An Approximation Algorithm for a Geometric Set Covering Problem

Sofia Kovaleva, Frits Spieksma
Maastricht University, Department of Mathematics, NL-6212 CA Maastricht
s.kovaleva@math.unimaas.nl, spieksma@math.unimaas.nl

Abstract

We consider a special set covering problem. The problem is a natural generalization of finding a minimum clique cover in an interval graph. We present a rounding algorithm that delivers a solution with a value of at most $2 + \epsilon$ times the value of an optimal solution for any $\epsilon > 0$.

1 Introduction

Given an integer k, and given an $n \times \infty$ grid with on each row at most k intervals, each of arbitrary integral length and placed at an arbitrary position, consider the following question.

What is the minimum number of rows and columns needed to stab all intervals?

It is not hard to see that this problem is a special case of the well-known set covering problem: let the intervals correspond to elements in the ground set, and let a set of intervals that are on a same row or a set of intervals that share a column correspond to a set in the collection. We will refer to this special geometric set covering problem as GEOSCk.

Applications. The following type of situation leads naturally to instances of GEOSCk. Each of n items (patients to receive treatment, products to undergo chemical processes, machines subject to inspection) has to undergo treatment on a regular basis. More precisely, for each item a set of intervals is given during which a treatment must take place. The treatment itself takes one time-unit and is provided by some kind of machine with unbounded capacity (that is, it can process any number of items), and consists of "turning the machine on" at some point in time, say t (this corresponds to selecting column t). Then the items corresponding to intervals that are stabbed by column t undergo the treatment. The objective is to minimize (a weighted combination of) the number of times the machine is turned on plus the number of items not processed (an item is not processed when at least one of its intervals has not undergone the treatment (this corresponds to selecting the row corresponding to that item)).

GEOSCk as a graph-theoretical problem. The question above can be formulated as a graph-theoretical problem. Indeed, given the grid with its intervals, construct a graph as follows. There is a node for each interval, and two nodes are connected if they share a column of the grid (blue edge) or if they are on the same row of the grid (red

edge). Notice that an edge can be red as well as blue in case two intervals of a same row share a column (if one finds bicolored edges awkward, one could alternatively work with parallel edges of different colors). Thus the graph constructed is the edge-union of an interval graph (the blue edges) and a k-partite graph. The covering question is now equivalent to finding a monochromatic minimum clique cover. Notice that in case $k = 1$ the covering question reduces to finding a minimum clique cover in an interval graph. Since intervals graphs are perfect (see for instance [2]) this number can be found in polynomial time.

Related results. This problem has been investigated in [3]. It is shown that the problem is MAX SNP-hard for each $k \geq 2$, and a $\frac{3}{2}$-approximation algorithm in case $k = 2$ is given.

Our results. In Section 2 we show that there is an LP-based algorithm that achieves a $2 + \epsilon$-approximation for GEOSCk for any k and for any $\epsilon > 0$.

2 An approximation algorithm for GEOSCk

In this section we describe an approximation algorithm for GEOSCk. Observe that by duplicating intervals one can assume wlog that instances of GEOSCk have precisely k intervals per row. Consider now the following {0,1} programming formulation for GEOSCk. Let f_l (s_l) be the last (first) column of the grid for which interval l is active, $l = 1, \ldots, kn$. We use y-variables for the columns and z-variables for the rows, where a variable is 1 iff the respective column, row is selected. Notice that we can restrict ourselves to using only those y-variables for which some interval ends in that column. Let us assume (wlog) that there are kn y-variables and let us further assume that they are ordered by nondecreasing f_l, $l = 1, \ldots, kn$. Let $S(l) = \{j : s_l \leq f_j \leq f_l\}$, and let interval l be on row $r(l)$, $l = 1, \ldots, kn$.

$$
\begin{array}{lll}
(IP)\ \text{Minimize} & \sum_{l=1}^{kn} y_l + \sum_{i=1}^{n} z_i & \\
\text{subject to} & z_{r(l)} + \sum_{j \in S(l)} y_j \geq 1 & \text{for all } l = 1, \ldots, kn, \\
& \text{all variables } \in \{0, 1\}. &
\end{array}
$$

Now, given an instance I of GEOSCk, let $OPT(I)$ be the value of the corresponding optimal solution, let $v_{LP}(I)$ be the value of the linear programming relaxation and let the corresponding solution of the linear program be described by $\hat{y}_l, \hat{z}_{r(l)}$. In addition to I let us also assume that some $\epsilon > 0$ has been specified, and let N be a - yet unspecified - large number.

We use an idea described in [1] to obtain the desired result.

Informally, the algorithm works as follows. Given the $\hat{y}_l$ and the $\hat{z}_{r(l)}$ values, round each of these fractional variables up to the nearest multiple of N, so that $\tilde{y}_j = a_l/N$ for some integral a_l, $l = 1, \ldots, kn$, and $\tilde{z}_j = b_j/N$ for some integral b_j, $j = 1, \ldots, n$.

In each column l where some interval ends, draw a_l vertical lines $l = 1, \ldots, kn$. Similarly, draw b_j horizontal lines in each row j, $j = 1, \ldots, n$.

Consider now the following question: is it possible to color all lines using $\frac{N}{2}$ colors such that each line receives exactly one color, and such that each color class (that is, all lines with a same color) stabs each interval? Observe that if the answer is yes, there would be, for each interval and for each color, a line with that color stabbing that interval; in other words, each interval is stabbed by each color and hence each set of lines corresponding to a color class is a feasible solution. The following proposition answers this question in the affirmative:

Proposition: There exists a feasible coloring using $\frac{N}{2}$ colors.
Proof: Let us first consider the vertical lines. Index these lines according to the index of the column they belong to in non-decreasing order. Then give line j color (j mod $\frac{N}{2}$), $j = 1, \ldots$. Apply a similar approach to the horizontal lines: index these lines according to the index of the row they belong to in non-decreasing order. Then give line j color (j mod $\frac{N}{2}$), $j = 1, \ldots$. Let us now argue why this is a feasible coloring. Consider some interval l. This interval is stabbed by columns of the set $S(l)$ or alternatively it is stabbed by $\sum_{j \in S(l)} a_j$ lines. If this number is equal to or larger than $\frac{N}{2}$ then we're done since the coloring approach described above ensures that $\frac{N}{2}$ vertical lines each with a different color stab interval l. Otherwise, we know that there are at least $\frac{N}{2}$ horizontal lines stabbing interval l (since $\tilde{y}_j, \tilde{z}_{r(l)}$ correspond to a feasible LP-solution) which implies by the coloring procedure above that interval l is stabbed by at least $\frac{N}{2}$ horizontal lines each with a different color. **End of proposition**

Thus, by the proposition above there are $\frac{N}{2}$ solutions which have a total value of at most $Nv_{LP} + (k+1)n/N$ (where the last term comes from rounding up the $\tilde{y}_j, \tilde{z}_{r(l)}$ values). Setting $N := \sqrt{\frac{2(k+1)n}{\epsilon}}$, it follows that there is at least one solution that has a value bounded by $2 + \epsilon$ times $OPT(I)$.

Finally, observe that this approach is valid for integral right-hand sides in the formulation (IP) and also for a weighted objective function.

References

[1] Bar-Noy, A., S. Gupta, J. Naor and B. Schieber (1999), "Approximating the Throughput of Multiple Machines in Real-Time Scheduling", Proceedings of the 31st ACM Symposium on Theory of Computing.

[2] Golumbic, M.C, *Algorithmic Graph Theory and Perfect Graphs*, Academic Press, San Diego, California, 1980.

[3] Spieksma, F.C.R. (1999), private communication.