1. Theory of Linear Inequalities

Let $a_0, a_1, \ldots a_n$ be integral vectors in $\mathbf{R}^d$ and let A be the matrix having $a_0, a_1, \ldots, a_n$ as columns. For a nonnegative integer λ, call $\{a_0, a_1, \ldots, a_n\}$ a (a_0, λ)-*Hilbert basis* if every integral vector b in cone($\{a_0, a_1, \ldots, a_n\}$) can be written as integral combination

$$b = \sum (\gamma_i a_i : i = 0, 1, \ldots n)$$

where $\gamma_0 + \lambda \geq 0$ and $\gamma_i \geq 0$ for $i = 1, 2, \ldots, n$. Suppose cone$\{a_0, a_1, \ldots, a_n\}$ is a pointed cone. Show that $\{a_0, a_1, \ldots, a_n\}$ is **not** a (a_0, λ)-Hilbert basis if and only if there is an integral vector b in cone($\{a_0, a_1, \ldots, a_n\}$) such that $b + \lambda a_0$ is not a nonnegative integer combination of $a_0, a_1, \ldots, a_n$ and

$$max\{\mathbf{1}^T x | Ax = b, x \geq 0\} < d.$$

Solution. A solution is available upon request.

2. Combinatorial Optimization

Let $G = (V, E)$ be a graph and let k be the cardinality of a maximum matching in G. Let E_1 and E_2 be non-empty subsets of E with $E_1 \cup E_2 = E$. For $i = 1, 2$, let k_i be the cardinality of a maximum matching of G contained in E_i. If $k_1 + k_2 = k$, then (E_1, E_2) is called a *matching separation* of G. If G has no matching separation then it is called *matching nonseparable*.

A graph $G = (V, E)$ is called *factor-critical* if for each vertex $v \in V$ the graph obtained by deleting v from G has a perfect matching.

Show that a graph G with no isolated vertices is matching nonseparable if and only if G is isomorphic to $K_{1,t}$ for some t **or** G is factor-critical and 2-vertex connected.

Solution. A solution is available upon request.

3. Analysis of Algorithms

(a) Given n points in the plane, design an $O(n^2 logn)$ algorithm that determines if any three points are colinear.

(b) Assume you have n square containers of different sizes with matching lids. Consider the following randomized algorithm for finding the largest container or lid.

Draw a lid uniformly at random from the set of n lids, and draw a container uniformly at random from the set of n containers. If the lid is smaller than the container, discard the lid, draw a new lid uniformly at random from the remaining lids, and repeat. If the container is smaller than the lid discard the container, draw a new container uniformly at random from the remaining containers, and repeat. If they are the same size flip a fair coin to decide which one to discard and repeat. Stop when you run out of containers or lids. When you stop, the container (or lid) you are holding is the largest.

What is the expected number of lid-container tests performed by this algorithm? Prove your answer is correct.

Solution:
(a) Let P be the set of n points. Notice that if three points p, q and r are collinear, with q being the central point, then the slopes of segments $\overline{pq}$ and $\overline{qr}$ are identical. Consider the following algorithm:

Select a point $q \in S$, compute the slope of $\overline{qs}$ for all $s \in S$ such that $s \neq q$. Now take the $n-1$ slopes and sort them. If any two slopes are the same then we have found our collinear points. Repeat for each point $q \in S$. This takes $O(nlogn)$ time for each point q for a total running time of $O(n^2 logn)$.

(b) Each time a lid-container test is performed an item (lid or container) is discarded. Let R be the number of items not discarded. Then the expected number of tests is equal to $2n - E(R)$. Let X_i be a random variable which is 0 if item i is discarded and 1 otherwise. By linearity of expectation $E(R) = \sum_{i=1}^{2n} E(X_i)$. Now let us consider any lid i that is not the largest lid. Let p_l be the probability i is not discarded given that we are left at the end of the process holding a lid and similarly let p_c be the probability i is not discarded given that we are left holding a container. With equal probability we will be left holding a lid or a container. Notice that if we are left holding a container then all lids have been discarded so $p_c = 0$. If we are left with a lid then the probabiliy that i has not been discarded is the probability that the largest lid was drawn before lid i. We can consider the process of drawing lids the same as selecting a random permutation of lids. In exactly half the permutations lid i is drawn after the largest lid so $p_l = \frac{1}{2}$. Therefore, $E(X_i) = \frac{1}{2}p_l + \frac{1}{2}p_c = \frac{1}{4}$. If i is the largest lid then if we are left holding a lid it is lid i so $E(X_i) = \frac{1}{2}$. This analysis is the same for containers so if we sum over all containers and lids, $E(R) = \sum_{i=1}^{2n} E(X_i) = 2(\frac{1}{2}) + 2(n-1)\frac{1}{4} = \frac{n+1}{2}$. Therefore the expected number of tests is $2n - \frac{n+1}{2}$.

4. Approximation Algorithms

1. Consider the $(2 - 2/k)$ factor algorithm for the multiway cut problem that operates by finding minimum cuts separating each terminal from all the rest. Show that the analogous algorithm for the node multiway cut problem, based on isolating cuts, does not achieve a constant factor. What is the best factor you can prove for this algorithm?
2. The multiway cut problem also possesses the half-integrality property. Give a suitable LP for the multiway cut problem and prove this fact.

Solution.
1. Consider the following graph $G = (V, E)$ on $2n + 1$ vertices of which $s_1, \ldots, s_n$ are terminals. In addition, there are vertices $v_1, \ldots, v_n$ and w. The edges are: (s_i, v_i) and (v_i, w), for $1 \leq i \leq n$. The costs of v_i's are $1 - \epsilon$ and the cost of w is 1. It is easy to see that the algorithm will produce a cut of cost $n(1 - \epsilon)$ by picking all v_i's and OPT is simply w of cost 1. The best factor for this algorithm is $O(n)$.

2. The LP for multiway cut problem we will use picks edges fractionally – let d_e be the edge variables – and ensures that each path between 2 terminals encounters a total d_e of 1. The dual is a maximum multicommodity flow between all pairs of terminals.

Consider any fractional optimal primal solution. Shrink edges having $d_e = 0$. Next, if v is a degree 2 vertex with edges (u, v) and (v, w), then replace the 2 edges by (u, w) with the sum of the two distance labels.

Now, any edges connecting 2 terminals must have distance label of 1. By optimality of the solution, the remaining edges must have distance label of < 1.

Next, we claim that any path carrying non-zero flow between 2 terminals consists of 1 or 2 edges. Suppose p is a path of 3 or more edges running between s_i and s_j. Clearly, an intermediate vertex has a path to a terminal, say s_k; it may be the same as s_i or s_j. Now, out of the 2 paths, s_i to s_k and s_k to s_j, one must have length < 1, contradicting feasibility.

Finally, we set edges on length 1 paths to 1 and those on length 2 paths to 1/2. By complementary slackness conditions and the previous fact, this solution must be optimal. Hence there is a half-integral optimal solution.

5. Graph Theory

For a graph G we use $e(G)$ to denote the number of edges of G. Let H be a spanning subgraph of a graph G such that every component of H is an induced subgraph of G and is bipartite. Prove that G has a bipartite subgraph with at least $e(G)/2 + e(H)/2$ edges.

Solution: Let H_i, $i = 1, \dots, k$, be the components of H. Then H_i is bipartite; so let A_i, B_i denote a partition of $V(H_i)$ such that all edges of H_i are between A_i and B_i.

We form a partition V_1, V_2 of $V(G)$ such that

(1) for $i = 1, \dots, k$, one of A_i, B_i is contained in V_1 and the other is contained in V_2, and

(2) for $i = 2, \dots, k$, at least half of the edges in $[H_i, \bigcup_{j=1}^{i-1} H_j]$ (the set of edges with one end in H_i and the other in $\bigcup_{j=1}^{i-1} H_j$) are in $[V_1, V_2]$.

This can be done by placing one of A_i, B_i in V_1 and the other in V_2, in order $i = 1, \dots, k$. At the ith stage, if (2) does not hold then switch A_i and B_i.

Since each H_i is an induced subgraph of G,

$$\sum_{i=2}^{k} e(H_i, \cup_{j=1}^{i-1} H_j) = e(G) - e(H).$$

Now the number of edges of G between V_1 and V_2 is at least

$$e(H) + \sum_{i=2}^{k} (1/2)|[H_i, \cup_{j=1}^{i-1} H_j]| = e(H) + (e(G) - e(H))/2 = e(G)/2 + e(H)/2.$$

6. Probability

Suppose $(X_n)_{n=1}^{\infty}$ is a sequence of independent standard Gaussian random variables. Prove:

$$\mathsf{P}\left\{\limsup_{n\to\infty}\frac{X_n}{\sqrt{2\ln n}}=1\right\}=1.$$

Hint. Use the asymptotic relation

$$\frac{1}{\sqrt{2\pi}}\int_x^{\infty}e^{-y^2/2}dy \;\sim\; \frac{1}{x\sqrt{2\pi}}e^{-x^2/2},\quad x\to\infty, \tag{1}$$

to show

$$\mathsf{P}\left\{\limsup_{n\to\infty}\frac{X_n}{\sqrt{2\ln n}}\le 1\right\}=1, \tag{2}$$

and

$$\mathsf{P}\left\{\limsup_{n\to\infty}\frac{X_n}{\sqrt{2\ln n}}\ge 1\right\}=1. \tag{3}$$

Solution: Relation (2) is equivalent to the fact that for any $\varepsilon>0$,

$$\mathsf{P}\left\{X_n\ge(1+\varepsilon)\sqrt{2\ln n}\quad\text{i.o.}\right\}=0. \tag{4}$$

Estimate (1) implies that for some constant $K>0$

$$\begin{aligned}\sum_{n=2}^{\infty}\mathsf{P}\left\{X_n\ge(1+\varepsilon)\sqrt{2\ln n}\right\}&\le K\sum_{n=2}^{\infty}\frac{1}{(1+\varepsilon)\sqrt{2\ln n}\sqrt{2\pi}}e^{-((1+\varepsilon)\sqrt{2\ln n})^2/2}\\&\le K\sum_{n=2}^{\infty}\frac{1}{(1+\varepsilon)\sqrt{2\ln n}\sqrt{2\pi}}n^{-(1+\varepsilon)^2}<\infty,\end{aligned}$$

The desired relation (4) follows by the Borel–Cantelli Lemma.

Relation (3) is equivalent to the fact that for any small $\varepsilon>0$,

$$\mathsf{P}\left\{X_n\ge(1-\varepsilon)\sqrt{2\ln n}\quad\text{i.o.}\right\}=1. \tag{5}$$

Estimate (1) implies that for some constant $C>0$

$$\begin{aligned}\sum_{n=2}^{\infty}\mathsf{P}\left\{X_n\ge(1-\varepsilon)\sqrt{2\ln n}\right\}&\ge C\sum_{n=2}^{\infty}\frac{1}{(1-\varepsilon)\sqrt{2\ln n}\sqrt{2\pi}}e^{-((1-\varepsilon)\sqrt{2\ln n})^2/2}\\&\ge C\sum_{n=2}^{\infty}\frac{1}{(1-\varepsilon)\sqrt{2\ln n}\sqrt{2\pi}}n^{-(1-\varepsilon)^2}=\infty,\end{aligned}$$

The desired relation (5) follows by the second part of Borel–Cantelli Lemma.

7. Algebra

Let a_n denote the Fibonacci sequence $a_0 = 0$, $a_1 = 1$, $a_n = a_{n-1} + a_{n-2}$, and let $b_n = (a_n)^2$. Prove that b_n satisfies a linear recursion relation.

Solution:

$$b_n = (a_{n-1} + a_{n-2})^2 = b_{n-1} + b_{n-2} + 2a_{n-1}a_{n-2}. \tag{6}$$

Now use a_{n-1} in terms of a_{n-2}, a_{n-3} to get

$$b_n = b_{n-1} + 3b_{n-2} + 2a_{n-2}a_{n-3}.$$

On the other hand, replacing n by $n-1$ in (6) gives

$$b_{n-1} = b_{n-2} + b_{n-3} + 2a_{n-2}a_{n-3}.$$

Subtracting the last two equations gives the desired linear recursion $b_n = 2b_{n-1} + 2b_{n-2} - b_{n-3}$.