Approximating covering points by lines

Raymond Chen

May 28, 2020

1 Introduction

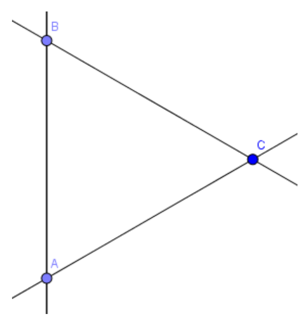
The **covering points by lines problem** seeks to minimize the number of lines needed such that every point is contained in at least one line. Formally, the problem can be written as follows:

Problem Statement: Given a set $A = \{a_1, a_2, ..., a_n\}$ of points in the two-dimensional plane, and a set $B = \{b_1, b_2, ..., b_m\}$ of lines, we want to minimize the number of lines needed to cover A.

Note that this problem is a specific instance of the set-coverage problem. It's a known fact that the set cover Linear Programming formulation has an integrality gap of $O(logn)$. Armed with this fact, I wanted to see if applying the set cover LP to the problem of covering a set of points by a set of lines could achieve a constant integrality gap. If so, it means that I might be able to use this LP to find a constant-factor approximation of the problem by rounding.

LP Formulation: Let y_j be a decision variable for picking line b_j. Then our goal is to minimize $\Sigma_{j=1}^{m} y_j$ subject to the constraint that for all a_i, $\Sigma_{j \in B_i} y_j \leq 1$, where B_i contains the indices of lines that contain point a_i.

Here's an **example** to show that the integrality gap is greater than 1:



In OPT, we need to pick two lines to cover all points. However, in the LP, we can let every decision variable be $\frac{1}{2}$, for a total of $\frac{3}{2}$. Thus, we can see from this simple example that the integrality gap is at least $\frac{4}{3}$. Extending this example to the k-clique, I saw that the integrality gap was $\frac{2(k-1)}{k}$ which approaches two as k increases.

I saw from Alon's [1] paper a construction that showed that a greedy algorithm for **covering lines by points** was $\Omega(logn)$ approximate. So I checked to see if this same construction could be used for the integrality gap. In fact, it did apply!

2 Alon's Paper

Covering lines by points is the dual of the covering points by lines, given a set $B = \{b_1, b_2, ..., b_m\}$ of lines, we want to minimize the number of points in $A = \{a_1, a_2, .., a_n\}$ such that every line is covered by a point.

The LP is constructed by decision variables x_i for picking point a_i. We want to minimize $\Sigma_{j=1}^{m} y_j$ given that for any line b_j, $\Sigma_{i \in A_j} x_i \leq 1$ where A_j is the set of indices of points that touch line b_j. Consider the same figure as before. Then we can see that assigning $\frac{1}{2}$ to every point covers every line for our LP. But OPT requires at least two points so our integrality gap is at least $\frac{4}{3}$.

In order to understand Alon's construction, I first need to introduce some notation. $[k]^d$ is the set of all d-dimensional coordinates with integer values between 1 and k.

A **combinatorial line** is a set of k points from $[k]^d$. Given a set $I \subset \{1, ..., d\} = \{c_1, c_2, ..., c_n\}$, Pick integers $x_{c_1}, ..., x_{c_n}$ from the set $\{1, ..., d\}$. Then the combinatorial line corresponds to the k points formed by fixing coordinate c_i to be x_{c_i}. For any other points set them all to the same value j between 1 and k. Thus, there are exactly k points in this combinatorial line.

As an example, consider in $[4]^4$, $I = \{1, 3\}$ with corresponding $2, 3$. Then our combinatorial line is the points:

$$(2, 1, 3, 1)$$
$$(2, 2, 3, 2)$$
$$(2, 3, 3, 2)$$
$$(2, 4, 3, 4)$$

Density Hales-Jewitt Theorem (Theorem 2.1 in Alon's Paper) says that given k, δ, there exists $d_0 = d_0(k, \delta)$ such that $\forall d \geq d_0$, any set of at least $\delta * k^d$ points in $[k]^d$ contains a combinatorial line.

Alon's construction We know that for any combinatorial line, there exists a unique line passing through all k points.

For any positive integer $k \geq 2$, let d be corresponding $d(k, \frac{1}{2})$ by Hales-Jewitt. Let A = $[k]^d$, and B be the set of all lines passing through combinatorial lines of A.

By Hales-Jewitt, we know that we know that any set of $\frac{k^d}{2}$ points contains a combinatorial line. And the complement of this set must also contain a combinatorial line, since it also has size $\frac{k^d}{2}$.

Thus, OPT must be at least $\frac{k^d}{2}$, since no set with less than $\frac{k^d}{2}$ points can contain all combinatorial lines. However, we can assign $\frac{1}{k}$ to every point in the LP, since every combinatorial line contains exactly k points. So our LP is at most k^{d-1}

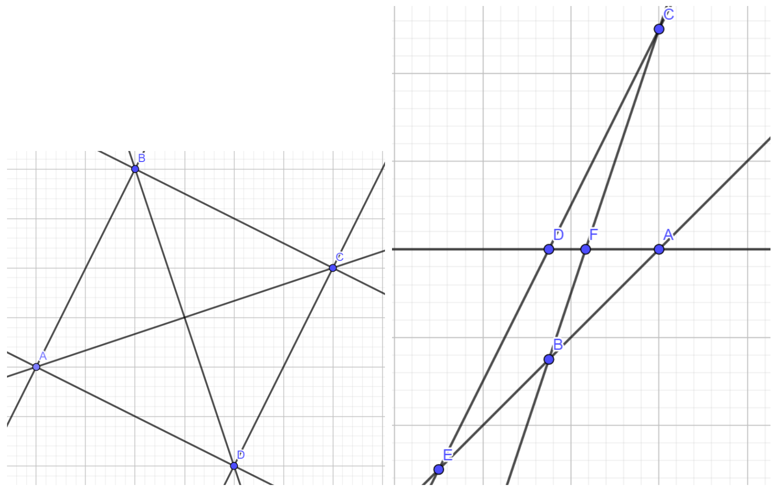
Thus, our integrality gap is at least $\frac{k}{2}$. Lemma 2.2 in [1] shows how to project this example onto the 2-D plane.

Since this holds true for all k, the integrality gap cannot be upper bounded by a constant.

3 Relating it back

So Alon's paper shows that the LP for covering lines by points can't be upper bounded by a constant. However, it didn't answer my original question about whether the LP for covering points by lines could be. That's when I came across this transformation that convert covering lines by points to covering points by lines[2].

For any line $y = mx + c$, transform it to the point $(m, -c)$. For any point (a, b) transform it to the line $y = ax + b$. Note that we can apply this transformation as long as there are no vertical lines. If there are, then we can simply rotate our plane.



(a) Covering lines by points (b) Covering points by lines

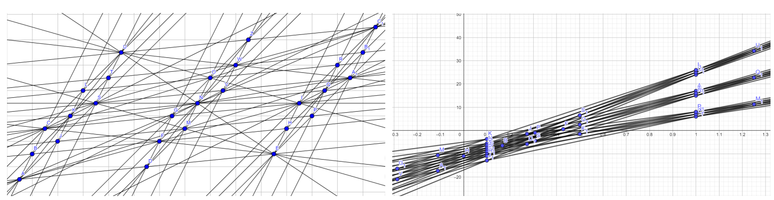
Example transformation of the 4-clique

This transformation has the property that n lines intersecting in Hitting Set are transformed to n colinear points in covering points by lines [2].

Moreover, if a line contains k points in the original image, k lines pass through its corresponding point in the transformation [k].

So the integrality gap from Hitting Set is preserved in this transformation.

So, I ended up proving my original question, but I still wanted to construct a concrete example with an integral gap greater than 2. That is, a worse gap than the cliques. However, Alon's example is very abstract and difficult to concretize. The following example corresponds to an integral gap of $\frac{3}{2}$.


(c) Covering lines by points (d) Covering points by lines

To begin with, there was no deterministic method to get the Hales-Jewitt number. Moreover, to show an integral gap of say 2.5, we would need to project the hyper lattice $[5]^d$ for some number d, which would require thousands of points.

References

[1] N. Alon, A non-linear lower bound for planar epsilon-nets; https://www.tau.ac.il/ nogaa/PDFS/epsnet1.pdf

[2] D. Mount, Computational Geometry, http://graphics.stanford.edu/courses/cs268-16-fall/Notes/cmsc754-lects.pdf, 43-44.