Approximation Algorithms CS599 **Homework 1**
Date: October 9 , 2014 **Due:** October 24, 2014

1. **Extended Formulations.** Consider a linear program defined over a feasible region given by $P = \{x \in \mathbb{R}^n : Ax \leq b\}$. As we have seen, the number of constraints can be very large even for simple optimization problems. $Q = \{(x, y) \in \mathbb{R}^{n+m} : Cx + Dy \leq e\}$ is called an extended formulation for P if $\{x \in \mathbb{R}^n : (x, y) \in Q\} = P$, i.e., the projection of Q on the x-coordinates equals P. In many cases, while P may have exponential constraints, there exists extended formulations which have polynomially many constraints and variables.

   (a) Consider the simplex embedding linear program for the multiway cut problem. Give a compact extended formulation for the problem.

$$
\begin{array}{ll}
\min \sum_{e \in E} c_e x_e & \\
\text{s.t.} & \\
\sum_{i=1}^{k} y_v^i = 1 & \forall v \in V \\
y_{t_i}^i = 1 & \forall v \in V \\
x_{uv} \geq \frac{1}{2} \sum_{i=1}^{k} |y_u^i - y_v^i| & \forall u, v \in V \\
0 \leq x_{uv} \leq 1 & \forall u, v \in V \\
0 \leq y_v^i \leq 1 & \forall 1 \leq i \leq k, v \in V
\end{array}
$$

Figure 1: Simplex Linear Program for Multiway Cut

   (b) Consider the path based formulation for the multiway cut problem. Give an extended formulation with polynomially many constraints and variables. Hint: Use distances as variables.

$$
\begin{array}{ll}
\min \sum_{e \in E} c_e x_e & \\
\text{s.t.} & \\
\sum_{e \in P_{ij}} x_e \geq 1 & \forall \text{ paths } P_{ij} \text{ from } t_i \text{ to } t_j, \forall 1 \leq i < j \leq k \\
0 \leq x_e \leq 1 & \forall e \in E
\end{array}
$$

Figure 2: Path Based Linear Program for Multiway Cut

2. **Randomized Rounding for Satisfiability**. An instance for MAX-k-SAT consists of clauses with k literals each where every literal is either a boolean variable x_i or its negation $\bar{x}_i$. An instance of MAX-SAT consists of clauses with any number of literals, including one. The goal in each of these problems is to find an assignment of variables to true or false to maximize the number of satisfied clauses.

   (a) Show that a random assignment gives a $(1 - \frac{1}{2^k})$-approximation for MAX-k-SAT and $\frac{1}{2}$-approximation for MAX-SAT.

(b) Consider the following linear program for MAX-SAT given in Figure 3. Let y_i be the indicator variable whether the boolean variable x_i is set to true. Let z_j be the indicator variable for clause j whether it is satisfied by the assignment given by y. Show that the linear program is a valid relaxation for the MAX-SAT problem.

$$\begin{array}{ll}
\max \sum_{j=1}^{m} z_j & \\
\text{s.t.} & \\
\sum_{i \in P_j} y_i + \sum_{i \in N_j} (1 - y_i) \geq z_j & \forall C_j = \vee_{i \in P_j} x_i \bigvee \vee_{i \in N_j} \bar{x_i} \\
0 \leq y_i \leq 1 & \forall 1 \leq i \leq n \\
0 \leq z_j \leq 1 & \forall 1 \leq j \leq m
\end{array}$$

Figure 3: Linear Program for MAX-SAT

Consider the rounding algorithm where we set the boolean variable x_i to true with probability $\frac{y_i}{2} + \frac{1}{4}$ independently. Show that it gives a $\frac{3}{4}$-approximation algorithm.

3. **Iterative Rounding.** In an instance of the scheduling problem, we are given a set of jobs J and machines M, a processing time p_j for each $j \in J$. We are also given a target T_i of maximum makespan for machine i and cost c_{ij} of scheduling job j on machine i. The goal is to find an assignment of jobs to machines of minimum cost such that the total processing time of jobs assigned to machine i is at most T_i. Consider the linear program for the problem.

$$\begin{array}{ll}
\min \sum_{i \in M, j \in J} c_{ij} x_{ij} & \\
\text{s.t.} & \\
\sum_{i \in M} x_{ij} = 1 & j \in J \\
\sum_{j \in J} p_j x_{ij} \leq T_i & i \in M \\
0 \leq x_{uv} \leq 1 & \forall (u, v) \in E
\end{array}$$

Figure 4: Linear Program for Scheduling Problem

We now outline an iterative algorithm which will given assignment whose cost is at most the cost of the LP and the total processing time of jobs assigned to machine i is at most $T_i + \max_j p_j$ using the following steps.

(a) Argue how to update the problem when we find either a variable set to $\{0, 1\}$.

(b) Show that for every extreme point solution either there is a variable with value 0 or 1 or a machine i such that $\cancel{|\{j \in J : x_{ij} > 0\}| \leq 2}$ $|\{j \in J : x_{ij} > 0\}| \leq 1$.

(c) (*) Use the above property to give an iterative algorithm with the claimed guarantee.

4. **Low Dimension Embeddings.** Consider a high-dimensional space $\mathbb{R}^N$ equipped with the ℓ_1 norm. Let V be an n-point subset of $\mathbb{R}^N$. The goal of this exercise is to show that V can

be isometrically embedded into $\binom{n}{2}$ dimensional ℓ_1 space i.e., there exists a map $\varphi : V \to \mathbb{R}^{\binom{n}{2}}$ such that for every $x, y \in V$,

$$\|x - y\|_1 = \|\varphi(x) - \varphi(y)\|_1,$$

where $\|x - y\|_1 = \sum_{i=1}^{N} |x_i - y_i|$ and $\|\varphi(x) - \varphi(y)\|_1 = \sum_{i=1}^{\binom{n}{2}} |\varphi(x)_i - \varphi(y)_i|$.

(a) First, show that V can be isometrically embedded into $\mathbb{R}^{2^n}$.

(b) Recall, that every ℓ_1 metric can be represented as a sum of cut metrics with nonnegative coefficients:

$$\|x - y\|_1 = \sum_{S \subset V} \lambda_S \delta_S(x, y),$$

where δ_S is the cut metric

$$\delta_S(x, y) = \begin{cases} 1, & \text{if } x \in S,\ y \notin S \text{ or } x \notin S,\ y \in S; \\ 0, & \text{if } x \in S,\ y \in S \text{ or } x \notin S,\ y \notin S. \end{cases}$$

Write a linear program for finding coefficients λ_S for a given $V \subset \mathbb{R}^N$. The LP may have exponentially many variables.

(c) How many non-zero variables can an extreme solution to the LP have?

(d) Using the LP solution construct map φ.