1. Computability, Complexity and Algorithms

Given a simple directed graph $G = (V, E)$, a cycle cover is a set of vertex-disjoint directed cycles that cover all vertices of the graph.

1. Show that there is a polynomial-time algorithm to find a cycle cover of a directed graph if one exists.
2. Show that deciding if a directed graph has a cycle cover with at most k cycles, for any fixed integer $k \geq 1$, is NP-complete.
3. Show that deciding if a directed graph has a cycle cover where each cycle has at least 1% of the vertices is NP-complete.

Solution:

1. Construct a new graph with two vertices u_1, u_2 for each vertex u of the original graph. For each directed edge (u, v), the new graph has an edge from u_1 to v_2. This graph is bipartite and has a perfect matching iff the original graph has a cycle over. The existence of a perfect matching in a bipartite graph can be checked either by the Hungarian algorithm or by viewing it as a max-flow problem.

2. Reduction from HAM. With $k = 1$, a cycle cover is a Hamilton cycle. So we can check Hamiltonicity by solving the cycle cover problem with $k = 1$. Since HAM is NP-complete, so is this problem. For any fixed integer $k \geq 1$, given a graph $G = (V, E)$, build a new graph H containing G and $k - 1$ vertex-disjoint triangles (for a total of $|V| + 3(k - 1)$ vertices and $|E| + 3(k-1)$ edges). The new graph has a cycle cover with at most k cycles iff G is Hamiltonian.

3. Reduction from HAM. Make a new graph with $100n$ vertices where n is the number of vertices in the original graph as follows: Take a clique of size $99n$ and attach it with a single edge to the given graph G. Then the new graph H has a cycle cover where each cycle has at least 1% of the vertices iff the original graph is Hamiltonian.

2. Analysis of Algorithms

The following LP-relaxation is exact for the maximum weight matching problem in bipartite graphs but not in general graphs. Give a primal–dual algorithm, relaxing complementary slackness conditions appropriately, to show that the integrality gap of this LP is $\geq 1/2$. What is the best upper bound you can place on the integrality gap?

$$
\begin{aligned}
\text{maximize} \quad & \sum_e w_e x_e && && (1)\\
\text{subject to} \quad & \sum_{e:\ e \text{ incident at } v} x_e \leq 1, && v \in V \\
& x_e \geq 0, && e \in E
\end{aligned}
$$

Solution: The dual LP is:

$$\begin{aligned} \text{minimize } & \sum_v y_v & & (2) \\ \text{subject to } & y_u + y_v \geq w_e, & & (u,v) \in E \\ & y_v \geq 0, & & v \in V \end{aligned}$$

The complementary slackness conditions are:

- $x_e > 0 \quad \Rightarrow \quad y_u + y_v = w_e$.
- $y_v > 0 \quad \Rightarrow \quad \sum_{e:\ e \text{ incident at } v} x_e = 1$.

Consider the following algorithm.

Start with the edges of max weight in the current graph, say W. Let the dual at the endpoints of these edges be set to W, find a maximal matching in this set of edges and remove all these vertices. Iterate on the rest of the graph.

It is easy to see that this algorithm is based on relaxing the second condition by a factor of 2, since each matched edge of weight w_e is covered to the extent of $2w_e$. Hence it places a lower bound of 1/2 on the integrality gap of the LP. An upper bound of 2/3 is obtained by considering a traingle of unit weight edges. In this graph, the best integral solution is to match one edge and the best fractional solution is to pick each edge to the extent of 1/2.

3. Theory of Linear Inequalities

Let $P \subseteq [0,1]^n$ be an integral polytope contained in the 0/1 cube, i.e., the polytope as 0/1 vertices. The goal is to maximize an objective $c \in \mathbb{Z}^n$ over P. You are given a feasible integral solution $\bar{x} \in P$ and access to the polytope P is restricted to querying the following oracle:

ℓ_1-penalty oracle:
Input: $x_0 \in P$ integral, $\lambda \in \mathbb{R}_+$, objective $c \in \mathbb{Z}^n$
Output: $x \in P$ integral with

$$c(x - x_0) - \lambda \|x - x_0\|_1 > 0,$$

if such an x exists, otherwise return INFEASIBLE.

Consider the following simple scaling algorithm, where $C := \|c\|_\infty$.

1. Initialize $\lambda \leftarrow 2C$ and $x_0 \leftarrow \bar{x}$.
2. Repeat

(a) Query oracle with x_0, c, λ.

(b) IF the oracle returns a point x, then set $x_0 \leftarrow x$.

(c) ELSE if the oracle returns INFEASIBLE, then set $\lambda \leftarrow \lambda/2$.

3. Until $\lambda < 1/n$.

4. Return x_0.

Task.

- Prove that the algorithm optimizes c over P with $O(n \log nC)$ oracle calls.
- Bonus: Can you further reduce the number of oracle calls to $O(n \log C)$, via a small modification to the algorithm?

Hint. Suppose that for a given choice $\lambda \in \mathbb{R}_+$ the oracle returns INFEASIBLE. Then in particular, also for the integral solution $x^* \in P$ that maximizes c, it holds:

$$\frac{c(x^* - x_0)}{\|x^* - x_0\|_1} \leq \lambda$$

Solution. *Analysis of the number of oracle calls.*

Let x^* denote the integral optimal solution. Let $y_0, \dots, y_i \dots$ denote the sequence of integral feasible solutions returned by the oracle within a number of iterations over which λ did not change. Note, that y_0 is the integral solution returned right after having halved λ or the very first iterate, i.e., $y_0 = \bar{x}$. By the hint/observation, we have

$$\frac{c(x^* - y_0)}{\|x^* - y_0\|_1} \leq 2\lambda,$$

this also clearly holds for the very first iterate by choice of λ.

If we now consider any two consecutive iterates y_i and y_{i+1}, we have

$$c(y_{i+1} - y_i) > \lambda \cdot \|y_{i+1} - y_i\|_1 \geq \lambda,$$

as the $y_i \neq y_{i+1}$. And further by the above

$$\lambda \geq \frac{1}{2} \frac{c(x^* - y_0)}{\|x^* - y_0\|_1} \geq \frac{1}{2n} c(x^* - y_0).$$

Putting this together implies that $c(y_{i+1} - y_i) \geq \frac{1}{2n} c(x^* - y_0)$, i.e., moving from y_i to y_{i+1} recovers at least a $1/2n$ fraction of the improvement of going from y_0 to the optimal solution x^*. Thus for each fixed λ we do at most $2n$ iterations.

Now for the number of changes of λ, observe that λ can be only halved $O(\log nC)$ times until $\lambda < 1/n$ and at that point, that last iterate x satisfies:

$$c(x^* - x) < 1/n\|x^* - x\|_1 < 1,$$

and hence $x^* = x$.

Bonus: further improving the number of oracle calls.

For this we simply observe as soon as $\lambda \leq 1$, then

$$c(x^* - x) \leq \|x^* - x\|_1 \leq n,$$

holds, thus the additive difference in cost is at most n. From here on we can simply continue with $\lambda = 0$, effectively turning the oracle into a simple improvement oracle. The remaining sequence of points satisfies now

$$c(y_{i+1} - y_i) > 0,$$

or equivalently $c(y_{i+1}) \geq c(y_i) + 1$, by integrality. This can happen at most n times yielding the required improvement.

4. Combinatorial Optimization

In the (fractional) multi-commodity flow problem, we are given a directed graph $G = (V, E)$ and pairs $(s_1, t_1), \ldots, (s_k, t_k)$ of vertices of G, a capacity function $c : E \to \mathbb{Q}_{\geq 0}$, and demands $d_1, \ldots, d_k$, and we seek to find for each $i = 1, \ldots, k$ an $s_i - t_i$-flow $x_i \in \mathbb{Q}^E_{\geq 0}$ so that x_i has value d_i and so that for each arc e of G: $\sum_{i=1}^k x_i(e) \leq c(e)$.

Question 1. Show with Farkas' Lemma that the multicommodity flow problem has a solution if and only if for each 'length' function $l : E \to \mathbb{Q}_{\geq 0}$ one has: $\sum_{i=1}^k d_i dist_l(s_i, t_i) \leq \sum_{e \in E} l(e)c(e)$. (Here $dist_l(s, t)$ denotes the length of a shortest $s - t$ path with respect to l.)

Question 2. The cut condition states that for each $W \subseteq V$, the capacity of $\delta^{out}(W)$ is not less than the demand of W, where the capacity of $\delta^{out}(W)$ is $\text{cap}(\delta^{out}(W)) := \sum(c(e) : e \in \delta^{out}(W))$ and the demand of W is $\sum(d_i : s_i \in W \text{ and } t_i \notin W)$. Interpret the cut condition as a special case of the condition in Question 1.

Solution. Question 1. Farkas' Lemma: $\exists x \geq 0 : Ax = b \iff \forall y, y^T A \geq 0 : y^T b \geq 0$.
Let

$$A = \begin{pmatrix} A_1 & A_2 & \cdots & A_k & I_{|E|} \\ 1 \cdots 1 & & & & 0 \\ & 1 \cdots 1 & & & 0 \\ & & \ddots & & \vdots \\ & & & 1 \cdots 1 & 0 \end{pmatrix}, \tag{3}$$

where A_i is the arc-path incidence matrix, in which the columns represent all paths from s_i to t_i. Let $b = (c, d)$ with c the capacity vector and d the demand vector. There exists a feasible solution to the multi-commodity flow problem if and only if there is a solution $x \geq 0$ such that $Ax = b$.

Assume there is a solution and let l be a length function, which we interpret as a vector in $Q^{|E|}$. Define $y^T = (l^T \mid -dist_l(s_1, t_1), -dist_l(s_2, t_2), \ldots, -dist_l(s_k, t_k))$. Now $y^T A \geq 0$, because $l^T A_i$ denotes the length of the paths P_{ij}. Farkas' Lemma now gives: $0 \leq y^T b = \sum_{e \in E} l(e)c(e) - \sum_{i=1}^{k} d_i dist_l(s_i, t_i)$.

Next, assume that $y^T = (p^T | q^T)$ is an arbitrary vector satisfying $y^T A \geq 0$. Then the coefficients of p cannot be negative, because $p^T I_{|E|} \geq 0$. These coefficients therefore define a length function and $p^T A_i$ gives the length of the paths from s_i to t_i. Now $y^T A \geq 0$ implies (using $l := p$) $-q_i \leq \sum_{e \in P_{ij}} l(e)$, for all paths P_{ij}. Thus, $-q_i \leq dist_l(s_i, t_i)$. Now we have

$$y^T b = \sum_{e \in E} l(e)c(e) + \sum_{i=1}^{k} q_i d_i \geq \sum_{[e \in E} l)e)c(e) - \sum_{i=1}^{k} dist_l(s_i, t_i) d_i \geq 0.$$

where the last inequality follows from the assumption in the question. According to Farkas' Lemma, there exists a solution to the multi-commodity flow problem.

Question 2. Let $W \subseteq V$ be a cut. Define the length function $l_W : E \to \mathbb{R}$ as follows: $l_W(e) = 0$ if $e \notin \delta_E^{out}(W)$ and $l_W(e) = 1$ if $e \in \delta_E^{out}(W)$. Then $dist_l(s_i, t_i) = 1$ if $s_i \in W$ and $t_i \notin W$, and and $dist_l(s_i, t_i) = 0$ otherwise. Thus

$$\sum_{s_i \in W, t_i \notin W} d_i = \sum_{i=1}^{k} d_i dist_{l_W}(s_i, t_i) \leq \sum_{e \in E} l_W(e)c(e) = \sum_{e \in \delta^{out}(W)} c(e) = \text{cap}(\delta^{out}(W)).$$

5. Graph Theory

We are given two square sheets of paper, each of area 2015. Each sheet is divided into 2015 polygons of area 1 (the divisions may be different). One sheet is placed on top of the other. Show that we can place 2015 pins in such a way that the interior of each of the 4030 polygons is pierced.

Solution: Let G be a graph with 4030 vertices, each representing a polygon. We place an edge between two vertices if their polygons are in different sheets of paper and when one sheet is placed on top of the other the interiors of these polygons have at least one point in common. Note that if S is a set of k polygons in the first sheet, they cover an area equal to k. Thus, to cover them we need at least k polygons of the second sheet of paper. This implies the condition of Hall's theorem. Thus, we have a perfect matching. Using this matching, we immediately deduce where to place the 2015 pins.

6. Probabilistic methods

Consider the random graph $G := G_{n,p}$ with $p := p(n) = 1/(6\sqrt{n})$, and let S be a fixed subset of $k \geq 2$ vertices of G, where $(k/6000 \ln k)^2 \leq n$. Let Y be the maximum size of a set of edge-disjoint triangles in G such that every triangle in the set has at least two vertices in S. Prove that for every positive integer t

$$\Pr(Y \geq t) \leq \frac{(30k \ln k)^t}{t!},$$

and deduce that

$$\Pr(Y \geq 120k \ln k) < k^{-3k}.$$

You may assume that k is sufficiently large.
Remark. The constant "3" in the last expression may be improved, but to do so may require a calculator. The stated bound can be derived using mental arithmetic only.

Solution: If $Y \geq t$, then there exists a collection of t unordered triples $\{a_1, b_1, c_1\}, \ldots, \{a_t, b_t, c_t\}$ of vertices such that (i) no pair of vertices lies in two triples, (ii) for each i, we have $a_i, b_i \in S$, and (iii) each triple forms a triangle in G. The expected number of such collections is at most

$$\binom{\binom{k}{2}}{t}(n-2)^t p^{3t} < \frac{(30k \ln k)^t}{t!},$$

and so, by Markov's inequality, we get the bound claimed on $\Pr(Y \geq t)$. Letting $t := \lceil 120k \ln k \rceil$ we obtain

$$\Pr(Y \geq 120k \ln k) \leq \frac{(30k \ln k)^{120k \ln k + 1}}{(120k \ln k/e)^{120k \ln k}} \leq 30k \ln k \left(\frac{e}{4}\right)^{120k \ln k} < k^{-3k},$$

because $(e/4)^6 < (3/4)^6 < (1/2)^2 < 1/e$, leaving a lot of room to spare.

7. Algebra

Which of the following rings are isomorphic? Justify your answers.

1. $R_0 = \mathbb{F}_5[X]/(X^2)$
2. $R_1 = \mathbb{F}_5[X]/(X^2 - 1)$
3. $R_2 = \mathbb{F}_5[X]/(X^2 - 2)$
4. $R_3 = \mathbb{F}_5[X]/(X^2 - 3)$

Solution: Since $X^2 - 2$ and $X^2 - 3$ are irreducible modulo 5 (neither 2 nor 3 is a square), both R_2 and R_3 are fields with 5^2 elements, and are hence isomorphic. By the Chinese remainder theorem,

$$R_1 = \mathbb{F}_5[X]/(X+1)(X-1) \cong \mathbb{F}_5[X]/(X+1) \times \mathbb{F}_5[X]/(X-1) \cong \mathbb{F}_5 \times \mathbb{F}_5.$$

This ring is not an integral domain, so it is not isomorphic to R_3 or R_4. Also R_0 is not an integral domain as $X^2 = 0$ but $X \neq 0$, and R_0 is not isomorphic to R_1 because R_1 has no nilpotent elements.