow \text{row } ⓘ \\ & -1 & \leftarrow \text{row } ⓙ \end{bmatrix}$$

$$= \begin{bmatrix} \cdots & \overset{(i,j)}{\boxed{y_i - y_j}} & \cdots \end{bmatrix} \leftarrow \text{row with entries } y_i - y_j \text{ for each edge } (i,j)$$

So : $y_i - y_j = 0 \quad \forall$ edge (i,j).

But network is connected. $\Rightarrow$ There is a path of edges between node ⓜ and any other node ⓘ.



$$y_m = 0 \Rightarrow y_i = 0 \quad \forall i.$$

In our example:

$$0 = y^T A = \begin{bmatrix} y_1 & y_2 & y_3 & y_4 & \overset{0}{\overset{\|}{y_5}} \end{bmatrix} \begin{array}{c} \begin{array}{ccccccc} (1,2) & (1,4) & (2,4) & (4,2) & (2,3) & (4,3) & (3,5) \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 \\ 0 & -1 & -1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \end{array}$$

$$= \begin{bmatrix} y_1 - y_2 & y_1 - y_4 & y_2 - y_4 & y_4 - y_2 & y_2 - y_3 & y_4 - y_3 & y_3 - y_5 \end{bmatrix}.$$

$$\begin{array}{ccccccc} & \Downarrow & & & \Downarrow & / & \Downarrow \;\; \overset{\|}{0} \\ & y_4 = 0 & & & y_2 = 0 & y_4 = 0 \Leftarrow & y_3 = 0 \\ & & & & \longleftarrow & & \end{array}$$

So the rows of A are linearly independent. □

We want to use the simplex method for solving (NFP).

How do we find an initial b.f.s? There is a way:

Basic solutions ⟷ spanning trees.

What is a spanning tree?

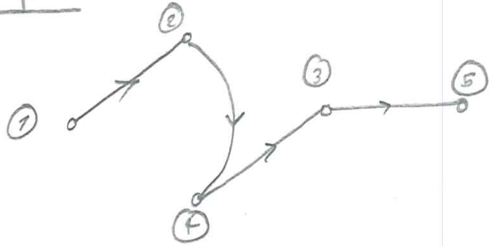
A subset T of edges of a network is a spanning tree if

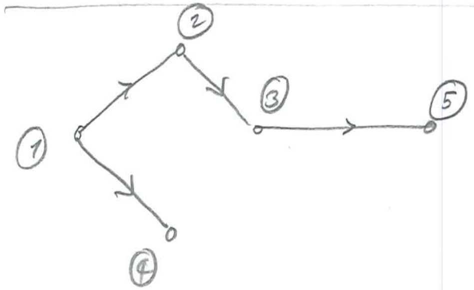
(1) every node touches at least one edge of T

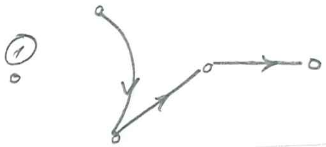
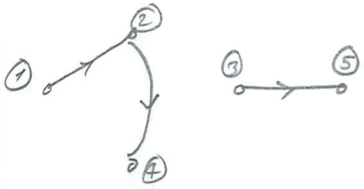
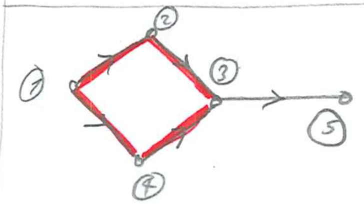
(2) T is connected

(3) there is no loop in T (neglecting directions of edges).

Examples





Non-examples	What goes wrong?
	Node ① doesn't touch any tree edge.
	Not connected
	∃ loop (shown in red)

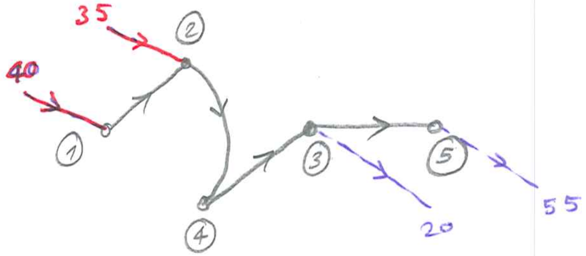
What is the connection between spanning trees and basic solutions?

Theorem. Consider $A \in \mathbb{R}^{(m-1)\times n}$ in the (NFP). Then

$$\boxed{\text{a set of } (m-1) \text{ columns of } A \text{ are linearly independent}} \iff \boxed{\text{the corresponding set of edges form a spanning tree for the network}}$$

Given a spanning tree, how do we find a basic solution?

Note that $x_{ij} = 0$ if (i,j) is not a tree edge. So just use flow balance in the tree!



By flow balance at node ① : $x_{12} = 40$

② : $x_{12} + 35 = x_{24} \Rightarrow x_{24} = 40+35 = 75$.

④ : $x_{24} = x_{43} \Rightarrow x_{43} = 75$.

③ : $x_{43} = 20 + x_{35} \Rightarrow x_{35} = 75 - 20 = 55$.

⑤ : $x_{35} = 55$

Basic part of basic solution

$x_{12} = 40 \geq 0$

$x_{24} = 75 \geq 0$

$x_{43} = 75 \geq 0$

$x_{35} = 55 \geq 0$

so the basic solution corresponding to this tree is feasible, i.e., it is a <u>b.f.s.</u>

Nonbasic part

$x_{14} = 0$

$x_{42} = 0$

$x_{23} = 0$.

6.13

Is this b.f.s. optimal?

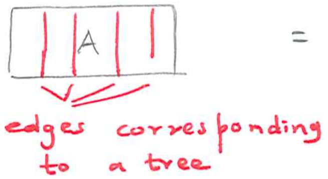
Are reduced costs for the nonbasic variables nonnegative,

i.e., is $r_\nu \geqslant 0$? $c^T x = c^T \bar{x} + r_\nu^T \underset{\geqslant 0}{x_\nu}$.

$r_\nu = c_\nu - A_\nu^T \textcircled{y}$.
└ need to calculate y first

Simplex multipliers vector y.

$A_\beta^T y = c_\beta$ i.e., $y^T A_\beta = c_\beta^T$.

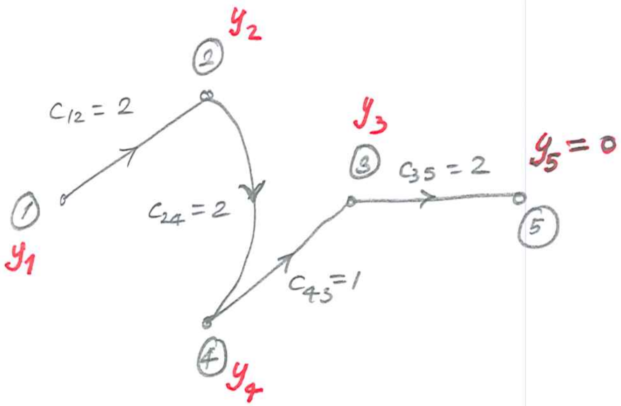
y^T 

$= [y^T \ \ 0] \underbrace{\begin{bmatrix} A \\ * \end{bmatrix}}_{\tilde{A}} = [y_1 \dots y_{m-1} \ \underset{\overset{\shortparallel}{0}}{y_m}] \tilde{A} = [\ \cdot \ \boxed{\ }-\boxed{\ } \ \overset{(i,j)}{\boxed{y_i - y_j}} \ \boxed{\ }-\boxed{\ }]$

$y^T A_\beta =$ row with $y_i - y_j$ for all edges $(i,j) \in T$ (tree)

$y^T A_\beta = c_\beta^T$ becomes

$$\boxed{\begin{array}{l} y_i - y_j = c_{ij} \quad \text{for} \quad (i,j) \in T \\ \quad y_m = 0. \end{array}}$$



$y_3 - y_5 = c_{35}$ $y_3 - 0 = 2$ $y_3 = 2$

$y_4 - y_3 = c_{43}$ $y_4 - 2 = 1$ $y_4 = 3$

$y_2 - y_4 = c_{24}$ $y_2 - 3 = 2$ $y_2 = 5$

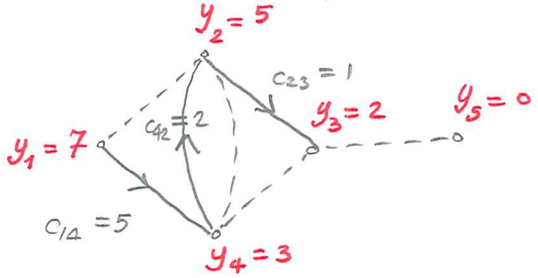
$y_1 - y_2 = c_{12}$ $y_1 - 5 = 2$ $y_1 = 7$.

Reduced costs of the nonbasic variables.

$r_\nu = c_\nu - A_\nu^T y$ $\quad$ $r_\nu^T = c_\nu^T - y^T A_\nu$

Recall : $y^T A_\nu$ = row with entries $y_i - y_j$ for (i,j)s not in T i.e., $(i,j) \in \underbrace{E \setminus T}_{\text{non tree edges.}}$

$$\boxed{\begin{array}{l} r_{ij} = c_{ij} - (y_i - y_j) \quad \text{for} \quad (i,j) \in E \setminus T. \\ y_m = 0 \end{array}}$$



$r_{14} = c_{14} - (y_1 - y_4) = 5 - (7 - 3) = 1$

$r_{42} = c_{42} - (y_4 - y_2) = 2 - (3 - 5) = 4$

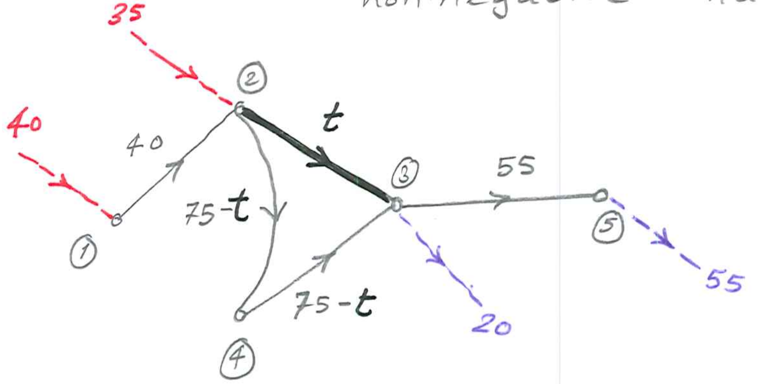
$r_{23} = c_{23} - (y_2 - y_3) = 1 - (5 - 2) = -2 < 0$

So it is not the case that $r_\nu \geqslant 0$. So we cannot conclude that the current b.f.s. is optimal!

(Note that the cost $\sum_{(i,j)\in E} \underbrace{c_{ij}}_{\geqslant 0} \underbrace{x_{ij}}_{\geqslant 0} \geqslant 0$ is bounded below (by zero), and so we cannot have a ray in the feasible set ~~se~~ along which the cost goes to $-\infty$.)

How do we find a new, better b.f.s. ?

Set $x_{23} = t$. Let t increase from 0, and let it be the largest possible nonnegative number while keeping all flows $\geqslant 0$.



$x_{12} = 40$

$x_{24} + t = 40 + 35 \Rightarrow x_{24} = 75 - t$.

$x_{43} = x_{24} = 75 - t$.

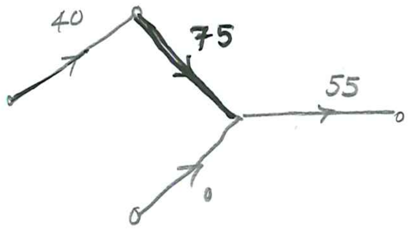
$\cancel{t} + 75 - \cancel{t} = 20 + x_{35} \Rightarrow x_{35} = 55$.

Q. What is the largest t can be while keeping all flows $\geqslant 0$?

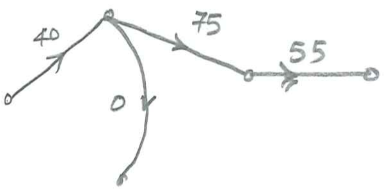
A. 75.

So $t_{max} = 75$. (Note in the simplex method for solving general LP problems in the standard form, we had a formula for this, namely $t_{max} = \min\left\{ \frac{\bar{b}_k}{\bar{a}_{v_q, k}} : \bar{a}_{v_q,k} > 0 \right\}$, but now we have replaced it by the above.)

So the new b.f.s. is:



(We could also take:

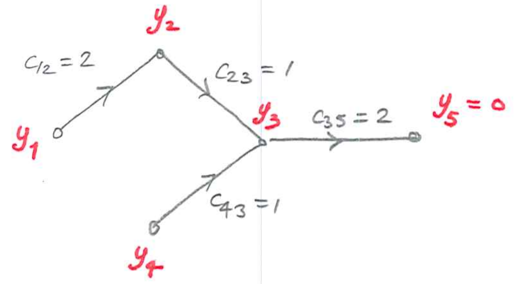


)

Notice that the edges form a tree. One iteration completed.

Is the new b.f.s. optimal?

Simplex multipliers. $y_i - y_j = c_{ij}$ for $(i,j) \in T$; $y_m = 0$



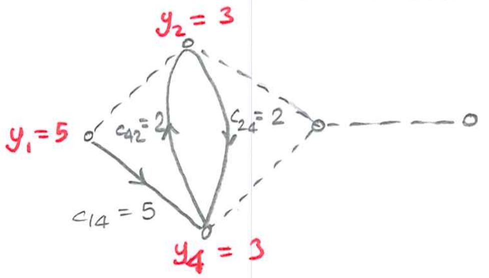
$y_3 - y_5 = c_{35}$ $\quad$ $y_3 - 0 = 2$ $\quad$ $y_3 = 2$

$y_4 - y_3 = c_{43}$ $\quad$ $y_4 - 2 = 1$ $\quad$ $y_4 = 3$

$y_2 - y_3 = c_{23}$ $\quad$ $y_2 - 2 = 1$ $\quad$ $y_2 = 3$

$y_1 - y_2 = c_{12}$ $\quad$ $y_1 - 3 = 2$ $\quad$ $y_1 = 5$.

Reduced costs of the non basic variables. $r_{ij} = c_{ij} - (y_i - y_j)$ for $(i,j) \in E \setminus T$.

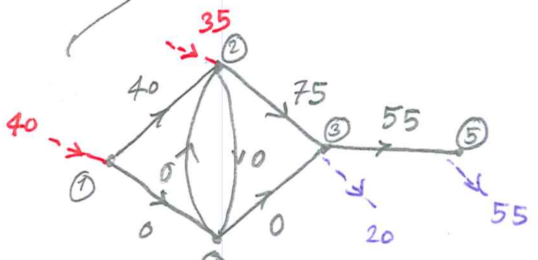


$r_{14} = c_{14} - (y_1 - y_4) = 5 - (5-3) = 5 - 2 = 3 \geq 0$

$r_{24} = c_{24} - (y_2 - y_4) = 2 - (3-3) = 2 \geq 0$

$r_{42} = c_{42} - (y_4 - y_2) = 2 - (3-3) = 2 \geq 0$.

So the current b.f.s. is optimal



The minimum cost is: $\sum_{(i,j) \in E} c_{ij} x_{ij}$

$= c_{12} x_{12} + c_{14} x_{14} + c_{24} x_{24} + c_{42} \cdot x_{42} + c_{23} \cdot x_{23} + c_{43} x_{43} + c_{35} x_{35}$

$= 2 \cdot 40 + 5 \cdot 0 + 2 \cdot 0 + 2 \cdot 0 + 1 \cdot 75 + 1 \cdot 0 + 2 \cdot 55$

$= 80 + 75 + 110 = 265$.