1. Analysis of Algorithms

Consider the following two graph problems:

Graph coloring: Given a graph $G = (V, E)$ and an integer $c \geq 0$, a c-coloring is a function $f : V \to \{1, 2, \ldots, c\}$ such that $f(u) \neq f(v)$ for all edges $uv \in E(G)$. We define $\chi(G)$ to be the minimum integer $c \geq 0$ such that G has a c-coloring. An α-approximation algorithm for the coloring problem is one that, given a graph G, returns a c-coloring, where $c \leq \alpha\chi(G)$.

Longest path: Given a graph $G = (V, E)$, a *path* P in G is a sequence of distinct vertices $v_1, \ldots, v_d$ such that $v_i v_{i+1} \in E$ for all $1 \leq i < d$, and the length of the path P is d. We wish to find a path which maximizes d. We write that the longest path uses $\ell(G)$ vertices. An α-approximation algorithm for the longest path problem is one that, given a graph G, returns a path of length at least $\ell(G)/\alpha$.

Unfortunately, unless P = NP, there is no polynomial time $n^{1-\epsilon}$-approximation algorithm for either problem, for any constant $\epsilon \in (0, 1)$, where $n = |V|$. Taken together, however, the problems are easier. In this question, you will give a linear time algorithm that, for any graph G and for any $0 < \epsilon < 1$, outputs either an $n^{1-\epsilon}$-approximation to the graph coloring problem, or a n^{ϵ}-approximation to the longest path problem.

(a) For any c-coloring f of a graph $G = (V, E)$, the color class of color i is defined to be $C(i) = \{v \in V : f(v) = i\}$. Given an n vertex graph, use DFS to find some parameter k such that there is a k-coloring as well as a path in the graph of length k so that each color class of the coloring contains exactly one vertex on the path.

(b) Using part (a), for any graph G and parameter $\epsilon \in (0, 1)$, show how to output in linear time either a path that is an n^{ϵ}-approximation for the longest path problem, or a c-coloring that is an $n^{1-\epsilon}$-approximation for the graph coloring problem. (Hint: Observe that each path has length at most n and each coloring has at least 1 color.)

Solution:
(a) Consider a DFS tree and let its depth be d. Assign the vertices at each level of the tree a separate color. Since there are no cross edges, there are no edges connecting two vertices at the same level of the tree, and this is a valid coloring. Taking the path from the deepest leaf to root gives a path whose length is the same as the number of colors.
(b) By part (a), there exist a path of length k as well as a coloring of the graph using k colors. If $k < n^{\epsilon}$, then the coloring is an $n^{\epsilon}/1 = n^{\epsilon}$-approximation to coloring. Otherwise $k \geq n^{\epsilon}$ and the path of length k is an $n/n^{\epsilon} = n^{1-\epsilon}$-approximation to longest path.

2. Approximation Algorithms

Minimizing Shipping Times. A factory produces m different kinds of items (say $\{1, \ldots, m\}$), and produces one item of each kind every time step. Let $\mathcal{R} = \{R_1, \ldots R_n\}$ be a set of customer requests. Every request contains at most one item of each kind, and hence a request R_i is specified by a subset $R_i \subseteq \{1, \ldots, m\}$.

The items produced by the factory are to be allocated to the requests $\mathcal{R}$. The shipping time of a request R_i is the time at which all its items are allocated. The goal is to find an allocation that minimizes the total shipping time of all the requests.

In particular, design a polynomial time algorithm that outputs an allocation whose total shipping time is within a factor 2 of the optimal allocation (a 2-approximation algorithm).

(Hint: Write a linear program with **only** the shipping times of the requests as variables.)

Solution.
Let c_i be the shipping time of request R_i. For each $i \in \{1, \ldots, m\}$, let S_i denote the subset of requests that need an item of kind i.

Consider the following linear program with exponentially many constraints.

$$\text{Minimize} \sum_{i=1}^{n} c_i \tag{1}$$

$$\text{Subject to} \sum_{s \in S} c_s \geq \frac{|S|(|S|+1)}{2} \qquad \text{for all } S \subseteq S_i, i \in \{1, \ldots, m\} \tag{2}$$

All the requests in a subset $S \subseteq S_i$ need an item of type i. Hence the sum of shipping times of the requests in S is at least $1 + 2 + \ldots + |S| = \frac{|S|(|S|+1)}{2}$. The linear program includes one such constraint for every subset $S \subseteq S_i$ of every S_i.

The linear program has exponentially many constraints, but there is a polynomial time separation oracle. Given a solution $\{c_i\}$, run the following separation oracle:

For each $i \in \{1, \ldots, m\}$ do

- Sort the numbers $\{c_s | s \in S_i\}$ in increasing order to get a sequence $(c_{j_1}, c_{j_2}, \ldots, c_{j_{|S_i|}})$.
- For every $t \in \{1, \ldots, |S_i|\}$, check that

$$\sum_{\ell=1}^{t} c_{j_\ell} \geq \frac{t(t+1)}{2}$$

. If for some t, the above inequality is unsatisfied, then output the corresponding constraint.

Rounding For every $i \in \{1, \ldots, m\}$, allocate goods of kind i to the requests in increasing order of shipping times.

Claim 1 *Every request R_i is shipped by time $2c_i$.*

Proof. Consider a type $j \in R_i$. Let A be the set of requests interested in item j that have shipping times at most c_i.

$$A = \{\ell | j \in R_\ell, c_\ell \leq c_i\}$$

Observe that,

$$\frac{|A|(|A|+1)}{2} \leq \sum_{\ell \in A} c_\ell \leq |A| c_i \ .$$

Hence we have

$$c_i \geq (|A|+1)/2 \ .$$

As items are allocated in increasing order of shipping times, item j is allocated to R_i by time $|A| < 2c_i$.

From the above claim, it follows that the allocation is a 2-approximation to the optimal allocation.

3. Theory of Linear Inequalities

Let P be a non-empty bounded polyhedron in R^n, let $c \in R^n$, and let $a_i \in R^n$ and $b_i \in R$ for $i = 1, \ldots, t$ for some positive integer t. Show that the linear programing problem

$$\max(c^T x : x \in P, a_1^T x = b_1, \ldots, a_t^T x = b_t)$$

has an optimal solution that is a convex combination of at most $t+1$ vertices of P.

Solution. A solution is available upon request.

4. Combinatorial Optimization

Let $G = (V, E)$ be a complete graph, with vertex set V and edge set E; let $n = |V|$. Consider the polytope P defined by

$$x_i + x_j \leq 1, \text{ for all } \{i, j\} \in E$$
$$0 \leq x_i \leq 1, \text{ for all } i \in V.$$

Show that the Chvátal rank of P is at least $\log_2(n-1)$.

Solution. A solution is available upon request.

5. Graph Theory

Let G be a connected graph. (a) Use a depth-first-search spanning tree to prove that if G is triangle-free, then G contains a bipartite subgraph H such that $|E(H)| \geq 3(|V(G)|-1)/4$ and every component of H is an induced subgraph of G. (b) Prove an analogous bound when G has odd girth $g \geq 5$ by replacing the constant $3/4$ by an appropriate function of g. [A graph G has *odd girth* g if g is the largest integer such that every odd cycle in G has length at least g.]

Solution: (a) Grow a DFS spanning tree T from a vertex u, the root. For each $x \in V(T)$, let $T[u, x]$ denote the unique path in T from u to x. Let $v \in V(T)$ such that that $|V(T[u, v])|$ is maximum. If $|V(T[u, v])| \leq 4$ then G is bipartite (as G is triangle-free), and we have (a) with $H = G$.

So assume $|V(T[u, v])| \geq 5$. Let $w \in V(T[u, v])$ such that the distance in T from w to v is 3. Define T_w as the subtree of T consisting of w and all it descendants. Then $T - T_w$ is a DFS spanning tree for $G' := G - T_w$.

By induction G' has a bipartite subgraph H' such that $|E(H')| \geq 3(|V(G)| - |V(T_w)| - 1)/4$ and every component of H' is an induced subgraph of G' (hence of G). Note that $G[T_w]$ is bipartite (as G is triangle-free). Hence, $H := H' \cup G[T_w]$ is bipartite, $|E(H)| \geq 3(|V(G)| - |V(T_w)| - 1)/4 + (|V(T_w)| - 1) \geq 3(|G| - 1)/4$, and every component of H is an induced subgraph of G.

(b) One can modify the proof to show the existence of a bipartite subgraph H of G such that $|E(H)| \geq \frac{g-2}{g-1}(|V(G)| - 1)$ and every component of H is an induced subgraph of G.

6. Probability

Let $(X_n)_{n=1}^{\infty}$ be a sequence of i.i.d. random variables taking values ± 1 with probability 1/2. Let

$$S_n = \begin{cases} X_1 + \ldots + X_n, & n = 1, 2, \ldots \\ 0, & n = 0. \end{cases}$$

Let $S_n^* = \max_{k=1,\ldots,n} S_k$, $n = 1, 2, \ldots$

Find the density of the limiting distribution for $\frac{S_n^*}{\sqrt{n}}$ and prove the convergence using the following "reflection principle": for any $n, r = 1, 2, \ldots$,

$$\mathsf{P}\{S_n^* > r\} = 2\mathsf{P}\{S_n > r+1\} + \mathsf{P}\{S_n = r+1\}.$$

You do not need to prove the reflection principle.

Solution: One of the ways to prove weak convergence of distributions is to check the convergence of distribution functions at each point of continuity of the limiting distribution function.

So, let us fix $x \in \mathbb{R}$, and try to find $\lim_{n\to\infty} \mathsf{P}\{S_n^*/\sqrt{n} \le x\}$. If $x \le 0$, the limit is clearly 0. For $x > 0$, we can write

$$\mathsf{P}\{S_n^*/\sqrt{n} \le x\} = 1 - \mathsf{P}\{S_n^*/\sqrt{n} > x\}.$$

Using the reflection principle, we can approximate this quantity by

$$1 - 2\mathsf{P}\{S_n/\sqrt{n} > x\},$$

which by the Central Limit Theorem converges to

$$F(x) = 1 - 2\mathsf{P}\{N > x\} = 1 - \frac{2}{\sqrt{2\pi}} \int_x^{\infty} e^{-y^2/2} dy,$$

where N is a standard Gaussian random variable. $F(\cdot)$ is a distribution function (we set $F(x) = 0$ for $x \le 0$), continuous at all points. Therefore, the distribution of $S_n^*/\sqrt{n}$ convergerges weakly to the distribution with distribution function $F(\cdot)$.

The density of the limiting distribution is

$$F'(x) = \begin{cases} \sqrt{\frac{2}{\pi}} e^{-x^2/2}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

7. Algebra

Let $G = GL(2, \mathbb{C})$ denote the group of invertible 2 by 2 matrices with complex entries, and B denote the subgroup of G that consists of all matrices with lower-left entry 0. Recall that the double-coset $[g]$ of $g \in G$ with respect to B is the set $\{bgb' \,|b, b' \in B\}$. Let $X = \{[g] \,| g \in G\}$ denote the set of double cosets of B in G. Prove that there exist two matrices g_0 and g_1 in G such that $X = \{[g_0], [g_1]\}$.

Solution: Take a matrix $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G$ and do row operations on g. A row operation writes $g = bg'$. There are two cases. Case 1: $a \neq 0$. Then, $g'_{21} = 0$. Case 2: $c \neq 0$. Then, $g'_{11} = 0$. In case 1, we have $b' := g' \in B$ and $g = bIb'$, where I is the identity matrix. In case 2, the calculation

$$\begin{pmatrix} 0 & b \\ c & d \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} c & d \\ 0 & b \end{pmatrix}$$

shows that $g = bwb'$ where

$$w = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

Let $g_0 = I$, $g_1 = w$. It is easy to see that the double cosets $[g_0]$ and $[g_1]$ are discjoint, and the above calculation shows that $X = \{[g_0], [g_1]\}$.

Note B is the standard Borel subgroup of G. For $G = GL(n, \mathbb{C})$ the double cosets are given by permutation matrices g_i with zeros in all but one entry of each row and column. This is nothing but the Gauss elimination procedure done to a matrix with row operations, and row-echelon forms.