Guillotine subdivisions approximate polygonal subdivisions: Part III – Faster polynomial-time approximation schemes for geometric network optimization*

Joseph S. B. Mitchell†

April 19, 1997; Last revision: May 29, 1997

1 Introduction

In this note, we show how a modification to our earlier results on guillotine subdivisions leads to an $n^{O(1)}$ time (deterministic) PTAS for Euclidean versions of various geometric network optimization problems on a set of n points in the plane. This improves the previous $n^{O(1/\epsilon)}$ time algorithms of Arora [1] and Mitchell [10]. Arora [2] has recently obtained even better bounds than our own: He obtains a randomized algorithm with expected running time $O(n \log^{O(1/\epsilon)} n)$. However, our new theorem on approximating guillotine subdivisions may be interesting in its own right, and may lead to yet further improvements or generalizations.

Our methods are based on the concept of an "m-guillotine subdivision", which were introduced by Mitchell [9, 10]. Roughly speaking, an "m-guillotine subdivision" is a polygonal subdivision with the property that there exists a line ("cut"), whose intersection with the subdivision edges consists of a small number ($O(m)$) of connected components, and the subdivisions on either side of the line are also m-guillotine. The upper bound on the number of connected components allows one to apply dynamic programming to optimize over m-guillotine subdivisions, as there is a succinct specification of how subproblems interact across a cut.

Key to our method is a theorem showing that any polygonal subdivision can be converted into an appropriate m-guillotine subdivision by adding a set of edges whose total length is small: at most $\frac{c}{m}$ times that of the original subdivision (where $c = 1, \sqrt{2}$, depending on the metric). Key to our improvement over previous results on approximating with guillotine subdivisions is the notion of a "grid-rounded" m-guillotine subdivision, in which each connected component is also required to contain one of a small number of regularly spaced grid points. (These notions are made precise in the next section.) Then, exactly as in [10], we use dynamic programming to optimize over an appropriate class of m-guillotine subdivisions, resulting in, for any fixed m, $(1 + \frac{c}{m})$-approximation algorithms that run in polynomial-time ($O(n^{O(1)})$), for various network optimization problems.

Related Work. There has been an abundance of work on the problems studied here, both on instances of the problems in graphs and on geometric instances. We refer the reader to some standard textbooks, such as [4, 6, 12]. For the particular problem of the TSP, there is a survey book edited by Lawler et al. [7], and for results on approximation theory and algorithms, there is the recent book edited by Hochbaum [5].

*Manuscript available at http://ams.sunysb.edu/~jsbm/jsbm.html

†`jsbm@ams.sunysb.edu`, Dept. of Applied Mathematics and Statistics, State University of New York, Stony Brook, NY 11794-3600. Supported in part by Hughes Research Laboratories and NSF Grant CCR-9504192.

While the geometric optimization problems considered here are known to be NP-hard, polynomial-time approximation algorithms have been known previously that get within a constant factor of optimal. Further, polynomial-time approximation schemes were discovered last year, by Arora [1] and by Mitchell [9, 10].

This paper represents a continuation of our previous work on guillotine subdivisions ([9, 10, 11]), which in turn is based on the concept of "division trees" introduced by Blum, Chalasani, and Vempala [3, 11], and the guillotine *rectangular* subdivision methods of Mata and Mitchell [8]. Here, we obtain substantially better bounds than before, and we generalize our previous results to apply to a much broader class of problem instances that include those weighted graphs whose edge lengths correspond to shortest path lengths among obstacles in the plane.

2 Grid-Rounded m-Guillotine Subdivisions

Definitions

We follow most of the notation of [9, 10]. We consider a polygonal subdivision ("planar straight-line graph") S that has n edges (and hence $O(n)$ vertices and facets). Let E denote the union of the edge segments of S, and let V denote the vertices of S. We can assume (without loss of generality) that S is restricted to the unit square, B (i.e., $E \subset int(B)$). Then, each facet (2-face) is a bounded polygon, possibly with holes. The *length* of S is the sum of the lengths of the edges of S. Assume, without loss of generality, that no two vertices of S have a common x- or y-coordinate.

A closed, axis-aligned rectangle W is a *window* if $W \subseteq B$. In the following definitions, we fix attention on a given window, W. We let E_W denote the subset of E consisting of the union of segments of E having at least one endpoint inside (or on the boundary of) W. The *combinatorial type* (*with respect to* E) of a window W is an ordered listing, for each of the four sides of W, of the identities of the line segments of E_W that intersect it. We say that W is *minimal* (with respect to E) if there does not exist a $W' \subset W$, strictly contained in W, having the same combinatorial type as W. By standard critical placement arguments, we see that, since it has four degrees of freedom, a minimal window is determined by four *contact pairs*, defined by a vertex $v \in V$ in contact with a side of W or by a corner of W in contact with a segment of E_W. (Thus, there are at most $O(n^4)$ minimal windows.) For any window W containing at least one vertex of E, we let $\overline{W}$ denote a minimal window, contained within W, having the same combinatorial type as W. (At least one such $\overline{W}$ must exist.)

A line ℓ is a *cut* for E (with respect to W) if $\ell \cap int(W) \neq \emptyset$. The intersection, $\ell \cap (E \cap int(W))$, of a cut ℓ with $E \cap int(W)$ (the restriction of E to the window W) consists of a discrete (possibly empty) set of subsegments of ℓ. (Some of these "segments" may be single points, where ℓ crosses an edge.) The endpoints of these subsegments are called the *endpoints along* ℓ (with respect to W). (The two points where ℓ crosses the boundary of W are not considered to be endpoints along ℓ.) Let ξ be the number of endpoints along ℓ, and let the points be denoted by $p_1, \ldots, p_\xi$, in order along ℓ.

For a positive integer m, we define the *m-span*, $\sigma_m(\ell)$, of ℓ (with respect to W) as follows. If $\xi \leq 2(m-1)$, then $\sigma_m(\ell) = \emptyset$; otherwise, $\sigma_m(\ell)$ is defined to be the (possibly zero-length) line segment, $p_m p_{\xi-m+1}$, joining the mth endpoint, p_m, with the mth-from-the-last endpoints, $p_{\xi-m+1}$.

Given a line segment $\sigma = pq$ ($p \neq q$) and a positive integer M, consider the set of subsegments obtained by cutting pq into M equal-length segments; we define the *M-grid* of $\sigma = pq$ to be the set of $M+1$ endpoints of these subsegments. (In particular, the M-grid contains the two points p and q.)

A line ℓ is an *(m, M)-perfect cut with respect to* W if $\sigma_m(\ell) \subseteq E$, *and* each connected component of $\ell \cap E$ contains an M-grid point of the 1-span, $\sigma_1(\ell)$. In particular, if $\xi \leq 2(m-1)$, then ℓ is trivially an (m,M)-perfect cut (since $\sigma_m(\ell)=\emptyset$). Similarly, if $\xi=2m-1$, then ℓ is m-perfect (since $\sigma_m(\ell)$ is a single point). Otherwise, if ℓ is m-perfect, and $\xi \geq 2m$, then $\xi=2m$.

In the remainder of this paper, we fix $M=m(m-1)$ and assume that $m \geq 2$.

Finally, we say that S is a *grid-rounded m-guillotine subdivision with respect to window W* if either (1) $V \cap int(W)=\emptyset$; or (2) there exists an (m,M)-perfect cut, ℓ, with respect to a minimal window, $\overline{W} \subseteq W$, such that S is grid-rounded m-guillotine with respect to windows $W \cap H^+$ *and* $W \cap H^-$, where H^+, H^- are the closed halfplanes induced by ℓ. (Note that, since $\overline{W}$ is minimal, necessarily windows $W \cap H^+$ and $W \cap H^-$ will each have a combinatorial type different from that of W.) We say that S is a *grid-rounded m-guillotine subdivision* if S is grid-rounded m-guillotine with respect to the unit square, B.

The Approximation Theorem

The theorem below shows that grid-rounded m-guillotine subdivisions can approximate arbitrary subdivisions arbitrarily closely (as a function of m). Its proof directly follows that of [9, 10], with relatively minor changes to incorporate the concept of (m,M)-perfect cuts, which allow us to strengthen the requirements from that of m-perfect cuts to include the effect of rounding to the M-grid of the 1-span.

Theorem 1 *Let S be a polygonal subdivision, with edge set E, of length L. Then, for any positive integer m, there exists a grid-rounded m-guillotine subdivision, S_G, of length at most $(1+\frac{2\sqrt{2}}{m})L$ whose edge set, E_G, contains E.*

Proof. We will convert S into a grid-rounded m-guillotine subdivision S_G by adding to E a new set of horizontal/vertical edges whose total length is at most $\frac{2\sqrt{2}}{m}L$. The construction is recursive: at each stage, we show that there exists a cut, ℓ, with respect to the current window W (which initially is the box B), such that we can afford to add the following set of segments to E:

- *("red" segment)* the m-span, $\sigma_m(\ell)$; and
- *("blue" segments)* a line segment on ℓ connecting each of the endpoints of $\ell \cap (E \cup \sigma_m(\ell))$ to a point of the M-grid of $\sigma_1(\ell)$.

By construction, once we add these segments to E, ℓ becomes an (m,M)-perfect cut with respect to W. The sense in which we can "afford" to add these segments is that we can charge off the lengths of the constructed segments to a portion of the length of the original edge set, E.

First, note that if an (m,M)-perfect cut (with respect to W) exists, then we can simply use it, and proceed, recursively, on each side of the cut. Thus, we assume that no (m,M)-perfect cut exists with respect to a given window, W.

We say that a point p on a cut ℓ is *m-dark with respect to ℓ and W* if, along $\ell^- \cap int(W)$, there are at least m endpoints (strictly) on each side of p, where ℓ^- is the line through p and perpendicular to ℓ.[1] We say that a subsegment of ℓ is *m-dark* (with respect to W) if all points of the segment are m-dark with respect to ℓ and W.

The important property of m-dark points along ℓ is the following: Assume, without loss of generality, that ℓ is horizontal. Then, if all points on subsegment pq of ℓ are m-dark, then we can charge the length of pq off to the bottoms of the first m subsegments, $E^+ \subseteq E$, of edges that lie above pq, and the tops of the first m subsegments, $E^- \subseteq E$, of edges that lie below pq (since we know that there are at least m edges "blocking" pq from the top/bottom of W). We charge pq's length half to E^+ (charging each of the m levels of E^+ *from below*, with $\frac{1}{2m}$ units of charge) and

[1] We can think of the edges E as being "walls" that are not very effective at blocking light — light can go through $m-1$ walls, but is stopped when it hits the mth wall; then, p on a line ℓ is m-dark if p is not illuminated when light is shone in from the boundary of W, along the direction of $\ell^\perp$.

half to E^- (charging each of the m levels of E^- *from above*, with $\frac{1}{2m}$ units of charge). We refer to this type of charge as the "red" charge.

In addition to charging off the length of the m-dark portion of ℓ, in order to round to the M-grid of $\sigma_1(\ell)$, we are also going to charge off $(1/m)$th of the 1-dark portion of ℓ: If pq is 1-dark, then we charge $(1/m)$th of pq's length, by charging half of this length (i.e., $(1/2m)$th of the length of pq) off to the level of E that lies above pq, and half of it to the level of E that lies below pq. We refer to this type of charge as "blue" charge.

The *chargeable length* of a cut ℓ is defined to be the length of the m-dark portion of ℓ, plus $(1/m)$ times the length of the 1-dark portion of ℓ.

The *cost* of a cut, ℓ, is defined to be the length of the segments we must add to make the cut (m, M)-perfect. Thus, the cost of a cut ℓ is at most the length, $|\sigma_m(\ell)|$, of the m-span "red" segment, $\sigma_m(\ell)$, plus the lengths of the "blue" segments on ℓ connecting each of the endpoints of $\ell \cap (E \cup \sigma_m(\ell))$ to a point of the M-grid of $\sigma_1(\ell)$. Since there are at most $2m$ endpoints of $\ell \cap (E \cup \sigma_m(\ell))$, and two of these (the endpoints of $\sigma_1(\ell)$) are already at M-grid points of $\sigma_1(\ell)$, the total number of blue segments is at most $2m - 2$. Further, each blue segment is at most one half of $\frac{|\sigma_1(\ell)|}{M}$, where $|\sigma_1(\ell)|$ is the length of the 1-span of ℓ. Thus, the overall cost of a cut ℓ is at most

$$|\sigma_m(\ell)| + (2m-2) \cdot \frac{1}{2} \cdot \frac{|\sigma_1(\ell)|}{M} = |\sigma_m(\ell)| + \frac{1}{m}|\sigma_1(\ell)|,$$

where we have used our choice o $M = m(m-1)$.

We call a cut ℓ *favorable* if the chargeable length of $\ell \cap W$ is at least as long as the cost of the cut.

The lemma below shows that a favorable cut always exists. For a favorable cut ℓ, we add its m-span to the edge set (charging off its length, as above), and recurse on each side of the cut, in the two new windows. After a portion of E has been charged *red* on one side, due to a cut ℓ, it will be within m levels of the boundary of the windows on either side of ℓ, and, hence, within m levels of the boundary of any future windows, found deeper in the recursion, that contain the portion. Thus, *no portion of E will ever be charged* red *more than once from each side, in each of the two directions (horizontal/vertical)*, so no portion of E will ever pay more than *twice* its total length, times $1/m$, in red charge ($\frac{1}{2m}$ from each side, for each of the two directions). Similarly, no portion of E will ever be charged *blue* more than once from each side, in each of the two directions, and when it is charged blue, it is charged at the rate of only $1/2m$ per unit length (per side, per direction); thus, no portion of E will ever pay more than its total length, times $2/m$, in blue charge.

So far, this charging scheme gives rise to a total charge of at most $\frac{4}{m}L$. This factor can be improved slightly by noting that each side of an inclined segment of E may be charged red (resp., blue) *twice*, once vertically and once horizontally, so the red (resp., blue) charge assigned to a segment is at most $\frac{1}{m}$ times the sum of the lengths of its x- and y-projections, i.e., at most $\frac{\sqrt{2}}{m}$ times its length. This gives the overall charge of $\frac{2\sqrt{2}}{m}L$, as claimed.

It is also important to note that we are always charging portions of the original edges set E: The new edges added are never themselves charged, since they lie on window boundaries and cannot therefore serve to make a portion of some future cut m-dark or 1-dark.

(Note too that, in order for a cut ℓ to be favorable, but not (m, M)-perfect, there must be at least one vertex of V in each of the two open halfplanes induced by ℓ; thus, the recursion must terminate in a finite number of steps.) $\square$

We now prove the lemma that guarantees the existence of a favorable cut. The proof of the lemma uses a particularly simple argument, based on elementary calculus (reversing the order of integration). It is based on the similar lemma that appears already in [9, 10], but we include its details here for completeness:

Lemma 1 *For any subdivision S, and any window W, there is a favorable cut.*

Proof. We show that there must be a favorable cut that is either horizontal or vertical.

Let $f(x)$ denote the cost of the vertical line, ℓ_x, through x; then,

$$f(x) = |\sigma_m(\ell_x)| + \frac{1}{m}|\sigma_1(\ell_x)|.$$

Then, $A_x = \int_0^1 f(x)dx$ is simply the area, $A_x^{(m)} = \int_0^1 |\sigma_m(\ell_x)|dx$, of the ($x$-monotone) region $R_x^{(m)}$ of points of B that are m-dark with respect to horizontal cuts, plus $(1/m)$ times the area, $A_x^{(1)} = \int_0^1 |\sigma_1(\ell_x)|dx$, of the ($x$-monotone) region $R_x^{(1)}$ of points of B that are 1-dark with respect to horizontal cuts. Similarly, define $g(y)$ to be the cost of the horizontal line through y, and let $A_y = \int_0^1 g(y)dy$.

Assume, without loss of generality, that $A_x \geq A_y$. We claim that there exists a horizontal favorable cut; i.e., we claim that there exists a horizontal cut, ℓ, such that its chargeable length (i.e., length of its m-dark portion, plus $(1/m)$ times the length of its 1-dark portion) is at least as large as the cost of ℓ ($|\sigma_m(\ell)| + \frac{1}{m}|\sigma_1(\ell)|$). To see this, note that A_x can be computed by switching the order of integration, "slicing" the regions $R_x^{(m)}$ and $R_x^{(1)}$ horizontally, rather than vertically; i.e., $A_x = \int_0^1 h(y)dy = \int_0^1 h_m(y)dy + \frac{1}{m}\int_0^1 h_1(y)dy$, where $h(y)$ is the chargeable length of the horizontal line through y, and $h^{(i)}(y)$ is the length of the intersection of $R_x^{(i)}$ with a horizontal line through y. (i.e., $h^{(m)}(y)$ (resp., $h^{(1)}(y)$) is the length of the m-dark (resp., 1-dark) portion of the horizontal line through y.) Thus, since $A_x \geq A_y$, we get that $\int_0^1 h(y)dy \geq \int_0^1 g(y)dy \geq 0$. Thus, it cannot be that for all values of $y \in [0,1]$, $h(y) < g(y)$, so there exists a $y = y^*$ for which $h(y^*) \geq g(y^*)$. The horizontal line through this y^* is a cut satisfying the claim of the lemma. (If, instead, we had $A_x \leq A_y$, then we would get a *vertical* cut satisfying the claim.) □

Algorithms

The dynamic programming algorithms of Mitchell [9, 10] carry over almost verbatim to the new setting of grid-rounded m-guillotine subdivisions. The main difference is in the complexity analysis.

A subproblem in the dynamic programming recursion is specified now by a rectangle ($O(n^4)$ choices), and, on each of the four sides, a segment corresponding to the 1-span ($O(n^2)$ choices per side), and a set of up to $2m$ M-grid points within each segment that specify the attachment points between this subproblem and neighboring subproblems. (Depending on the problem instance, other information, of constant size for fixed m, is also specified for a subproblem; see [10].) The key to the improvement given in this paper is that there are now only $\binom{M}{2m} = O(m^{4m})$ choices for these grid points on any one side, and this number is constant for fixed m. (Compare this to the $O(n^{2m})$ choices of crossing points in [10].) Thus, there are overall $O(n^{12})$ subproblems. We then optimize over all $O(n)$ choices of cuts, $O(n^2)$ choices of 1-spans along the cut, and $O(m^{4m})$ choices of grid points on the cut. The overall complexity of the dynamic programming algorithm is therefore $O(n^{15})$. By rounding the 1-span intervals up to be intervals of lengths that are power-of-two factors smaller than the dimensions of the window, it is not hard to improve this complexity to $O(n^{10}\log^5 n)$, without significantly changing the approximation factor.

Corollary 1 *Given any fixed positive integer m, and any set of n points in the plane, there is an $O(n^{O(1)})$ algorithm to compute a Steiner spanning tree (or Steiner k-MST), or a traveling salesperson tour, whose length is within a factor $(1+\frac{c}{m})$ of minimum, for constant c.*

Acknowledgements

I thank Avrim Blum and Santosh Vempala for useful discussions on the subject of this paper.

References

[1] S. Arora, *Polynomial time approximation schemes for Euclidean TSP and other geometric problems*, Manuscript, March 30, 1996. Appears in Proc. 37th Annu. IEEE Sympos. Found. Comput. Sci.(1996), pp. 2–12.

[2] S. Arora, *More efficient approximation schemes for Euclidean TSP and other geometric problems*, Unpublished manuscript, January 9, 1997.

[3] A. Blum, P. Chalasani, and S. Vempala, *A constant-factor approximation for the* k-MST *problem in the plane*, Proc. 27th Annu. ACM Sympos. Theory Comput. (1995), pp. 294–302.

[4] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, The MIT Press, Cambridge, Mass., 1990.

[5] D. Hochbaum, editor, *Approximation Problems for NP-Complete Problems*, PWS Publications, 1996.

[6] E. Lawler, *Combinatorial Optimization: Networks and Matroids*, Holt, Rinehart and Winston, New York, 1976.

[7] E. L. Lawler, J. K. Lenstra, A. H. G. Rinnooy Kan, and D. B. Shmoys, editors, *The Traveling Salesman Problem*, Wiley, New York, NY, 1985.

[8] C. Mata and J. S. B. Mitchell, *Approximation algorithms for geometric tour and network design problems*, Proc. 11th Annu. ACM Sympos. Comput. Geom. (1995), pp. 360–369.

[9] J. S. B. Mitchell, *Guillotine subdivisions approximate polygonal subdivisions: A simple new method for the geometric* k-MST *problem*, Proc. 7th ACM-SIAM Sympos. Discrete Algorithms (1996), pp. 402–408.

[10] J. S. B. Mitchell, *Guillotine subdivisions approximate polygonal subdivisions: Part II — A simple polynomial-time approximation scheme for geometric k-MST, TSP, and related problems*, SIAM J. Comp., to appear.

[11] J. S. B. Mitchell, A. Blum, P. Chalasani, and S. Vempala, *A constant-factor approximation for the geometric k-MST problem in the plane*, SIAM J. Comp., to appear.

[12] C. H. Papadimitriou and K. Steiglitz, *Combinatorial Optimization: Algorithms and Complexity*, Prentice Hall, Englewood Cliffs, NJ, 1982.