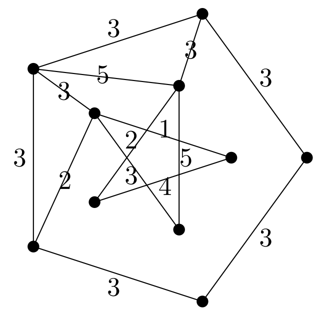
Bioinformatics

Discrete Mathematics and Optimization. Partial exam: Fall 2019

---

1. (Enumeration (25 points)) A DNA sequence is a word on the letters $\{A, T, G, C\}$. Let a_n be the number of DNA sequences of length n such that the letter 'T' is not followed by the letter 'G'.
   (a) Write a recurrence relation for a_n.
   (b) Solve the recurrence relation and give an explicit formula for a_n. What is the number of these sequences of length 10?

2. (Graphs (25 points))
   (a) What is a spanning tree? Given a graph G, is there a unique spanning tree? Explain your answer.
   (b) Run the Kruskal algorithm on the weighted graph G given below. Write down the individual steps and the output.
   (c) Will the output of the Kruskal algorithm be unique when it is run on graphs with distinct positive edge weights? Explain why.



3. (Linear programing (25 points)) We must run a program P and we have two available machines M_1 and M_2. Each execution of P spends $4s$ on M_1 and $1s$ on M_2. The cost per execution on M_1 is 1 cents and in M_2 is 4 cent. The energy consumption per execution on M_1 is $3mW$ and on M_2 it is $1mW$. We must run at least 10 executions in total but we can not exceed $37s$ of running time nor 28 cents of cost. We want to minimize the energy consumption.
   (a) Write a Linear Program to solve the problem.
   (b) Draw the feasible region of the problem and identify a solution graphically.
   (c) Write the Linear Program in equational form and run the simplex method on it.

Solutions

Enumeration.

1. For each sequence of length $n-1$ which ends in a letter different from T, one obtains four sequences of length n with no T followed by G. From all the sequences of length $n-1$ which end in T one can add only three symbols at the end. There are a_{n-2} sequences of length $n-1$ which end in T. Hence

$$a_n = 4(a_{n-1} - a_{n-2}) + 3a_{n-2} = 4a_{n-1} - a_{n-2}, n \geq 3.$$

   We clearly have $a_1 = 4$ and $a_2 = 15$ (from the 16 sequences of length two only the sequence TG is forbidden).

2. The characteristic polynomial of the recurrence is

$$x^2 - 4x + 1,$$

   which has roots $x_1 = 2 + \sqrt{3}$ and $x_2 = 2 - \sqrt{3}$. The general solution is

$$a_n = \alpha(2 + \sqrt{3})^n + \beta(2 - \sqrt{3})^n.$$

   In order to determine the values of α and β we use the initial values:

$$\begin{aligned} 4 &= \alpha(2 + \sqrt{3}) + \beta(2 - \sqrt{3}) \\ 15 &= \alpha(2 + \sqrt{3})^2 + \beta(2 - \sqrt{3})^2. \end{aligned}$$

   By multiplying the first equation by $(2 - \sqrt{3})$ and substracting it from the second one we obtain $\alpha = \frac{7+4\sqrt{3}}{2\sqrt{3}(2+\sqrt{3})}$ and $\beta = \frac{4-\alpha(2+\sqrt{3})}{2-\sqrt{3}} = -\frac{7-4\sqrt{3}}{2\sqrt{3}(2-\sqrt{3})}$. Therefore

$$a_n = \frac{12 + 7\sqrt{3}}{6}(2 + \sqrt{3})^{n-1} + \frac{12 - 7\sqrt{3}}{6}(2 - \sqrt{3})^{n-1}.$$

   In particular

$$a_{10} = 564719.$$

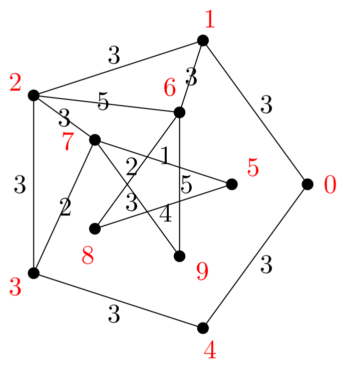
Graph Theory.

1. A spanning tree is a tree T which is a subgraph of G and $V(T) = V(G)$.

   If G is not connected then clearly G has no spanning trees (a tree is connected).

   A graph is connected if and only if it has a spanning tree. If G is not connected there are no spanning trees in G. If G is connected and it is a tree itself then it has an only spanning tree. Otherwise, it has a cycle and, from any given spanning tree $T \subset G$ one can build a different spanning tree by adding one edge from $E(G) \setminus E(T)$ (which will create a cycle) and removing one edge of T in this cycle: what remains is still connected and has $n-1$ edges, so it is a tree. Therefore G has more than one spanning tree.

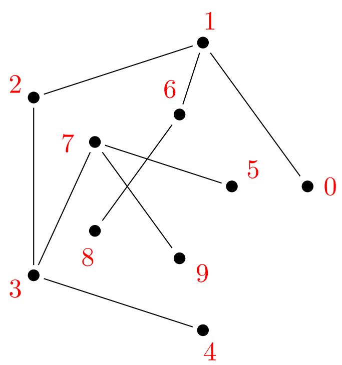
2. Label the vertices as in the Figure,



and sort the edges in nondecreasing order (the ordering in this case is not unique). The output of the algorithm is described in the following table, where at each step we add the next edge in the ordering as long as it produces an acyclic graph:

i	e_i	weight	$E(X_i)$
1	{5,7}	1	{5,7}
2	{6,8}	2	{5,7}, {6,8}
3	{3,7}	2	{5,7}, {6,8},{3,7}
4	{0,1}	3	{5,7}, {6,8},{3,7},{0,1}
5	{1,2}	3	{5,7}, {6,8},{3,7},{0,1},{1,2}
6	{2,3}	3	{5,7}, {6,8},{3,7},{0,1},{1,2},{2,3}
7	{3,4}	3	{5,7}, {6,8},{3,7},{0,1},{1,2},{2,3},{3,4}
8	{4,0}	3	{5,7}, {6,8},{3,7},{0,1},{1,2},{2,3},{3,4}
9	{2,7}	3	{5,7}, {6,8},{3,7},{0,1},{1,2},{2,3},{3,4}
10	{7,9}	3	{5,7}, {6,8},{3,7},{0,1},{1,2},{2,3},{3,4},{7,9}
11	{1,6}	3	{5,7}, {6,8},{3,7},{0,1},{1,2},{2,3},{3,4},{7,9} {1,6}
12	{5,8}	4	{5,7}, {6,8},{3,7},{0,1},{1,2},{2,3},{3,4},{7,9} {1,6}
13	{2,6}	5	{5,7}, {6,8},{3,7},{0,1},{1,2},{2,3},{3,4},{7,9} {1,6}
14	{3,9}	5	{5,7}, {6,8},{3,7},{0,1},{1,2},{2,3},{3,4},{7,9} {1,6}

The output is the spanning tree containing the edges listed at the last step:



with total weight 23.

3. If the edges have pairwise distinct weights then the minimum spanning tree is unique. Indeed, suppose that T_1, T_2 are two distinct trees with miniimum weight. And suppose that T_1 contains the edge e with smallest weight in $E(T_1) \cup E(T_2)$. Add this edge to T_2 and remove one edge of T_2 forming a cycle distinct from e. We obtain a spanning tree with smaller weight than T_2 (and T_1), a contradiction.

Linear programing.

1. Let x and y be the number of executions of P on M_1 and M_2 respectively. The Linear Program is

$$\begin{aligned} \text{Minimize:} \quad & 3x + y \\ \text{Subject to:} \quad & 4x + y \leq 37 \\ & x + 4y \leq 28 \\ & x + y \geq 10 \\ & x \geq 0 \\ & y \geq 0 \end{aligned}$$

2. The feasible region is indicated in the next Figure. The objective function is represented as a dashed line, so its minimum occurs at point A in the figure.

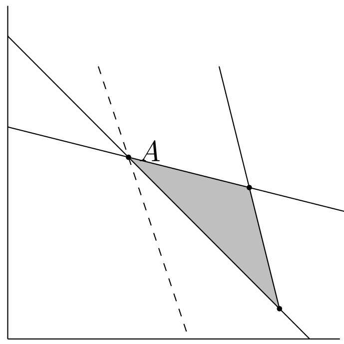


Figure 1: The feasible region and the objective function.

3. We introduce slack variables z, w, t and write the problem as

$$\begin{aligned} \text{Maximize:} \quad & -3x - y \\ \text{Subject to:} \quad & 4x + y + z = 37 \\ & x + 4y + w = 28 \\ & -x - y + t = 10 \\ & x, y, z, w, t \geq 0 \end{aligned}$$

We start with the basic feasible solution $(0, 0, 37, 28, 10)$ on this problem and organize it in a tableau

$$\begin{array}{llll|l} z & = 37 & -x & -4y & \\ w & = 28 & -x & -4y & \\ t & = 10 & +x & +y & \\ \hline c & = -3x & -y & & c = 0 \end{array}$$

Choose x as entering variable. The remaining variables put the restrictions $x \leq 37$ from z, $x \leq 28$ form w a and none for t, the most restrictive is w. We put $x = 28$ and get the new solution $(28, 0, 37, 0, 10)$ and the new tableau

$$\begin{array}{llll|l} z & = 37 & -x & -4y & \\ x & = 28 & & -4y & -w \\ t & = 10 & +x & +y & \\ \hline c & = 3x & +y & & c = 84 \end{array}$$