Hitting Set, Spanning Trees, and the Minimum Length Corridor Problem

Tomás Feder* Carlos Subi†

December 23, 2008

1 Introduction

We consider a problem that is a combination of hitting set and minimum spanning trees that generalizes the minimum length corridor problem. In the *minimum length corridor problem*, we are given a rectangle aligned with the axes in the plane that is subdivided into smaller rectangles also aligned with the axes in the plane. This defines a planar graph whose vertices are the corners of the rectangles and whose edges are segments of the sides joining them, where the edges have edge lengths. The aim is to find a subtree of this graph that touches all the faces, including the outer face, in at least one vertex, and of minimal total edge length.

The corridor problem is known to be NP-complete [2, 3], and constant factor approximable [1]. If the planar graph has k faces corresponding to rectangles and n vertices, then $k \leq n \leq 2k$. We give a framework for a more general problem that infers only a polynomial time approximation factor of $O(\sqrt{n \log n})$ for this problem.

The approach is based on consider the following more general *combined problem*, that combines hitting set and spanning trees. An instance consists of a clique on n vertices x_i, $1 \leq i \leq n$, where the edges $e_{ij} = (x_i, x_j)$ have distances $d_{ij} \geq 0$ satisfying the triangle inequality $d_{ij} + d_{j\ell} \leq d_{i\ell}$. An instance also has k sets of vertices C_i or *colors*, $1 \leq i \leq k$. Notice that a vertex may have any number of colors. The aim in the *combined problem* is to find a minimum total distance tree in the clique that spans vertices so that every color C_i is hit by the tree at some vertex.

The combined problem is shown to be NP-complete even if $d_{ij} = 1$ for all $i \neq j$. The reduction is from minimum hitting set. In fact, the combined problem is essentially equivalent to hitting set, and therefore $\Theta(\log k)$ easyness and hardness approximation factors follow.

We then show that the combined problem can be approximated within a factor of $\min(k-1, 2\sqrt{n \log k})$. The corridor problem can be viewed as a special case of the combined problem where the rectangles are the colors C_i that are sets of vertices to be touched by the tree, and distances among vertices are defined in the corresponding planar graph accordingly. Since $k \leq n \leq 2k$, we obtain an $O(\sqrt{n \log n})$ approximation algorithm for the corridor problem.

*268 Waverley St., Palo Alto, CA 94301. E-mail: `tomas@theory.stanford.edu`.
†E-mail: `carlos subi@hotmail.com`.

2 Easyness and Hardness Bounds for Combined Problem with $d_{ij} = 1$

In this section we consider the combined problem with $d_{ij} = 1$.

Theorem 1 *Consider the combined problem with $d_{ij} = 1$ for all vertices x_i, x_j, $i \neq j$. The problem is NP-complete, in fact $\Omega(\log k)$ approximation hard, and $O(\log k)$ easy to approximate, as the problem is essentially equivalent to the hitting set problem.*

PROOF: If we hit the k colors C_i, $1 \leq i \leq k$ at $r + 1$ vertices, we have a spanning tree of total edge distance r. Thus the problem is as easy and as hard as the hitting set problem, and the result follows. ■

3 $\min(k - 1, 2\sqrt{n \log k})$ Approximation Bound for the General Combined Problem

In this section we prove the main result of this paper.

Theorem 2 *The general combined problem has a polynomial time algorithm that achieves a $\min(k - 1, 2\sqrt{n \log k})$ approximation bound.*

PROOF: Consider the tree T in the optimal solution, and let u, v be the two farthest points in this tree. If we fix u, v, then a lower bound to the solution will be $L = d(u, v)$, and for the $k - 2$ colors not covered by u, v, we can consider the minimum edge joining them to u, of edge length at most L, for a total edge length at most $(k - 1)L$, proving the $k - 1$ approximation bound.

Now set $S = \sqrt{n/\log k}$. Shrink all the edges of edge length at most L/S and potentially include them in the solution. This will cost at most nL/S for the corresponding forest of shrunk edge, at most a factor n/S over the lower bound L. It remains to hit the k colors with shrunk vertices, including the shrunk vertices for u and v. If we can hit with r shrunk vertices, then a lower bound is rL/S. A hitting set algorithm will hit $r \log k$ vertices, all at distance at most L from u, for an upper bound of $r(\log k)L$, at most a factor $S \log k$ over the lower bound rL/S. We thus achieve a factor of $(n/S) + S \log k = 2\sqrt{n \log k}$. ■

4 Application to the Corridor Problem

We first observe how large the distance of a solution can be in the corridor problem.

Proposition 1 *Suppose the corridor problem has an outer rectangle that is a square of side lengths equal to 1. Then the maximum possible solution has cost $\Theta(k) = \Theta(n)$, where $k \leq n \leq 2k$.*

PROOF: As each of the k rectangles has perimeter at most 4, we obtain a bound of $4k$ for the total perimeter, thus $O(k) = O(n)$ is the bound for the total distance in the tree.

For the lower bound, divide the rectangle into k' horizontal strips of equal height, and alternatingly divide the strips into two or three rectangles of by vertical edges. Thus $k = O(k')$, and each of the middle rectangles of the alternating three rectangle strips, will require an edge of length at least $1/6$ within two edge steps, proving a lower bound on the total edge length that is $\Omega(k) = \Omega(n)$. ■

We finally give the first approximation result known for the corridor problem, as a corollary from Theorem 2.

Corollary 1 *The minimum length corridor problem can be approximated within* $O(\sqrt{n \log n}) = O(\sqrt{k \log k})$.

References

[1] A. Gonzalez-Gutierrez and T.F. Gonzalez, Approximation algorithms for the minimum-length corridor and related problems, TR UCSB 2007–03.

[2] H.L. Bodlanender, C. Feremans, A. Grigoriev, E. Penninkx, R. Sitters, and T. Wolle, On the minimum corridor connection and other generalized geometric problems, In 4th Workshop on Approximation and Online Algorithms (WAOA) (Zurich, Switzerland, September 2006), T. Erlebach and C. Kaklamanis, Eds., vol. 4368 of Lecture Notes in Computer Science, Springer, pp. 69–82.

[3] A. Gonzalez-Gutierrez, and T.F.Gonzalez, Complexity of the minimum-length corridor problem. Comput. Geom. Theory Appl. 37, 2 (2007) 72–103.

[4] Garey and Johnson.