Geodesic Spanners for Points on a Polyhedral Terrain

SIAM J. COMPUT.
Vol. 48, No. 6, pp. 1796–1810

GEODESIC SPANNERS FOR POINTS ON A POLYHEDRAL TERRAIN*

MOHAMMAD ALI ABAM†, MARK DE BERG‡, AND
MOHAMMAD JAVAD REZAEI SERAJI†

Abstract. Let S be a set of n points on a polyhedral terrain $\mathcal{T}$ in $\mathbb{R}^3$, and let $\varepsilon > 0$ be a fixed constant. We prove that S admits a $(2+\varepsilon)$-spanner with $O(n \log n)$ edges with respect to the geodesic distance. This is the first spanner with constant spanning ratio and a near-linear number of edges for points on a terrain. On our way to this result, we prove that any set of n weighted points in $\mathbb{R}^d$ admits an additively weighted $(2+\varepsilon)$-spanner with $O(n)$ edges; this improves the previously best known bound on the spanning ratio (which was $5+\varepsilon$) and almost matches the lower bound.

Key words. geometric spanners, approximation algorithms, polyhedral terrain

AMS subject classifications. 52B99, 65D1512

DOI. 10.1137/18M119358X

1. Introduction.

Background and motivation. When designing networks on a given set of nodes—whether they are road or railway networks, or computer networks, or some other type of networks—there are often two conflicting desiderata. On the one hand, one would like to have fast connections between any pair of nodes, and, on the other hand, one would like the network to be sparse. This leads to the concept of *spanners*, as defined next.

In an abstract setting, one is given a metric space $\mathcal{M} = (S, \mathbf{d}_\mathcal{M})$, where the elements from S are called *points*—the points represent the nodes in the network—and $\mathbf{d}_\mathcal{M}$ is a metric on S. A t-*spanner* for $\mathcal{M}$, for a given $t > 1$, is an edge-weighted graph $\mathcal{G} = (S, E)$ where the weight of each edge $(p, q) \in E$ is equal to $\mathbf{d}_\mathcal{M}(p, q)$ and the following condition is satisfied: for all pairs $p, q \in S$, we have that $\mathbf{d}_\mathcal{G}(p, q) \leqslant t \cdot \mathbf{d}_\mathcal{M}(p, q)$, where $\mathbf{d}_\mathcal{G}(p, q)$ denotes the distance between p and q in $\mathcal{G}$. (The distance between p and q in $\mathcal{G}$ is defined as the minimum weight of any path connecting p and q in $\mathcal{G}$.) In other words, the distance between any two points in the spanner $\mathcal{G}$ approximates their original distance in the metric space $\mathcal{M}$ up to a factor t. The factor t is called the *spanning ratio* (or dilation, or stretch factor) of $\mathcal{G}$. The question now becomes the following: can we construct a sparse graph with small spanning ratio? Or, stated differently, given a desired spanning ratio t, how many edges do we need to obtain a t-spanner?

Previous work. As mentioned, the concept of spanners arises naturally in the design of efficient networks. Spanners have also been used as a tool in solving a variety of other problems. It is not surprising therefore that spanners have been

*Received by the editors June 11, 2018; accepted for publication (in revised form) September 24, 2019; published electronically December 12, 2019. The abstract version of this paper appeared in *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA), 2017.
https://doi.org/10.1137/18M119358X
Funding: The second author's research was supported by the Netherlands Organisation for Scientific Research (NWO) under project 024.002.003.
†Department of Computer Engineering, Sharif University of Technology, Tehran, Iran (abam@sharif.ir, http://sharif.edu/~abam; mjrezaei@ce.sharif.edu, http://ce.sharif.edu/~mjrezaei).
‡Department of Computer Science, TU Eindhoven, 5600 MB Eindhoven, The Netherlands (m.t.d.berg@tue.nl, http://www.win.tue.nl/~mdberg/).

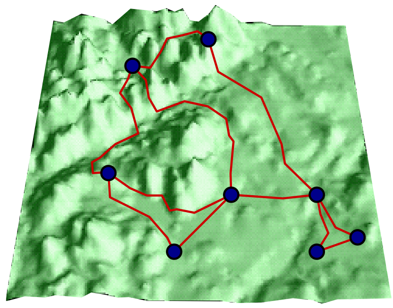

FIG. 1. *A set of points on a terrain. Some shortest paths are highlighted.*

studied extensively. Many papers on spanners focus on obtaining spanners of small size, i.e., that is, with a small number of edges. This is also the focus of our paper. However, other properties—spanners in which the total weight of the edges is small, or spanners in which the maximum degree of the nodes is small—are also of interest. Dynamic and kinetic spanners have been considered as well; see [2, 14] for some recent results.

In the most general version, where we do not have any additional properties of the underlying metric space, one can get a $(2k-1)$-spanner of size $O(n^{1+1/k})$, for any integer $k > 0$, by the method given in [5] and an improvement on its main lemma (Lemma 6 in [5]) in [4]. In particular, in general metric spaces no methods are known to obtain constant spanning ratio with a spanner of size $O(n \operatorname{polylog} n)$. For several special types of metric spaces, much better results can be obtained, however.

One important case is when S is a set of n points in $\mathbb{R}^d$ and the *Euclidean metric* is used. For any fixed $\varepsilon > 0$, one can then obtain a $(1+\varepsilon)$-spanner with $O(n)$ edges. More precisely, there is a $(1+\varepsilon)$-spanner with $O(n/\varepsilon^{d-1})$ edges. See the book by Narasimhan and Smid [13] for an extensive discussion on geometric spanners. Another special case that received ample attention [7, 8, 9, 10, 15] are metric spaces of *bounded doubling dimension*. (A metric space $\mathcal{M} = (S, \mathbf{d}_{\mathcal{M}})$ has doubling dimension d if any ball of radius r in the space can be covered by 2^d balls of radius $r/2$.) Also for spaces whose doubling dimension is a constant—note that this is a generalization of Euclidean spaces—it is possible to obtain, for any fixed $\varepsilon > 0$, a $(1+\varepsilon)$-spanner with $O(n)$ edges.

Another natural generalization to study is the case where the points in S lie on a polyhedral terrain $\mathcal{T}$ and the *geodesic distance* is used; see Figure 1. A polyhedral terrain is the graph of a piecewise linear function $f : D \to \mathbb{R}^3$, where D is a convex polygonal region in the plane. Polyhedral terrains (also called TINs) are often used in GIS to model mountainous landscapes [11]. The geodesic distance, $\mathbf{d}_{\mathcal{T}}(p,q)$, between two points $p, q \in \mathcal{T}$ is the length on the shortest path on the terrain between p and q. We call a spanner for points on a terrain with respect to the geodesic distance a *geodesic spanner*. At first sight, it may seem that geodesic spanners are very similar to Euclidean spanners. This is not the case: a crucial difference is that the metric space $(\mathcal{T}, \mathbf{d}_{\mathcal{T}})$ does not have bounded doubling dimension. Indeed, it is unknown whether any set of points on a terrain admits a spanner with constant spanning ratio and of size $O(n \operatorname{polylog} n)$. (Kapoor and Li [12] studied geodesic spanners on polyhedral surfaces, but their spanner construction uses Steiner points and its size depends on the so-called geometric dilation factor of the surface; in the worst case, their Steiner spanner has size $\Omega(n^2)$.)

There are two recent results that deal with special cases of geodesic spanners on terrains.

First, consider the case where the terrain is completely flat except for n needle-like peaks, and the points in S are located on the top of these peaks. This leads to the concept of *additively weighted spanners*, as studied by Abam et al. [3]. Here one is given a set S of points in $\mathbb{R}^2$ (or, more generally, in $\mathbb{R}^d$), where each $p \in S$ has a nonnegative weight $\mathrm{w}(p)$; the weights model the heights of the peaks. The additively weighted distance $\mathbf{d}_w(p,q)$ between two points $p, q \in S$ is now defined as

$$\mathbf{d}_w(p,q) = \begin{cases} 0 & \text{if } p = q, \\ \mathrm{w}(p) + |pq| + \mathrm{w}(q) & \text{if } p \neq q, \end{cases}$$

where $|.|$ denote the Euclidean distance. A t-spanner $\mathcal{G}$ for the metric space $(S, \mathbf{d}_w)$ is called an *additively weighted t-spanner*. Note that $(S, \mathbf{d}_w)$ does not necessarily have bounded doubling dimension. (To see this, take a set S of n points inside a unit disk in the plane, each having unit weight.) Nevertheless, Abam et al. [3] showed that there exists a $(5 + \varepsilon)$-spanner $\mathcal{G}$ with a linear number of edges for the metric space $(S, \mathbf{d}_w)$. They also proved that for any $\varepsilon > 0$, there are weighted point sets S such that any $(2 - \varepsilon)$-spanner of $(S, \mathbf{d}_w)$ has $\Omega(n^2)$ edges.

A second special case of spanners on a terrain is where the terrain is again completely flat, except for a number of polygonal plateaus at very high elevations, and the points on S are located on the flat part of the terrain. If the plateaus are sufficiently high, then this terrain can be seen as a domain with polygonal holes. Abam et al. [1] recently showed that for a set of n points in a polygonal domain with $h > 0$ holes, there exists a $(5 + \epsilon)$-spanner of size $O(n\sqrt{h}\log^2 n)$. When $h = 0$, they obtain a $(\sqrt{10} + \epsilon)$-spanner with $O(n \log^2 n)$ edges. The main question is still open, however: is there a geodesic spanner with $O(n \operatorname{polylog} n)$ edges and constant spanning ratio for any set of n points on a terrain?

Our results. We answer the question above affirmatively by showing that, for any constant $\varepsilon > 0$, there exists a $(2+\varepsilon)$-spanner with $O(n \log n)$ edges—we only show the existence of such a spanner and leave its computation to future research. Note that our result not only generalizes the recent result of Abam et al. [1], but it also improves both the spanning ratio and the size of the spanner. Also note that the lower bound for additively weighted spanners implies that we cannot hope to get spanning ratio $2 - \varepsilon$ with a subquadratic number of edges. On the way to proving this result, we present a new algorithm to construct an additively weighted spanner. This spanner has $O(n)$ edges, like the one of Abam et al. [3], but its spanning ratio is $2 + \varepsilon$, an improvement over the previously known bound of $5 + \varepsilon$. Given the lower bound and the fact that our spanner uses $O(n)$ edges, this is essentially optimal. Our method to obtain a $(2 + \varepsilon)$-spanner on a terrain uses, besides the additively weighted spanners in $\mathbb{R}^1$, another tool that we believe is of independent interest: we show that for any set of n points on a terrain, there is a *balanced shortest-path separator*: a shortest path connecting two points on $\partial\mathcal{T}$, or a triangle whose sides are shortest paths, that partitions the point set S into two subsets of size at least $2n/9$.

Our method works not only for points on a terrain, but it works for points on any polyhedral surface in $\mathbb{R}^3$ that is a topological disk. The main property we require is that between any two points s and t on the surface there are finitely many shortest paths.

2. Additively weighted spanners for points in $\mathbb{R}^d$. In this section, we present our improved spanner construction for additively weighted point sets. We use the same global approach as Abam et al. [3]—we cluster the points in a suitable way, then we construct a spanner on the cluster centers, and finally we connect the points to the cluster centers—but the implementation of the various steps is different.

Let S be the given weighted set of n points in $\mathbb{R}^d$ for which we want to construct a spanner. We will partition S into a number of clusters $C_i \subseteq S$, each with a designated center $c_i \in C_i$, such that the clusters have the following two properties. Let $\mathcal{C}$ be the set of all cluster centers:

(i) The metric space $(\mathcal{C}, \mathbf{d}_w)$ has doubling dimension $O(d \log(1/\varepsilon))$, which will be shown in Lemma 2.1.

(ii) For any cluster C_i and any point $p \in C_i$, we have $\mathbf{d}_w(p, c_i) \leqslant (2+\varepsilon) \cdot \mathrm{w}(p)$, which will be shown in Theorem 2.2).

The following algorithm takes as input the weighted point set S and a parameter $\varepsilon > 0$ and computes a clustering of S with these properties:

1. Sort the points of S in nondecreasing order of their weight, with ties broken arbitrarily. Let $p_1, p_2, \ldots, p_n$ be the resulting sorted sequence.
2. Initialize the first cluster as $C_1 := \{p_1\}$, define $c_1 := p_1$ to be its center, and initialize the set of cluster centers as $\mathcal{C} := \{c_1\}$. Set $m := 1$ to be the current number of clusters.
3. Handle the points $p_2, \ldots, p_n$ in order as follows. Let p_i be the point to be handled:
   (a) Compute a center $c_j \in \mathcal{C}$ with $1 \leqslant j \leqslant m$ whose Euclidean distance to p_i is minimum.
   (b) If $|c_j p_i| \leqslant \varepsilon \cdot \mathrm{w}(p_i)$, then add p_i to cluster C_j. Otherwise, start a new cluster with p_i as its center: set $m := m+1$, set $C_m = \{p_i\}$ and $c_m := p_i$, and set $\mathcal{C} := \mathcal{C} \cup \{p_i\}$.
4. Return the collection $\{C_1, \ldots, C_m\}$ of clusters, with $\mathcal{C}$ as cluster centers.

LEMMA 2.1. *The metric space $(\mathcal{C}, \mathbf{d}_w)$ has doubling dimension $O(d \log(1/\varepsilon))$.*

Proof. Consider a $\mathbf{d}_w$-ball $B(c_i, r)$ with radius r centered at a point $c_i \in \mathcal{C}$. We must show that $B(c_i, r)$ can be covered by $2^{O(d \log(1/\varepsilon))}$ balls of radius $r/2$. To this end, let $\mathcal{C}^* \subseteq B(c_i, r)$ be a maximal set of centers such that $\mathbf{d}_w(c_j, c_k) > r/2$ for every pair $c_j, c_k \in \mathcal{C}^*$. Then the set of balls $\{B(c_j, r/2) : c_j \in C^*\}$ covers $B(c_i, r)$. Hence, it suffices to prove that $|\mathcal{C}^*| = O(1/\varepsilon^d)$.

Define $\mathcal{C}_1^* := \{c_j \in \mathcal{C}^* : \mathrm{w}(c_j) \leqslant r/8\}$ and $\mathcal{C}_2^* := \mathcal{C}^* \setminus \mathcal{C}_1^*$. Since the $\mathbf{d}_w$-distance of any two points $c_j, c_k \in \mathcal{C}^*$ is at least $r/2$, we have $|c_j c_k| \geqslant r/2 - \mathrm{w}(c_j) - \mathrm{w}(c_k)$. For $c_j, c_k \in \mathcal{C}_1^*$, this implies that $|c_j c_k| \geqslant r/4$. A simple packing argument shows that we can only put $O(t^d)$ points whose mutual distances are at least r/t into a ball with radius r in $\mathbb{R}^d$. We conclude that $|\mathcal{C}_1^*| = O(4^d)$. To bound the size of $\mathcal{C}_2^*$, we use the fact that in our construction any two centers $c_j, c_k \in \mathcal{C}$ satisfy $|c_j c_k| > \varepsilon \cdot \min(\mathrm{w}(c_j), \mathrm{w}(c_k))$. For $c_j, c_k \in \mathcal{C}_2^*$, this implies that $|c_j c_k| > \varepsilon \cdot (r/8)$. The packing argument now implies $|\mathcal{C}_2^*| = O((8/\varepsilon)^d)$. Therefore, we have $|\mathcal{C}^*| = |\mathcal{C}_1^*| + |\mathcal{C}_2^*| = O(4^d + 1/\varepsilon^d) = O(1/\varepsilon^d)$. ∎

We can now compute a $(2+\varepsilon)$-spanner $\mathcal{G}$ on S, for a given $0 \leqslant \varepsilon \leqslant \sqrt{2} - 1$, as follows:

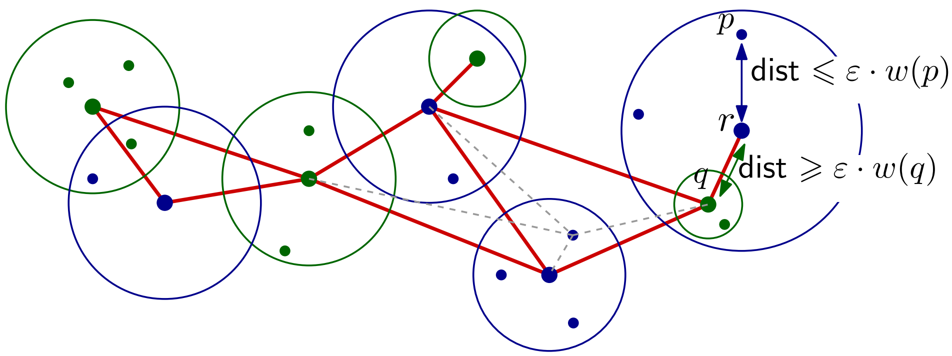

FIG. 2. *The clustering and spanner construction.*

I. Compute a clustering $\{C_1, \ldots, C_m\}$ and a set $\mathcal{C}$ of cluster centers as described above.

II. Construct a $(1+\varepsilon)$-spanner $\mathcal{B}$ on $\mathcal{C}$ using the method given by Gottlieb and Roditty [9] for computing spanners in spaces of bounded doubling dimension. The spanner produced by this method has the special property that the maximum degree in $\mathcal{B}$ is $O((2+1/\varepsilon)^{O(\text{dim})})$, where dim is the doubling dimension. We call $\mathcal{B}$ the *backbone* of our spanner.

III. To obtain our final spanner $\mathcal{G}$, we connect each noncenter point p to the backbone: we connect p to the center c_i of the cluster C_i containing p, and in addition we connect p to all the neighbors of c_i in $\mathcal{B}$.

Figure 2 shows an example of both clustering and spanner construction. The large dots show center points, and the small dots show the other points. The red (solid) segments show the edges of a $(1+\varepsilon)$-spanner on centers, and gray (dashed) segments show the extra edges added to the spanner for a noncenter point. The Euclidean distance of a noncenter point p to its assigned center r is at most $\varepsilon \cdot \mathrm{w}(p)$, and the Euclidean distance of a center point q to a center point r whose weight is smaller is at least $\varepsilon \cdot \mathrm{w}(q)$.

We conclude this section with the following theorem.

THEOREM 2.2. *Let S be a set of n weighted points in $\mathbb{R}^d$, and let $\varepsilon > 0$ be a fixed constant. There exists a $(2+\varepsilon)$-spanner with $(2+1/\varepsilon)^{O(d \log(1/\varepsilon))} n$ edges for the metric space $(S, \mathbf{d}_w)$.*

Proof. The bound on the number of edges follows immediately from Lemma 2.1 together with the fact that the maximum degree in the backbone $\mathcal{B}$ is $O((2+1/\varepsilon)^{O(\text{dim})})$, where dim is the doubling dimension. It remains to prove the bound on the spanning ratio. Let $\mathcal{G} = (S, E)$ be the computed spanner. We must prove that $\mathbf{d}_{\mathcal{G}}(p, q) \leqslant (2+\varepsilon) \cdot \mathbf{d}_w(p, q)$. If $(p, q) \in E$, this is obviously true. If both p and q are centers, then this is also true since the backbone $\mathcal{B}$ is a $(1+\varepsilon)$-spanner on $\mathcal{C}$. So now consider the case where one or both of p, q are noncenter points. Let C_i and C_j be the clusters containing p_i and p_j, respectively. Note that our construction guarantees that $|pc_i| \leqslant \varepsilon \cdot \mathrm{w}(p)$ and that $\mathrm{w}(p) \geqslant \mathrm{w}(c_i)$; two similar properties hold for q and c_j. (These properties are used several times in the derivations below.) We consider two cases:

- The first case is that p and q belong to the same cluster, and so $c_i = c_j$. We then have

$$\begin{aligned}
\mathbf{d}_{\mathcal{G}}(p,q) &\leqslant \mathbf{d}_{\mathcal{G}}(p,c_i) + \mathbf{d}_{\mathcal{G}}(c_i,q) \\
&= \big(\,\mathrm{w}(p) + |pc_i| + \mathrm{w}(c_i)\,\big) \\
&\quad + \big(\,\mathrm{w}(c_i) + |c_i q| + \mathrm{w}(q)\,\big) \\
&\leqslant \big(\,\mathrm{w}(p) + \varepsilon \cdot \mathrm{w}(p) + \mathrm{w}(p)\,\big) \\
&\quad + \big(\,\mathrm{w}(q) + \varepsilon \cdot \mathrm{w}(q) + \mathrm{w}(q)\,\big) \\
&= (2+\varepsilon)\cdot(\mathrm{w}(p) + \mathrm{w}(q)) \\
&\leqslant (2+\varepsilon)\cdot(\mathrm{w}(p) + |pq| + \mathrm{w}(q)) \\
&= (2+\varepsilon)\cdot \mathbf{d}_w(p,q).
\end{aligned}$$

- The second case is that p and q belong to different clusters, and so $c_i \neq c_j$. Since the backbone $\mathcal{B}$ is a $(1+\varepsilon)$-spanner on $\mathcal{C}$, the shortest path in $\mathcal{B}$ from c_i to c_j has length at most $(1+\varepsilon)\cdot \mathbf{d}_w(c_i,c_j)$. Define c_s and c_t to be the neighbors of c_i and c_j along this path, respectively. (If the path consists of two edges, then $c_s = c_t$, and if it consists of a single edge, then we define $c_s = c_t = c_j$.) Note that (p, c_s) and (c_t, q) are edges in $\mathcal{G}$. Hence,

$$\begin{aligned}
\mathbf{d}_{\mathcal{G}}(p,q) &\leqslant \mathbf{d}_w(p,c_s) + \mathbf{d}_{\mathcal{B}}(c_s,c_t) + \mathbf{d}_w(c_t,q) \\
&= \mathrm{w}(p) + |pc_s| + \mathrm{w}(c_s) \\
&\quad + \mathbf{d}_{\mathcal{B}}(c_s,c_t) \\
&\quad + \mathrm{w}(c_t) + |c_t q| + \mathrm{w}(q) \\
&\leqslant \mathrm{w}(p) + |pc_i| + |c_i c_s| + \mathrm{w}(c_s) \\
&\quad + \mathbf{d}_{\mathcal{B}}(c_s,c_t) \\
&\quad + \mathrm{w}(c_t) + |c_t c_j| + |c_j q| + \mathrm{w}(q).
\end{aligned}$$

Moreover, because $(c_i, c_s, \ldots, c_t, c_j)$ is a shortest path in $\mathcal{B}$, we have

$$\begin{aligned}
\mathbf{d}_{\mathcal{B}}(c_i,c_j) &= \mathrm{w}(c_i) + |c_i c_s| + \mathrm{w}(c_s) \\
&\quad + \mathbf{d}_{\mathcal{B}}(c_s,c_t) \\
&\quad + \mathrm{w}(c_t) + |c_t c_j| + \mathrm{w}(c_j).
\end{aligned}$$

Since $\mathcal{B}$ is a $(1+\varepsilon)$-spanner, we thus get

$$\begin{aligned}
&|c_i c_s| + \mathrm{w}(c_s) + \mathbf{d}_{\mathcal{B}}(c_s,c_t) + \mathrm{w}(c_t) + |c_t c_j| \\
&\quad = \mathbf{d}_{\mathcal{B}}(c_i,c_j) - \mathrm{w}(c_i) - \mathrm{w}(c_j) \\
&\quad \leqslant (1+\varepsilon)\cdot \mathbf{d}_w(c_i,c_j) - \mathrm{w}(c_i) - \mathrm{w}(c_j).
\end{aligned}$$

It follows that

$$\begin{aligned}
\mathbf{d}_{\mathcal{G}}(p,q) &\leqslant \mathrm{w}(p) + |pc_i| + (1+\varepsilon)\cdot \mathbf{d}_w(c_i,c_j) \\
&\quad - \ \mathrm{w}(c_i) - \mathrm{w}(c_j) + |c_j q| + \mathrm{w}(q) \\
&\leqslant \mathrm{w}(p) + \varepsilon\cdot \mathrm{w}(p) + (1+\varepsilon)\cdot \mathbf{d}_w(c_i,c_j) \\
&\quad - \ \mathrm{w}(c_i) - \mathrm{w}(c_j) + \varepsilon\cdot \mathrm{w}(q) + \mathrm{w}(q) \\
&\leqslant \mathrm{w}(p) + \varepsilon\cdot \mathrm{w}(p) \\
&\quad + (1+\varepsilon)\cdot(\mathrm{w}(c_i) + |c_i c_j| + \mathrm{w}(c_j)) \\
&\quad - \ \mathrm{w}(c_i) - \mathrm{w}(c_j) + \varepsilon\cdot \mathrm{w}(q) + \mathrm{w}(q) \\
&= (1+\varepsilon)\cdot(\mathrm{w}(p) + \mathrm{w}(q)) \\
&\quad + \ (1+\varepsilon)\cdot |c_i c_j| \\
&\quad + \ \varepsilon\cdot(\mathrm{w}(c_i) + \mathrm{w}(c_j)) \\
&\leqslant (1+\varepsilon)\cdot(\mathrm{w}(p) + \mathrm{w}(q)) \\
&\quad + \ (1+\varepsilon)\cdot(|c_i p| + |pq| + |q c_j|) \\
&\quad + \ \varepsilon\cdot(\mathrm{w}(p) + \mathrm{w}(q)) \\
&\leqslant (1+2\varepsilon)\cdot(\mathrm{w}(p) + \mathrm{w}(q)) \\
&\quad + \ (1+\varepsilon)\cdot(\varepsilon\cdot \mathrm{w}(p) + |pq| + \varepsilon\cdot \mathrm{w}(q)) \\
&= (1+3\varepsilon+\varepsilon^2)\cdot(\mathrm{w}(p) + \mathrm{w}(q)) \\
&\quad + \ (1+\varepsilon)\cdot |pq| \\
&\leqslant (1+3\varepsilon+\varepsilon^2)\cdot(\mathrm{w}(p) + \mathrm{w}(q) + |pq|) \\
&= (1+3\varepsilon+\varepsilon^2)\cdot \mathbf{d}_w(p,q) \\
&\leqslant (2+\varepsilon)\cdot \mathbf{d}_w(p,q),
\end{aligned}$$

where the last inequality holds because we can assume without loss of generality that $\varepsilon \leqslant \sqrt{2}-1$.

Thus, in both cases we have $\mathbf{d}_{\mathcal{G}}(p,q) \leqslant (2+\varepsilon)\cdot \mathbf{d}_w(p,q)$. ∎

3. Spanners for points on a polyhedral terrain. Let $\mathcal{T}$ be a polyhedral terrain (or a polyhedral surface in $\mathbb{R}^3$ that is a topological disk), and let S be a set of n points on $\mathcal{T}$. In this section, we show that there is a $(2+\varepsilon)$-spanner for S with respect to $\mathbf{d}_{\mathcal{T}}$, the geodesic distance on $\mathcal{T}$. Our global approach is divide-and-conquer: we partition S into two subsets of roughly equal size, compute spanners for these subsets recursively, and then generate a set of edges to connect the points from the two subsets. For the latter step, it is important that the two subsets are separated in a suitable way. In particular, we need to separate the subsets by shortest paths (not necessarily between points in S). Next we define the two types of separators that we allow more precisely, and we show that a suitable separator always exists.

The first type of separator is a shortest path $\sigma(u,v)$ that connects two points $u, v \in \partial\mathcal{T}$. Such a shortest path partitions $\mathcal{T}$ into two regions: the closed region $\sigma^+(u,v)$ consisting of all points $q \in \mathcal{T}$ that lie to the right of the (directed) path $\sigma(u,v)$, and the open region $\sigma^-(u,v)$ consisting of all points to the left of $\sigma(u,v)$; see Figure 3(i). Note that parts of $\sigma(u,v)$ may lie on $\partial\mathcal{T}$, and so $\mathrm{Int}(\sigma^+(u,v))$ and, similarly, $\mathrm{Int}(\sigma^-(u,v))$ need not be connected where $\mathrm{Int}(.)$ denotes the interior.

The second type of separator that we allow is defined as follows. Consider three points $u, v, w \in \mathcal{T}$ with shortest paths $\sigma(u,v)$, $\sigma(v,w)$, and $\sigma(w,u)$ connecting them, and assume these paths are pairwise disjoint except at shared endpoints. We call the closed region Δ bounded by such a triple of paths a *shortest-path triangle*, or *sp-*

FIG. 3. (i) *The shortest path* $\sigma(u,v)$ *(in red) partitions* $\mathcal{T}$ *into two regions,* $\sigma^+(u,v)$ *(in grey) and* $\sigma^-(u,v)$ *(in white).* (ii) *The paths* $\sigma(u,v^*)$ *(solid) and* $\sigma'(u,v^*)$ *(dashed, except the part that overlaps* $\sigma(u,v^*)$*) enclose a number of regions (in grey). If* R_i *contains more than* $4n/9$ *points, it is turned into an sp-triangle by adding a point* w_0 *on one of the paths from* u_0 *to* v_0*. Color is available online only.*

triangle for short. The paths $\sigma(u,v)$, $\sigma(v,w)$, and $\sigma(w,u)$ are called the *sides* of Δ. A *degenerate sp-triangle* is either a shortest path or a path along $\partial\mathcal{T}$. In the following, when we talk about sp-triangles we also allow degenerate sp-triangles.

We call a separator of one of the two types defined above an *sp-separator*. The main tool in our spanner construction is the following theorem.

THEOREM 3.1. *For any set of* n *points on a polyhedral terrain* $\mathcal{T}$*, there is a balanced sp-separator. More precisely, there is either a shortest path* $\sigma(u,v)$ *connecting two points* $u,v \in \partial\mathcal{T}$ *such that* $2n/9 \leqslant |\sigma^+(u,v)\cap S| \leqslant 2n/3$*, or there is an sp-triangle* Δ *such that* $2n/9 \leqslant |\Delta \cap S| \leqslant 2n/3$.

To prove the theorem, we first try to find a balanced sp-separator of the first type. If this fails, we argue that a suitable sp-triangle exists.

Let $u \notin S$ be an arbitrary point on $\partial\mathcal{T}$. Now move a point v around $\partial\mathcal{T}$, starting at u and traversing $\partial\mathcal{T}$ counterclockwise, until v reaches u again. Let $\sigma(u,v)$ be the shortest path from u to v that maximizes the area of $\sigma^+(u,v)$ (the grey region in Figure 3(i)). As we continuously move v along $\partial\mathcal{T}$, the shortest path $\sigma(u,v)$ moves forward and changes continuously, except at certain *breakpoints*. More precisely, we can partition $\partial\mathcal{T}$ into finitely many pieces—the breakpoints are the endpoints of these pieces—such that as v moves along one such boundary piece, we can deform $\sigma(u,v)$ continuously. Initially, when v is still infinitesimally close to u, we have $\sigma^+(u,v)\cap S = \emptyset$; at the very end, when v approaches u again, we have $\sigma^+(u,v)\cap S = S$.

If at some point during the traversal v reaches a position such that $2n/9 \leqslant |\sigma^+(u,v)\cap S| \leqslant 2n/3$, then we have found our balanced sp-separator. Otherwise, there is a breakpoint v^* at which $\sigma(u,v)$ jumps over more than $2n/3 - 2n/9 = 4n/9$ points from S. In this case, there are two shortest paths $\sigma(u,v^*)$ and $\sigma'(u,v^*)$ such that the (open) region R enclosed by $\sigma(u,v^*)$ and $\sigma'(u,v^*)$ contains more than $4n/9$ points from S.

Note that R may consist of more than one connected component, as $\sigma(u,v^*)$ and $\sigma'(u,v^*)$ may overlap at some points (see Figure 3(ii)). If all of them contain at most $4n/9$ points from S, then we can obtain a balanced separator by only jumping over a subset of the components. Otherwise, there is a single component, R_i, that contains more than $4n/9$ points. Let u_0 and v_0 be the two points on ∂R_i where $\sigma(u,v^*)$ and $\sigma'(u,v^*)$ meet. Let w_0 be an arbitrary point on ∂R_i that is distinct from u_0 and v_0; see Figure 3(ii). Then the triple u_0, v_0, w_0, together with ∂R_i, defines an sp-triangle Δ_0 containing at least $4n/9$ points from S. Next we show how to construct a sequence of sp-triangles $\Delta_0 \supset \Delta_1 \supset \cdots \supset \Delta_k$ such that $2n/9 \leqslant |\Delta_k \cap S| \leqslant 2n/3$. We will maintain the invariant that $|\Delta_i \cap S| \geqslant 2n/9$. Note that this is indeed satisfied for Δ_0.

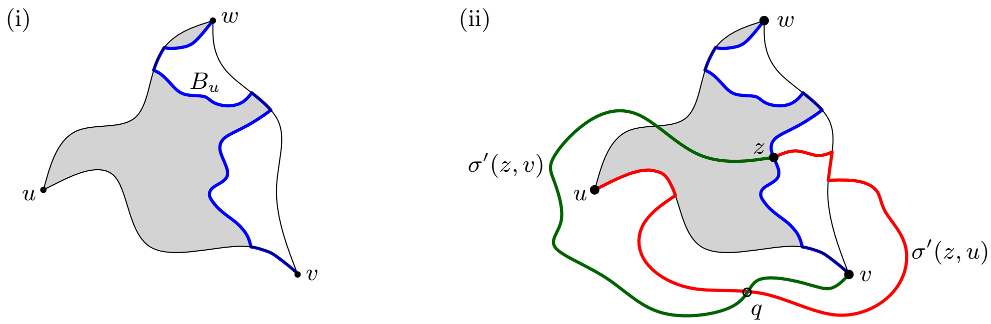

FIG. 4. (i) *The blue curve is* B_u. *The interior of the region* Z_u *is indicated in grey; the curve* B_u *is also part of* Z_u. (ii) *The paths* $\sigma'(z,u)$ *and* $\sigma'(z,v)$ *intersect. Color is available online only.*

In the following, we denote the vertices of the sp-triangle Δ_i by u_i, v_i, w_i, its interior by $\text{Int}(\Delta_i)$, and its boundary by $\partial\Delta_i$.

Suppose we have constructed Δ_i. We distinguish three cases:

1. If $|\Delta_i \cap S| \leqslant 2n/3$, then Δ_i is the final triangle in our construction and we are done.
2. Otherwise, if Δ_i is a degenerate sp-triangle (i.e., $\text{Int}(\Delta_i)$ is empty), then we can immediately find an sp-triangle $\Delta_{i+1} \subset \Delta_i$ with the required properties: we just take a subpath containing $2n/9$ points.
3. If the previous two cases do not apply, then Δ_i is a nondegenerate sp-triangle containing more than $2n/3$ points. This is the difficult case, and next we show how to handle it.

So now assume Δ_i is a nondegenerate sp-triangle containing more than $2n/3$ points. Note that if Δ_i has a side containing at least $2n/9$ points from S, we can again finish the construction by taking a suitable subpath of this side as our next (and final) sp-triangle. Hence, we can assume that each side contains less than $2n/9$ points, which implies there is at least one point—actually, at least four points—in the interior of Δ_i. Next we show how to construct an sp-triangle $\Delta_{i+1} \subset \Delta_i$ containing at least $2n/9$ points such that either $|\Delta_{i+1} \cap S| < |\Delta_i \cap S|$ or $|\text{Int}(\Delta_{i+1}) \cap S| < |\text{Int}(\Delta_i) \cap S|$. Note that the condition that $|\Delta_{i+1} \cap S| < |\Delta_i \cap S|$ or $|\text{Int}(\Delta_{i+1}) \cap S| < |\text{Int}(\Delta_i) \cap S|$ implies that our process terminates. Indeed, when $|\text{Int}(\Delta_{i+1}) \cap S| = 0$ and Δ_{i+1} contains more than $2n/3$ points, we have a side with more than $2n/9$ points, and so we can finish the construction as described above.

To simplify the notation, we will drop the subscript i from now on. Thus, we are given a nondegenerate sp-triangle Δ with corners u, v, w that contains more than $2n/3$ points from S and has at least one point in its interior. For a point $z \in \Delta$, we call a path from z to one of the corners of Δ a *good path* if it is a shortest path that stays within Δ (although not necessarily in its interior). Define $Z_u \subseteq \Delta$ to be the set of all points $z \in \Delta$ such that there is a good path $\sigma(z,u)$ to the corner u. The region Z_u is simply connected, and its boundary consists of the sides $\sigma(u,v)$ and $\sigma(u,w)$, and a curve B_u from v to w; see Figure 4(i). Note that B_u may overlap partially (or fully) with $\sigma(u,v)$ and/or $\sigma(u,w)$.

LEMMA 3.2. *For any point* $z \in B_u$, *there are good paths* $\sigma(z,u)$, $\sigma(z,v)$, *and* $\sigma(z,w)$ *to the three corners of* Δ.

Proof. By the definition of B_u, there is a good path from z to u. If z lies on $\sigma(v,w)$, the side of Δ opposite u, then we also have good paths to v and w. Now

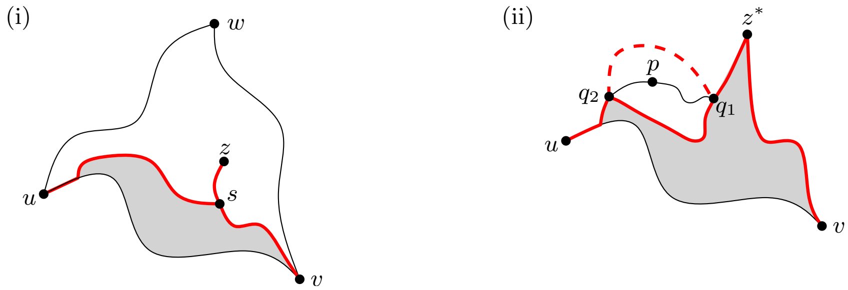

FIG. 5. (i) *A good-path tree (in red) with root* z *and leaves* u *and* v*. The interior of* Δ_z *is indicated in grey.* (ii) *A subpath of the shortest path* $\sigma(z,u)$ *jumps at* z *from the dashed subpath to the solid subpath from* q_1 *to* q_2*. Region* R *is split into two sp-triangles by connecting* p *to* q_1 *and* q_2 *by shortest paths. Color is available online only.*

assume $z \notin \sigma(v,w)$.

Since $z \in \partial Z_u$ and $z \notin \sigma(v,w)$, we have not only a good path from z to u, but we also have a shortest path $\sigma'(z,u)$ that does not stay inside Δ and that cannot be shortcut (while maintaining its length) to do so. Note that as soon as $\sigma'(z,u)$ exits Δ through one of the sides $\sigma(u,v)$ or $\sigma(u,w)$, it could also follow that side to u. Hence, we can assume $\sigma'(z,u)$ is as follows: it exits Δ through $\sigma(v,w)$ (possibly after following $\sigma(v,w)$ for awhile), and then it moves through $\text{Ext}(\Delta)$ until it hits one of the sides incident to u, say $\sigma(u,v)$, which it then follows to u. See Figure 4(ii). (The portion along $\sigma(u,v)$ may be empty.) Note that within $\text{Ext}(\Delta)$ the path $\sigma'(z,u)$ separates u either from v or from w. Assume without loss of generality that the former is the case, as in Figure 4(ii). We now argue the existence of good paths $\sigma(z,v)$ and $\sigma(z,w)$.

First consider a shortest path $\sigma'(z,v)$. If $\sigma'(z,v)$ already stays inside Δ, we have a good path to v. Otherwise, it must go through $\text{Ext}(\Delta)$ and, hence, cross $\sigma'(z,u)$ at a point q. Note that the distances from z to q along $\sigma'(z,u)$ and along $\sigma'(z,v)$ must be equal. But then the path from z to v that follows $\sigma'(z,u)$ until it hits the side $\sigma(v,w)$ and then follows that side to v cannot be longer than $\sigma'(z,v)$. Hence, we have a good path from z to v.

Now consider a shortest path $\sigma'(z,w)$. If it does not already stay inside Δ, it exists Δ through the side $\sigma(u,v)$ and it crosses $\sigma'(z,u)$. In the latter case, we can use the same argument as above and find a good path to w. ∎

Now imagine moving a point z from w to v along B_u. Initially the sp-separator defined by vertices u, v, and z is the sp-triangle Δ which contains all points, and at the end of the movement it has changed to the degenerate sp-separator with vertices u, v, and v which does not contain any points. Therefore, during the movement of z from w to v, we will find a sp-separator with the desired number of points. Next we make this idea precise.

By the previous lemma, at any point we have good paths $\sigma(z,u)$ and $\sigma(z,v)$. Now consider a shortest-path tree[1] T_z with z as the root and u and v as leaves that consist of good paths $\sigma(z,u)$ and $\sigma(z,v)$ to u and v; see Figure 5(i). We call such a shortest-path tree a *good-path tree*. When $z = w$, we take $\sigma(z,u)$ and $\sigma(z,v)$ to be the sides $\sigma(w,u)$ and $\sigma(w,v)$ of Δ. When $z = v$, we take $\sigma(z,u)$ to be the side

[1] The two shortest paths $\sigma(z,u)$ and $\sigma(z,v)$ may not form a tree because they meet more than once, but by rerouting we can always get rid of this situation.

$\sigma(v,u)$; the path $\sigma(z,v)$ is then empty. Let Δ_z denote the sp-triangle with v as one of its corners that is bounded by (part of) T_z and (part of) the side $\sigma(u,v)$.

As we continuously move z along B_u, the good-path tree T_z also changes continuously, except at certain breakpoints. Initially, when $z = w$, we have $T_z = \sigma(w,u) \cup \sigma(w,v)$, and so $\Delta_z = \Delta$. At the end, when $z = v$, we have $T_z = \sigma(u,v)$, and so $\mathrm{Int}(\Delta_z)$ is empty. (Thus, Δ is a degenerate sp-triangle.) Now consider the first moment when either $|\Delta_z \cap S|$ decreases or $|\,\mathrm{Int}(\Delta_z) \cap S|$ decreases. Let z^* be the point at which this happens. We have two cases:

- If $|\Delta_{z^*} \cap S| \geqslant 2n/9$, then we can take $\Delta_{i+1} := \Delta_{z^*}$.
- Otherwise, the number of points we just lost is more than $2n/3 - 2n/9 = 4n/9$. This can happen when one of the two good paths forming T_z jumps. In this case, z^* must be a breakpoint, and there are two different shortest paths from z^* to u, or two different shortest paths from z^* to v (or both). Assume we have two shortest paths from z^* to u. The points that we lose as z moves over the breakpoint are located in between these two paths. The area between the two paths may consist of various regions. If one of these regions contains at least one and at most $4n/9$ points, then we can find a suitable sp-triangle Δ_{i+1} by only jumping over this region. Otherwise, we have a region R that contains more than $4n/9$ points. This region is bounded by two shortest paths that meet at shared endpoints, q_1 and q_2. We then take any point $p \in \mathrm{Int}(R) \cap S$ and connect p by shortest paths that stay in R to q_1 and q_2; see Figure 5(ii). This partitions R into two sp-triangles. At least one of them contains more than $2n/9$ points. We take this sp-triangle to be Δ_{i+1}. Note that Δ_{i+1} contains fewer points in its interior than Δ since $p \notin \mathrm{Int}(\Delta_{i+1})$.

In both cases we find an sp-triangle with the required properties, thus finishing the proof of Theorem 3.1. Next we describe how to use this theorem to compute a spanner for a set S of points on a terrain.

The spanner construction. Next we show how to compute a spanner $\mathcal{G}(S) = (S, E_S)$ for a set S of n points on a terrain $\mathcal{T}$. We first describe how to obtain a $(6+\varepsilon)$-spanner; then we show how to improve the construction to reduce the spanning ratio to $(2+\varepsilon)$.

Let $\varepsilon > 0$ be a given constant. If $|S| \leqslant 3$, we connect all pairs of points in S; that is, $\mathcal{G}(S)$ is the complete graph. If $|S| > 3$, we proceed as follows:

1. Take a balanced sp-separator as in Theorem 3.1. If the separator is a shortest path $\sigma(u,v)$ connecting points $u, v \in \partial\mathcal{T}$, then define $S_{\mathrm{in}} := \sigma^+(u,v) \cap S$; if the separator is an sp-triangle Δ, then define $S_{\mathrm{in}} := \Delta \cap S$. If the first case applies, we will from now on call the shortest path $\sigma(u,v)$ the *side* of the separator and denote it by λ. Thus, a separator has one or three sides. Note that a side is always a shortest path on $\mathcal{T}$. Set $S_{\mathrm{out}} := S \setminus S_{\mathrm{in}}$.
2. Process each side λ of the separator as follows:
   (i) For each $p \in S$, let $p_\lambda \in \lambda$ be a point whose geodesic distance $\mathbf{d}_{\mathcal{T}}(p, p_\lambda)$ to p is minimum; see Figure 6. Assign each point p_λ a weight $\mathrm{w}(p_\lambda) := \mathbf{d}_{\mathcal{T}}(p, p_\lambda)$, and define $S_\lambda := \{p_\lambda : p \in S\}$ to be the resulting weighted (multi-)set. Note that for two distinct points p and q, points p_λ and q_λ may be equal, but we pretend that they are different points with their Euclidean distance being zero.
   (ii) We view λ as a 1-dimensional Euclidean space and S_λ as a weighted point set in this space. (Note that because λ is a shortest path on $\mathcal{T}$, distances d_λ in the 1-dimensional space λ are the same as distances $\mathbf{d}_{\mathcal{T}}$ on $\mathcal{T}$.) We now construct an additively weighted $(2+\varepsilon_1)$-spanner $\mathcal{G}_\lambda$ for

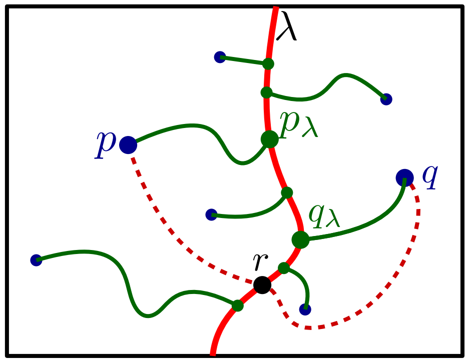

FIG. 6. *Spanner construction.*

S_λ, where $\varepsilon_1 = \varepsilon/3$, using the method from Theorem 2.2.

(iii) For each edge (p_λ, q_λ) in the spanner $\mathcal{G}_\lambda$, we add (p, q) to E_S.

3. Recursively compute spanners $\mathcal{G}(S_{\text{in}}) = (S_{\text{in}}, E_{\text{in}})$ and $\mathcal{G}(S_{\text{out}}) = (S_{\text{out}}, E_{\text{out}})$, and add the edge sets E_{in} and E_{out} to E_S.

LEMMA 3.3. *The construction above gives a* $(6+\varepsilon)$*-spanner with respect to the geodesic distance. The spanner has* $O(c_\varepsilon n \log n)$ *edges, where* c_ε *is a constant depending on* ε.

Proof. By Theorem 2.2, the number of edges we add to the spanner in step 2 is $O(n)$. Hence, if $A(n)$ denotes the total number of edges we have in our spanner on n points, then we have $A(n) = O(c_\varepsilon n) + A(n_1) + A(n_2)$, where $n_1 + n_2 = n$ and $n_1, n_2 \leqslant 7n/9$ and $c_\varepsilon = (2 + 1/\varepsilon)^{O(\log(1/\varepsilon))} n$. Hence, $A(n) = O(c_\varepsilon n \log n)$ as claimed.

Let p, q be two arbitrary points in S. If both points are in S_{in} or both points are in S_{out}, then we have a $(6+\varepsilon_1)$-path between them by induction. So assume $p \in S_{\text{in}}$ and $q \in S_{\text{out}}$. Let $\sigma(p, q)$ be a shortest path on $\mathcal{T}$ from p to q (the dotted path in Figure 6). Let λ be a side of Δ intersected by $\sigma(p, q)$. Consider the points p_λ and q_λ. Then there is a path π in $\mathcal{G}_\lambda$ of length at most $(2 + \varepsilon_1) \cdot \mathbf{d}_{\lambda,w}(p_\lambda, q_\lambda)$, where $\mathbf{d}_{\lambda,w}$ denotes the additively weighted distance in the 1-dimensional space λ. The same path, with each point x_λ replaced by its original $x \in S$, also exists in our spanner $\mathcal{G}$. Note that for any two points $x, y \in S$ we have

$$\begin{aligned} \mathbf{d}_\mathcal{T}(x, y) &\leqslant \mathbf{d}_\mathcal{T}(x, x_\lambda) + \mathbf{d}_\mathcal{T}(x_\lambda, y_\lambda) + \mathbf{d}_\mathcal{T}(y_\lambda, y) \\ &= \mathrm{w}(x_\lambda) + d_\lambda(x_\lambda, y_\lambda) + \mathrm{w}(y_\lambda) \\ &= \mathbf{d}_{\lambda,w}(x, y). \end{aligned}$$

Let r be a point in $\lambda \cap \sigma(p, q)$. Then $\mathbf{d}_\mathcal{T}(p, q) = \mathbf{d}_\mathcal{T}(p, r) + \mathbf{d}_\mathcal{T}(r, q)$. We also have $\mathbf{d}_\mathcal{T}(p, p_\lambda) \leqslant \mathbf{d}_\mathcal{T}(p, r)$ and $\mathbf{d}_\mathcal{T}(q, q_\lambda) \leqslant \mathbf{d}_\mathcal{T}(q, r)$ by the definition of p_λ and q_λ. Hence,

$$\begin{aligned} \mathbf{d}_\mathcal{G}(p, q) &\leqslant \text{length}(\pi) \\ &\leqslant (2 + \varepsilon_1) \cdot \mathbf{d}_{\lambda,w}(p_\lambda, q_\lambda) \\ &= (2 + \varepsilon_1) \cdot \big(\mathrm{w}(p_\lambda) + d_\sigma(p_\lambda, q_\lambda) + \mathrm{w}(q_\lambda) \big) \\ &= (2 + \varepsilon_1) \cdot \big(\mathbf{d}_\mathcal{T}(p, p_\lambda) + \mathbf{d}_\mathcal{T}(p_\lambda, q_\lambda) + \mathbf{d}_\mathcal{T}(q, q_\lambda) \big) \\ &\leqslant (2 + \varepsilon_1) \cdot \big(\mathbf{d}_\mathcal{T}(p, r) + \mathbf{d}_\mathcal{T}(p_\lambda, q_\lambda) + \mathbf{d}_\mathcal{T}(q, r) \big) \\ &= (2 + \varepsilon_1) \cdot \left(\mathbf{d}_\mathcal{T}(p, q) + \mathbf{d}_\mathcal{T}(p_\lambda, q_\lambda) \right). \end{aligned}$$

Moreover,

$$\begin{aligned}\mathbf{d}_{\mathcal{T}}(p_\lambda, q_\lambda) &\leqslant \mathbf{d}_{\mathcal{T}}(p_\lambda, p) + \mathbf{d}_{\mathcal{T}}(p, q) + \mathbf{d}_{\mathcal{T}}(q, q_\lambda)\\ &\leqslant \mathbf{d}_{\mathcal{T}}(r, p) + \mathbf{d}_{\mathcal{T}}(p, q) + \mathbf{d}_{\mathcal{T}}(q, r)\\ &= 2\,\mathbf{d}_{\mathcal{T}}(p, q).\end{aligned}$$

Thus, $\mathbf{d}_{\mathcal{G}}(p,q) \leqslant (2+\varepsilon_1)\cdot 3\,\mathbf{d}_{\mathcal{T}}(p,q) = (6+\varepsilon)\cdot \mathbf{d}_{\mathcal{T}}(p,q)$. □

We now refine our construction to reduce the spanning ratio to $2+\varepsilon$. The idea behind the improvement is as follows. As follows from the proof of Lemma 3.3, we would get a $(2+\varepsilon_1)$-spanner if we had $\mathbf{d}_{\lambda,w}(p_\lambda, q_\lambda) = \mathbf{d}_{\mathcal{T}}(p,q)$. In the above construction, however, we have $\mathbf{d}_{\lambda,w}(p_\lambda, q_\lambda) \leqslant 3\,\mathbf{d}_{\mathcal{T}}(p,q)$, giving a $(6+\varepsilon)$-spanner. We can obtain $\mathbf{d}_{\lambda,w}(p_\lambda, q_\lambda) \leqslant (1+\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p,q)$, for a suitable $\varepsilon_2 = O(\varepsilon)$, by modifying step 2 as follows.

In step 2(i), we take for each $p \in S$ not a single point $p_\lambda \in \lambda$ but a collection $S(p,\lambda)$ defined as follows. As before, let p_λ be a point on λ that is closest to p. Let $\lambda(p) \subseteq \lambda$ be the set of points on λ whose distance to p_λ is at most $(1+2/\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p, p_\lambda)$, that is,

$$\lambda(p) := \{x \in \lambda : \mathbf{d}_{\lambda,w}(p_\lambda, x) \leqslant (1+2/\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p, p_\lambda)\}.$$

We partition $\lambda(p)$ into $O(1/\varepsilon_2^2)$ pieces $\lambda_i(p)$, each of length at most $\varepsilon_2 \cdot \mathbf{d}_{\mathcal{T}}(p, p_\lambda)$. For each such piece $\lambda_i(p)$, let $p_\lambda^{(i)}$ be a point on $\lambda_i(p)$ that is closest to p. We now take $S(p,\lambda)$ to be the set of all such points $p_\lambda^{(i)}$, where we set $\mathrm{w}(p_\lambda^{(i)}) := \mathbf{d}_{\mathcal{T}}(p, p_\lambda^{(i)})$. Note that $|S(p,\lambda)| = O(1/\varepsilon_2^2)$.

In step 2(ii), we now compute a $(2+\varepsilon_1)$ spanner $\mathcal{G}_\lambda$ on the set $\bigcup_{p\in S} S(p,\lambda)$. In step 2(iii), we then add for each edge $(p_\lambda^{(i)}, q_\lambda^{(j)})$ in $\mathcal{G}_\lambda$ the edge (p,q) to E_S. This leads to the following result.

THEOREM 3.4. *Let S be a set of n points on a polyhedral terrain $\mathcal{T}$ in $\mathbb{R}^3$, and let $\varepsilon > 0$ be a fixed constant. Then there exists a $(2+\varepsilon)$-spanner with $O(c_\varepsilon n \log n)$ edges with respect to the geodesic distance, where c_ε is a constant depending on ε.*

Proof. To prove the bound on the spanner size, we observe that as compared to our previous construction the number of points for which we compute an additively weighted $(2+\varepsilon_1)$-spanner in step 2(ii) has increased from $O(n)$ to (n/ε_2^2). Hence, if we set $\varepsilon_2 = O(\varepsilon)$, the overall number of edges increases by a factor $O(1/\varepsilon^2)$.

It remains to prove the bound on the spanning ratio of our spanner $\mathcal{G}$. To this end, consider two points $p \in S_{\text{in}}$ and $q \in S_{\text{out}}$. As in the proof of Lemma 3.3, let r be a point where the shortest path $\sigma(p,q)$ crosses λ. If $r \notin \lambda(p)$, set $p' := p_\lambda$. Otherwise, set p' to be the closest point in $S(p,\lambda)$ to r. Similarly, define q' for point q. Note that

$$\begin{aligned}\mathbf{d}_{\lambda,w}(p', q') &= \mathrm{w}(p') + d_\lambda(p', q') + \mathrm{w}(q')\\ &= \mathbf{d}_{\mathcal{T}}(p, p') + d_\lambda(p', q') + \mathbf{d}_{\mathcal{T}}(q, q')\\ &\leqslant \mathbf{d}_{\mathcal{T}}(p, p') + d_\lambda(p', r) + d_\lambda(r, q') + \mathbf{d}_{\mathcal{T}}(q, q').\end{aligned}$$

We next prove that $\mathbf{d}_{\mathcal{T}}(p, p') + d_\lambda(p', r) \leqslant (1+\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p, r)$. We have two cases:

- *Case* A: $r \notin \lambda(p)$. In this case, $p' = p_\lambda$ and $d_\lambda(p', r) > (1+2/\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p, p_\lambda)$.

Hence,

$$\begin{aligned}
&\mathbf{d}_{\mathcal{T}}(p,p') + d_\lambda(p',r) \\
&\quad\leqslant \mathbf{d}_{\mathcal{T}}(p,p_\lambda) + (\mathbf{d}_{\mathcal{T}}(p,p_\lambda) + \mathbf{d}_{\mathcal{T}}(p,r)) \\
&\quad\leqslant 2\cdot \mathbf{d}_{\mathcal{T}}(p,p_\lambda) + \mathbf{d}_{\mathcal{T}}(p,r) \\
&\quad\leqslant 2\cdot(\varepsilon_2/2)\cdot(d_\lambda(p',r) - \mathbf{d}_{\mathcal{T}}(p,p_\lambda)) + \mathbf{d}_{\mathcal{T}}(p,r) \\
&\quad\leqslant \varepsilon_2\cdot(\mathbf{d}_{\mathcal{T}}(p_\lambda,p) + \mathbf{d}_{\mathcal{T}}(p,r) - \mathbf{d}_{\mathcal{T}}(p,p_\lambda)) + \mathbf{d}_{\mathcal{T}}(p,r) \\
&\quad= (1+\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p,r).
\end{aligned}$$

- *Case* B: $r \in \lambda(p)$. Now we have

$$\mathbf{d}_\lambda(p',r) \leqslant \varepsilon_2 \cdot \mathbf{d}_{\mathcal{T}}(p,p_\lambda),$$

and so

$$\begin{aligned}
\mathbf{d}_{\mathcal{T}}(p,p') + d_\lambda(p',r) &\leqslant \mathbf{d}_{\mathcal{T}}(p,r) + \varepsilon_2\cdot \mathbf{d}_{\mathcal{T}}(p,p_\lambda) \\
&\leqslant \mathbf{d}_{\mathcal{T}}(p,r) + \varepsilon_2\cdot \mathbf{d}_{\mathcal{T}}(p,r) \\
&= (1+\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p,r).
\end{aligned}$$

So in both cases we have

$$\mathbf{d}_{\mathcal{T}}(p,p') + d_\lambda(p',r) \leqslant (1+\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p,r).$$

In a similar way, we can prove that

$$\mathbf{d}_{\mathcal{T}}(q,q') + d_\lambda(q',r) \leqslant (1+\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(q,r).$$

Hence, $\mathbf{d}_{\lambda,w}(p',q') \leqslant (1+\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p,q)$. Combining this with the proof of Lemma 3.3, which now holds with p_λ replaced by p' and q_λ replaced by q', we obtain

$$\mathbf{d}_{\mathcal{G}}(p,q) \leqslant (2+\varepsilon_1)\cdot(1+\varepsilon_2)\cdot \mathbf{d}_{\mathcal{T}}(p,q).$$

Picking $\varepsilon_1 = \varepsilon_2 = \varepsilon/4$ now gives us the desired spanning ratio (assuming without loss of generality that $\varepsilon \leqslant 1$). ∎

4. Concluding remarks. We have shown that any set of n points on a polyhedral terrain $\mathcal{T}$ admits a geodesic spanner of spanning ratio $2+\varepsilon$ and with $O(n \log n)$ edges. This is the first geodesic spanner for points on a terrain. In fact, our method works in a more general setting than for polyhedral terrains: it suffices to have a piecewise-linear surface that is a topological disk. We focused on proving the existence of the spanner, leaving the computation of such spanners to future research.

REFERENCES

[1] M. A. Abam, M. Adeli, H. Homapour, and P. Zafar Asadollahpoor, *Geometric spanners for points inside a polygonal domain*, in Proceedings of the 31st International Symposium on Computational Geometry (SoCG), 2015, pp. 186–197.

[2] M. A. Abam and M. de Berg, *Kinetic spanners in* $\mathbb{R}^d$, Discrete Comput. Geom., 45 (2011), pp. 723–736.

[3] M. A. Abam, M. de Berg, M. Farshi, J. Gudmundsson, and M. Smid, *Geometric spanners for weighted point sets*, Algorithmica, 61 (2011), pp. 207–225.

[4] N. Alon, S. Hoory, and N. Linial, *The Moore bound for irregular graphs*, Graphs Combin., 18 (2002), pp. 53–57.

[5] I. Althöfer, G. Das, D. Dobkin, D. Joseph, and J. Soares, *On sparse spanners of weighted graphs*, Discrete Comput. Geom., 9 (1993), pp. 81–100.

[6] S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Wu, *An optimal algorithm for approximate nearest neighbor searching in fixed dimensions*, J. ACM, 45 (1998), pp. 891–923.
[7] H. T.-H. Chan, A. Gupta, B. M. Maggs, and S. Zhou, *On hierarchical routing in doubling metrics*, in Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), 2005, pp. 762–771.
[8] R. Cole and L. A. Gottlieb, *Searching dynamic point sets in spaces with bounded doubling dimension*, in Proceedings of the 38th Annual ACM Symposium on Theory of Computing (STOC), 2006, pp. 574–583.
[9] L.A. Gottlieb and L. Roditty, *An optimal dynamic spanner for doubling metric spaces*, in Proceedings of the 16th European Symposium on Algorithms (ESA), Lecture Notes in Comput. Sci. 5193, Springer, Berlin, Heidelberg, 2008, pp. 478–489.
[10] S. Har-Peled and M. Mendel, *Fast construction of nets in low-dimensional metrics and their applications*, SIAM J. Comput., 35 (2006), pp. 1148–1184, https://doi.org/10.1137/S0097539704446281.
[11] I. Heywood, S. Cornelius, and S. Carver, *An Introduction to Geographical Information Systems*, 3rd ed., Prentice–Hall, Upper Saddle River, NJ, 2006.
[12] S. Kapoor and X.-Y. Li, *Geodesic spanners on polyhedral surfaces*, in Proceedings of the 20th International Symposium on Algorithms and Computational (ISAAC), Lecture Notes in Comput. Sci. 5878, Springer, Berlin, Heidelberg, 2009, pp. 213–223.
[13] G. Narasimhan and M. Smid, *Geometric Spanner Networks*, Cambridge University Press, Cambridge, UK, 2007.
[14] L. Roditty, *Fully dynamic geometric spanners*, Algorithmica, 62 (2012), pp. 1073–1087.
[15] K. Talwar, *Bypassing the embedding: Algorithms for low dimensional metrics*, in Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC), 2004, pp. 281–290.