BELLAIRS WORKSHOP ON APPROXIMATION ALGORITHMS
OPEN PROBLEM SESSION #1

(1) **Mohit Singh**. Mihalis Yannakakis (JCSS, 1991) proved that there is no compact symmetric extended formulation for matching and TSP.

Is there a result like this for approximation?

Conjecture. *For vertex cover, every compact extended symmetric formulation has integrality gap* $\geqslant 2 - \varepsilon$.

[SF: Volker Kaibel, Kanstantsin Pashkovich and Dirk Oliver Theis (IPCO 2010) proved that symmetry can force the size of *exact* extended formulations to be exponential, while polynomial size extended formulations exist.]

(2) **Bruce Shepherd**. *Assuming rational data, is the first Chvátal-Gomory closure of a compact convex set of a polyhedron?* Recently, Santanu Dey and Juan Pablo Vielma proved this for ellipsoids (IPCO 2010). The following older question is still open: Is the first Chvátal-Gomory closure of a non-rational polyhedron a polyhedron?

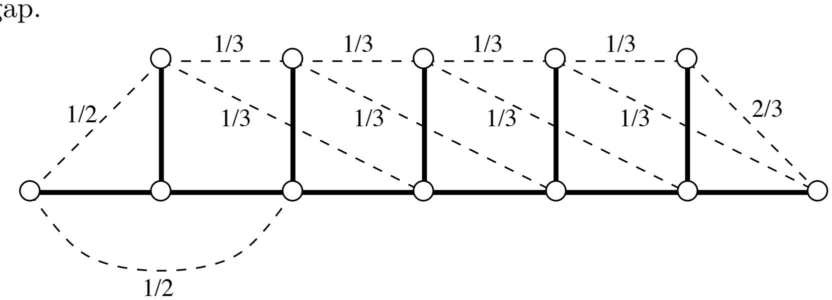
(3) **Howard Karloff**. Consider the following tree augmentation problem. We are given a spanning tree T in an undirected graph G. Find a minimum weight set of edges of $E(G) \setminus E(T)$ such that adding these edges to T gives a 2-edge connected spanning subgraph of G.

There are many ways to get a 2-approximation for the problem.

In the unweighted case, Guy Even, Jon Feldman, Guy Kortsarz and Zeev Nutov (APPROX 2001) obtained a 3/2-approximation. The original proof is long and difficult to read. [JK: Apparently, their latest proof is 15 pages long.]

Can you find a $(2 - \varepsilon)$*-approximation algorithm for the weighted case?*

The natural cut LP has an integrality gap of at least 3/2, as proved by the following instance. In the example the edges of $E(T)$ are the bold edges. The edges of $E(G) \setminus E(T)$ are the dashed edges. The values define a feasible fractional solution of the cut LP. A result of Joseph Cheriyan, Tibor Jordán and R. Ravi (ESA 1999) imply that inspecting half-integral solutions of the cut LP would not yield a better than 4/3 lower bound on the integrality gap.



(4) **Deeparnab Chakrabarty**. *Can we get better than logarithmic approximation for set cover instances where the set system has low (constant) VC dimension?* A case where this arises is in geometric settings. Given a set of rectangles and points on the plane, the (d-dimensional) rectangle cover problem asks for the minimum set of rectangles which covers every point.

One way to look at these problems might be via priority versions of the line/tree cover problem. In the d-priority line cover (PLC) problem, each edge and segment is associated a priority d-dimensional vector, and a segment now covers an edge if it contains it, and furthermore, the priority vector of the segment coordinate wise dominates that of the edge. The d-priority rooted tree cover is defined analogously where a rooted tree needs to be covered by segments going from child to ancestor.

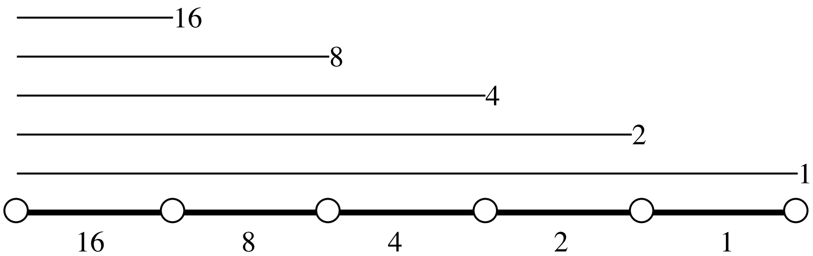
What can be shown is that 1-PLC is a special case of 1-PTC which is a special case of 2-PLC. What can also be shown is that 1-PLC is a special case of 2d rectangle cover, which is a special case of 2-PLC and so on. Complexity wise, 1-PLC is in P and 1-PTC is APX-hard. The latter implies 3 dimensional and higher rectangle cover is APX-hard. 1-PTC can be approximated to factor 2, but no other better than $O(\log n)$ approximation is known for anything more general, or the rectangle cover problem. Integrality gap wise, the 1-PLC has an integrality gap bounded by 2, and nothing more is known for any other problem.

[CC: For geometric set cover problems in 2d some interesting results are known. For example, there is an $O(1)$ approximation for covering points by disks (unweighted case). And an $O(\log\log n)$ approximation for covering by fat triangles. Recently Varadarajan (STOC 2010) generalized some of these results to the weighted case. These results are based on ϵ-nets which are in turn proved via union complexity bounds. The union complexity of a set of n discs in 2d is $O(n)$, while that of a set of n rectangles can be $\Omega(n^2)$ (think of a grid). Thus the geometric set cover ideas do not work for rectangles; note however they do work for square and rectangles of bounded aspect ratio.]

(5) **Chandra Chekuri**. This problem is a packing version of a problem we have seen in Jochen Könemann's talk. There is a path P, and capacities c_e on edges of the path. There are also n subpaths $P_1, \ldots, P_n$ of this path, each coming with its own weight/profit w_i and demand d_i. The goal is to pack the maximum profit subset of the segments while not exceeding the capacities. The natural LP relaxation for this problem is:

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} w_i x_i \\ \text{s.t.} \quad & \sum_{i:P_i \ni e} d_i x_i \le c_e \quad \forall e \in E(P) \\ & 0 \le x_i \le 1 \quad \forall i \in [n]. \end{aligned}$$

This LP has a $\Omega(n)$ integrality gap even for unit weights, as shown by the following example. (The path P is in bold. The capacities are indicated just below P. The other paths P_i are shown above P, together with their weights. Here, $OPT = 1$ while $LP \ge n/2$.) Remarks: t rounds of Sherali-Adams with the LP still leaves a $\Omega(n/t)$ integrality gap; If we make the no bottleneck assumption, that is, $\max_{e \in E(P)} c_e \le \min_{i \in [k]} w_i$, then the LP has a $O(1)$ integrality gap.



Nikhil Bansal *et al.* (SODA 2009) gave a combinatorial $O(\log n)$-approximation for the problem. Chandra Chekuri *et al.* (APPROX 2009) proved that combining a simple greedy algorithm with a partitioning trick from the paper of Bansal *et al.* gives an $O(\log^2 n)$ approximation. In addtion Chekuri *et al.* described a stronger LP relaxation with $O(\log n)$ integrality gap. The new LP has an inequality for each subset of the subpaths through a particular edge e; see the paper for more details.

Is the integrality gap of the new LP $O(1)$?

(6) **Vahab Mirrokni.** In the document clustering problem, we have set of documents $D_1, \ldots,$ D_n and k bins $S_1, \ldots, S_k$. Each document is seen as a set of words taken from a dictionary. We have to put the n documents into the k bins in order to minimize the maximum number of words present in any single bin. Regarding bins as subcollections of $\mathcal{D} := \{D_1, \ldots, D_n\}$, the goal is to determine

$$\min_{S_1,\ldots,S_k} \max_{i\in[k]} | \cup_{j\in S_i} D_j|.$$

The best we can do currently is a $O(\min\{k, \sqrt{n}\})$-approximation. *Can you do better?* Note that this approximation factor is also achievable even if we replace $| \cup_{j\in S_i} D_j|$ with $f(S_i)$ where f is any *monotone* submodular function on $\mathcal{D}$. However, for general monotone submodular functions, we cannot hope for a better approximation factor.