SIAM J. COMPUT.
Vol. 41, No. 3, pp. 587–617

APPROXIMATING TSP ON METRICS WITH BOUNDED GLOBAL GROWTH*

T.-H. HUBERT CHAN† AND ANUPAM GUPTA‡

Abstract. The traveling salesman problem (TSP) is a canonical NP-complete problem which is proved by Trevisan [*SIAM J. Comput.*, 30 (2000), pp. 475–485] to be MAX-SNP hard even on high-dimensional Euclidean metrics. To circumvent this hardness, researchers have been developing approximation schemes for "simpler" instances of the problem. For instance, the algorithms of Arora and of Talwar show how to approximate TSP on low-dimensional metrics (for different notions of metric dimension). However, a feature of most current notions of metric dimension is that they are "local": the definitions require *every* local neighborhood to be well-behaved. In this paper, we define a global notion of dimension that generalizes the popular notion of doubling dimension, but still allows some small dense regions; e.g., it allows some metrics that contain cliques of size $\sqrt{n}$. Given a metric with global dimension $\dim_C$, we give a $(1+\varepsilon)$-approximation algorithm that runs in *subexponential time*, i.e., in $\exp(O(n^{\delta}\varepsilon^{-4\dim_C}))$-time for every constant $0<\delta<1$. As mentioned above, metrics with bounded $\dim_C$ may contain metrics of size $O(\sqrt{n})$ on which the TSP problem is hard to approximate to within $(1+\varepsilon)$. Hence, to do better than a running time of $\Omega(\exp\{\sqrt{n}\})$, our algorithms find $O(1)$-approximations to some portions of the tour, and $(1+\varepsilon)$-approximations for other portions, and stitch them together. Moreover, we show that such globally bounded metrics have spanners that preserve distances to arbitrary accuracy and have size $\Theta(n^{1.5})$.

Key words. traveling salesman problem, approximation algorithms, metric spanners, global notion of dimension

AMS subject classifications. 11K55, 68W25

DOI. 10.1137/090749396

1. Introduction.

This paper presents a definition for the *global dimension* of a metric space, shows that it generalizes the notion of *doubling dimension*, and gives a subexponential approximation algorithm for the traveling salesman problem (TSP) on metrics with bounded global dimension. Recall that a metric space $M=(V,d)$ is a set of points V with a distance function $d: V\times V\to\mathbb{R}_{\geq 0}$ such that that distances are symmetric, satisfy the triangle inequality, and $d(x,y)=0 \iff x=y$. Unless specified otherwise, we assume that the set V is finite.

TSP has long been known to be NP-hard [GJ79], and Papadimitriou and Yannakakis showed that TSP is MAX-SNP hard even for metrics where all distances are either 1 or 2 [PY93]. To show that an underlying geometric structure may not make the problem tractable, Trevisan proved that TSP remains MAX-SNP hard if the Euclidean dimension is unbounded [Tre00]. On the other hand, Arora [Aro98] gave a polynomial-time approximation scheme (PTAS) for TSP on bounded dimensional Euclidean metrics. Moreover, TSP admits exact algorithms on metrics which arise from graphs that have bounded treewidth [AP89] and also admits a PTAS for

*Received by the editors February 11, 2009; accepted for publication (in revised form) March 19, 2012; published electronically May 31, 2012. A conference version of the paper appeared in *Proceedings of the* 19*th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, San Francisco, CA, 2008, pp. 690–699. This research was partly supported by NSF grant CCR-0122581 (the ALADDIN project), NSF CAREER award CCF-0448095, and an Alfred P. Sloan Fellowship.

http://www.siam.org/journals/sicomp/41-3/74939.html

†Department of Computer Science, The University of Hong Kong, Hong Kong (hubert@cs.hku.hk). This research was done while this author was at Carnegie Mellon University.

‡Computer Science Department, Carnegie Mellon University, Pittsburgh, PA 15213 (anupamg@cs.cmu.edu).

metrics arising from weighted planar graphs (see, e.g., [Kle05]). If the graph excludes some fixed minor, it is known how to obtain a $(1+\varepsilon)$-approximation to TSP using an algorithm that runs in quasi-polynomial time for every constant ε [Gri00].

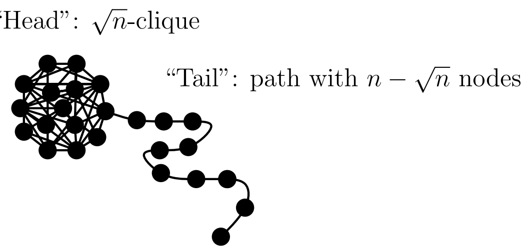
The above algorithms show that if a metric can be represented in some specific way (e.g., if a metric can be represented as a set of points in bounded-dimensional Euclidean space or as the shortest-path metric of a planar graph), it admits optimal or near-optimal algorithms for TSP. Consequently, if a metric is such that distances can be stretched by at most D to get another metric that is representable in some special way, one gets an approximation for TSP that is at most a factor of D worse. A slightly different approach has been to define notions of dimension for arbitrary metric spaces (and not just for geometric ones) and then to parameterize the time- or space-complexity of algorithms using not just the number $n = |V|$ of points in the metric space, but also the dimension of the metric space itself. For example, building on the study of "bounded growth-rate" metrics [PRR99, KR02, HKRZ02] and a definition of [Ass83], researchers considered the *doubling dimension* $\dim_D(M)$ of a metric M [Cla99, GKL03]; a formal definition appears in section 2. The doubling dimension of a metric space has proved to be a very useful parameter in algorithm design on metric spaces. It generalizes the notion of dimension in geometric spaces, i.e., $\dim_D(\mathbb{R}^d, \ell_p) = \Theta(d)$. Moreover, the performance (run-time, space) of many algorithms can be given as functions of $|V|$ and $\dim_D(M)$, hence allowing a more nuanced quantification than those obtained only in terms of the number of points. For the TSP problem, Talwar [Tal04] gave a $(1+\varepsilon)$-approximation algorithm such that for metrics with doubling dimension $\dim_D(M)$ at most k, the algorithm runs in time $\exp((\frac{k}{\varepsilon}\log n)^{O(k)})$. Hence, for metrics with constant doubling dimension, and for constant ε, this gives a $(1+\varepsilon)$-approximation to TSP in quasi-polynomial time.

If a metric has bounded doubling dimension, this implies that for any subset S of points which are nearly equidistant from one another, the cardinality of S is bounded; indeed, if the doubling dimension of the metric is k, the cardinality of any set S is bounded by $\Delta(S)^{\dim_D}$, where the *aspect ratio* $\Delta(S)$ is the ratio between the largest and smallest distances between distinct points in S. Since this has to hold for all subsets of points in the metric space, the introduction of even a small but nonconstant set of equidistant points into a space with bounded doubling dimension would cause the doubling dimension to become unbounded. This is the starting point of our investigation: in this paper, we define a new notion of dimension called the *correlation dimension*, which attempts to circumvent this problem. We give some basic structural results about how the correlation dimension is related to various metric properties; we then present algorithms for spanners and TSP for metrics with low correlation dimension. Our definitions are inspired by work on relaxed notions of dimensions such as the *correlation fractal dimension* in physics [GP83] and in databases [BF95].

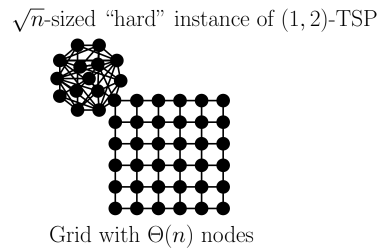
Note that correlation dimension is not the first idea to incorporate dense regions in graphs (e.g., see [KL06] for a different direction, which also gives good approximations for TSP). But it gives a different, global notion of dimension, and can be useful in contexts where strict, local ways of measuring dimension may not be applicable.

1.1. Our results and techniques. Given a metric $M = (V, d)$, for any $x \in V$ and $r \geq 0$, let $\mathbf{B}(x, r) = \{y \in V \mid d(x, y) \leq r\}$ denote the ball around point x of radius r. The *correlation dimension* is defined as the smallest constant $k > 0$ such that for every $r > 0$,

$$\sum_{x \in V} |\mathbf{B}(x, 2r)| \leq 2^k \cdot \sum_{x \in V} |\mathbf{B}(x, r)|, \tag{1.1}$$



(a) Lollipop



(b) Augmented Grid

FIG. 1.1. *Examples of metrics with low correlation dimension but high doubling dimension.*

and moreover, this inequality must hold under taking any net of the metric M. (Formal definitions of this and other relevant notions of dimension are given in section 2.) Note that if, instead of taking a sum over all $x \in V$, we were to take a maximum over all $x \in V$, we would get the notion of strong doubling dimension based on bounded-growth rate used by [PRR99, KR02]; it implies that the notion of correlation dimension generalizes the notion of pointwise dimension. While not so immediate, we show in section 2.2 the following relationship between correlation dimension and doubling dimension.

THEOREM 1.1 (correlation generalizes doubling). *Given any metric M, its correlation dimension is bounded above by nine times its doubling dimension.*

Note that it is easy to construct metrics where the correlation dimension is asymptotically much less than the doubling dimension: adding a clique of size $O(\sqrt{n})$ to a doubling metric may not change its correlation dimension by much but definitely increases its doubling dimension to $\Omega(\log n)$. (More examples are given in Figure 1.1.) Observe that a subset of a metric with bounded correlation dimension could have large correlation dimension: for instance, we can just take the $\Theta(\sqrt{n})$-sized clique as the subset in the above example. Moreover, we show as a corollary of Theorem 3.1 that correlation dimension is not preserved even when the metric is modified with distortion 2.

The following theorems show the algorithmic potential of this definition. (Please refer to [IM04, Mat02] for the definitions of embedding and distortion and to [Die00] for the definition of treewidth.)

THEOREM 1.2 (embedding into small treewidth graphs). *Given any constant $0 < \varepsilon < 1$ and k, metrics with correlation dimension at most k can be embedded into a distribution of graphs with treewidth $\tilde{O}_{k,\varepsilon}(\sqrt{n})$ and distortion $1 + \varepsilon$.*

This result implies $(1 + \varepsilon)$-approximate randomized algorithms with nontrivial running times for problems that can be solved efficiently on small-treewidth graphs, which include the TSP, facility location problems, and many other NP-hard optimization problems on metrics. Moreover, we can do much better for the TSP despite the presence of these $O(\sqrt{n})$-sized cliques (or other complicated metrics of that size); we can make use of the global nature of the TSP problem (and the corresponding global nature of $\dim_C$) to get the following result.

THEOREM 1.3 (approximation schemes for TSP). *Given any metric M with $\dim_C(M) = k$, there is a randomized algorithm that runs in time $2^{O(n^\delta \varepsilon^{-O(k)})}$ for any constant $\delta > 0$ and outputs a traveling salesman tour with expected cost within a $(1+\varepsilon)$-factor of the optimum.*

Hence, given constants ε, k, the algorithm runs in *subexponential* time. (Recall that subexponential time is $\cap_{\delta>0}\mathsf{TIME}(2^{n^\delta})$.) As we will see later, the best exponent in the expression above that we can show is $(\varepsilon^{-1}2^{\sqrt{\log n \log\log n}})^{4k}$; substantially improving this expression remains an interesting open problem.

Finally, we show that while metrics with bounded correlation dimension do not necessarily admit $(1+\varepsilon)$-stretch spanners with a linear number of edges, we can get better spanners than those possible for general metrics.

THEOREM 1.4 (sparse spanners). *Given any $0 < \varepsilon < 1$, any metric with correlation dimension k admits a spanner with $O(n^{3/2}\varepsilon^{-O(k)})$ edges that has stretch $(1+\varepsilon)$. Moreover, there exist metrics whose correlation dimension is 2, for which any 1.5-stretch spanner has $\Omega(n^{3/2})$ edges.*

1.2. Related work. Many notions of dimension for metric spaces (and for arbitrary measures) have been proposed; see the survey by Clarkson [Cla06] for the definitions and for their applicability to near-neighbor (NN) search. Some of these give us strong algorithmic properties which are useful beyond NN-searching. For instance, the *strong doubling dimension* of a metric $M = (V, d)$ is the smallest value k such that for all $x \in V$ and all r, $|\mathbf{B}(x, 2r)| \le 2^k \cdot |\mathbf{B}(x, r)|$. This notion was used in [PRR99, KR02, HKRZ02] to develop algorithms for object location in general metrics and in [KK77, AM05] for routing problems.

The notion of *doubling dimension* [Ass83, Cla99, GKL03] has been used for a large number of algorithms, e.g., for NN-searching [Cla99, KL04, KL05, BKL06, HPM05, CG06], TSP and other optimization problems [Tal04], low-stretch compact routing [Tal04, CGMZ05, Sli05, AGGM06, KRX06a, KRX06b], sparse spanners [CGMZ05, HPM05], and other applications [KSW04, KMW05]. Many algorithms that were originally developed for Euclidean spaces have subsequently been extended to work for doubling metrics.

For TSP on low-dimensional Euclidean spaces, the first PTASs were given by Arora [Aro98] and Mitchell [Mit99]; see, e.g., [CL98, ARR99, CLZ02, KR99] for algorithms for other problems in low-dimensional spaces. The run-time of Arora's algorithm [Aro98] for points in $\mathbb{R}^k$ was $O(n(\log n)^{O(\sqrt{k}\cdot\frac{1}{\varepsilon})^{k-1}})$, which was subsequently improved to $2^{(\frac{k}{\varepsilon})^{O(k)}} n + O(kn\log n)$ [RS99]. For $(1+\varepsilon)$-approximation for TSP on metrics with doubling dimension k, the best running time currently known is $2^{(\frac{k}{\varepsilon}\log n)^{O(k)}}$ [Tal04]. There has been much work on algorithms for TSP on other restricted classes of inputs: for instance, Arnborg and Proskurowski [AP89] studied the TSP for graphs with bounded treewidth and gave a dynamic program that solves TSP on the induced metrics exactly in linear time. For metrics induced by weighted planar graphs, Klein [Kle05] gave a $(1+\varepsilon)$-approximation algorithm that runs in time $O(c^{1/\varepsilon^2}n)$. Grigni [Gri00] gave quasi-PTASs (QPTAS)[1] for metrics induced by minor-forbidding graphs and bounded-genus graphs.

The concept of *correlation fractal dimension* [GP83] was used by physicists to distinguish between a chaotic source and a random source; while it is closely related to other notions of fractal dimension, it has the advantage of being easily computable. Let us define it here, since it may be useful to compare our definitions with the intuition behind the original definitions. Consider an infinite set V. If $\sigma = \{x_i\}_{i\geq 1}$ is a sequence of points in V, the *correlation sum* is defined as $C_n(r) = \frac{1}{n^2}|\{(i,j) \in [n]\times[n] \mid d(x_i,x_j) \leq r\}|$ (i.e., the fraction of pairs at distance at most r from each other). The *correlation integral* is then $C(r) = \lim_{n\to\infty} C_n(r)$, and the *correlation fractal dimension* for σ is defined to be $\sup_{r>0} \lim_{\epsilon\to 0} \frac{\log C((1+\epsilon)r) - \log C(r)}{\log(1+\epsilon)}$. Hence, given a set of points, the correlation fractal dimension quantifies the rate of growth in the number of points which can see each other as their range-of-sight increases. This notion was studied by Belussi and Faloutsos [BF95] (see also [PKF00]) for estimating the selectivity of spatial queries; Faloutsos and Kamel [FK94] also used fractal dimension to analyze R-trees. In the next section, we will define a version of this definition for finite sets.

2. Correlation dimension: Definition and motivation.

Given a finite metric $M = (V,d)$, we denote the number of points $|V|$ by n. For radius $r > 0$, we define the ball $\mathbf{B}(x,r) = \{y \in V \mid d(x,y) \leq r\}$. Given $U \subseteq V$, define $\mathbf{B}_U(x,r) = \mathbf{B}(x,r) \cap U$. Recall that a subset $N \subseteq V$ is an *ε-cover* for V if for all points $x \in V$, there is a covering point $y \in N$ with $d(x,y) \leq \varepsilon$. A subset $N \subseteq V$ is an *ε-packing* if for all $x \neq y \in N$, $d(x,y) > \varepsilon$. A subset $N \subseteq V$ is an *ε-net* if it is both an ε-cover and an ε-packing. A set $N \subseteq V$ is a *net* if it is an ε-net for some ε. Inspired by the definitions mentioned in section 1.2, we give the following definition.

DEFINITION 2.1 (correlation dimension). *The* correlation dimension $\dim_C(M)$ *of a metric $M = (V,d)$ is the least $k \geq 0$ such that for all $r > 0$ and for all nets $N \subseteq V$, the following inequality holds:*

$$\sum_{x\in N} |\mathbf{B}_N(x,2r)| \leq 2^k \cdot \sum_{x\in N} |\mathbf{B}_N(x,r)|. \tag{2.1}$$

In other words, we want to ensure that the average growth rate of the metric M is not too large, and the same holds for any net N of the metric. A natural question is whether we can remove this requirement in Definition 2.1 that the bounded average growth property (2.1) hold for *every net*. Unfortunately, just requiring property (2.1) to hold for the entire metric M is too weak, as we show in section 2.3.

The *doubling constant* is the least λ such that for all $r > 0$ and every ball $\mathbf{B}(x,2r)$ of radius $2r$, there exist at most λ balls $\{B(x_i,r)\}$ with radius r such that $\mathbf{B}(x,2r) \subseteq \cup_i B(x_i,r)$; the *doubling dimension* is defined to be $\dim_D(M) = \log_2 \lambda$ [GKL03]. The *strong doubling dimension* (this name is due to [BKL06]; the notion has also been called the *KR-dimension* in [GKL03] and the doubling measure in [Cla06]) is the least k such that for all $x \in V$ and radius r,

$$|\mathbf{B}(x,2r)| \leq 2^k |\mathbf{B}(x,r)|. \tag{2.2}$$

It is known that the doubling dimension is at most four times the strong doubling dimension [GKL03, Proposition 1.2]. Moreover, it follows directly from the definitions

[1] An algorithm runs in quasi-polynomial time if there exists a constant $c > 0$ such that for every problem instance of size n, the algorithm runs in time $2^{O((\log n)^c)}$.

that the correlation dimension is no more than the strong doubling dimension. In section 2.2, we prove Theorem 1.1, which claims that $\dim_C(M) \le O(\dim_D(M))$ for any metric space M. This implies that the class of bounded correlation dimension metrics contains the class of doubling metrics; however, the converse is not true, since there exist metrics with bounded correlation dimension but unbounded doubling dimension, as we show in section 2.2.1.

2.1. Small nets. A useful property of correlation dimension is that it has "small" nets. Since metrics with bounded correlation dimension allow large cliques, the nets cannot be as small as for bounded doubling dimension.

LEMMA 2.2 (small nets). *Consider a metric $M = (V, d)$ with $\dim_C(M) \le k$. Suppose N is an R-net and $S \subseteq N$ is a subset with diameter at most D. Then, the size $|S| \le (2D/R)^{k/2} \cdot \sqrt{|N|}$.*

Proof. Observe that $|S|^2 \le \sum_{x\in N} |\mathbf{B}_N(x, D)|$. By applying the definition of correlation dimension repeatedly, we have for each integer t,

$$\sum_{x\in N} |\mathbf{B}_N(x, D)| \le 2^{kt} \sum_{x\in N} |\mathbf{B}_N(x, D/2^t)|. \tag{2.3}$$

Setting $t = \lceil \log_2(D/R) \rceil$ gives the required result. □

Define the *aspect ratio* of a set S to be the ratio of the maximum to the minimum interpoint distances $\frac{\max_{x,y\in S} d(x,y)}{\min_{x\ne y\in S} d(x,y)}$, and call a set S *near-uniform* if it has bounded aspect ratio. Then the above lemma implies that, given any metric whose correlation dimension is $O(1)$, any near-uniform subset S of points has cardinality $O(\sqrt{n})$, and hence the doubling constant λ of this metric is also $O(\sqrt{n})$.

2.2. Correlation dimension generalizes doubling. In this section, we prove Theorem 1.1, relating correlation dimension to doubling dimension; recall that the theorem claims that the correlation dimension of any metric is bounded above by nine times its doubling dimension.

Proof of Theorem 1.1. Observe that if M contains at least two points, then $\dim_D(M) \ge 1$. Hence, it suffices to show that for any nontrivial metric space $M = (V, d)$, $\dim_C(M) \le 8 \dim_D(M) + 1$. Suppose the doubling dimension $\dim_D(M) = k$ and the doubling constant $\lambda = 2^k$; it suffices to show that

$$\sum_{x\in V} |\mathbf{B}(x, 2r)| \le 2\lambda^4 \sum_{x\in V} |\mathbf{B}(x, r)|. \tag{2.4}$$

This implies that $\dim_C(M) \le \log_2 2\lambda^4 = 4\log_2 \lambda + 1 = 4\dim_D(M) + 1$. To show (2.1) for every net $N \subseteq V$, we can then apply (2.4) to the submetric $M' = (N, d|_{N\times N})$ to infer

$$\sum_{x\in N} |\mathbf{B}(x, 2r)| \le 2^{4\dim_C(M')+1} \sum_{x\in N} |\mathbf{B}(x, r)|,$$

and finally use that $\dim_D(M') \le 2\dim_D(M)$ to complete the proof of the theorem.

Let us now prove (2.4). We first obtain an upper bound for each $\mathbf{B}(x, 2r)$. Suppose Y is an $\frac{r}{2}$-net of V. Defining $Y_x := Y \cap \mathbf{B}(x, 3r)$ and $B_y := \mathbf{B}(y, \frac{r}{2})$, we can observe that

$$\mathbf{B}(x, 2r) \subseteq \bigcup_{y\in Y_x} B_y. \tag{2.5}$$

Since Y_x is contained in a ball of radius $4r$ centered at x and the interpoint distance of Y_x is greater than $\frac{r}{2}$, it follows from $\dim_D(M) = k$ that $|Y_x| \le \lambda^4$. Hence *if each* B_y *were small*, i.e., $|B_y| \le |\mathbf{B}(x,r)|$, the right-hand side would be $\le \lambda^4 \cdot |\mathbf{B}(x,r)|$.

However, we may be unlucky and have several $y \in Y_x$ such that $|B_y| > |\mathbf{B}(x,r)|$. Define the *small* centers $S_x = \{y \in Y_x \mid |\mathbf{B}(y, \frac{r}{2})| \le |\mathbf{B}(x,r)|\}$ and the set of the *large* centers $L_x := Y_x \setminus S_x$. Note that $|S_x|, |L_x| \le |Y_x| \le \lambda^4$. Combining this with (2.5), we get

$$\begin{aligned}|\mathbf{B}(x,2r)| &\le \sum_{y\in Y_x} |B_y| \le \sum_{y\in S_x} |B_y| + \sum_{y\in L_x} |B_y| \\ &\le \sum_{y\in S_x} |\mathbf{B}(x,r)| + \sum_{y\in L_x} |B_y| \le \lambda^4 |\mathbf{B}(x,r)| + \sum_{y\in L_x} |B_y|.\end{aligned}$$

Hence, summing over all $x \in V$, we have

$$\sum_{x\in V} |\mathbf{B}(x,2r)| \le \lambda^4 \sum_{x\in V} |\mathbf{B}(x,r)| + \sum_{x\in V}\sum_{y\in L_x} |B_y|. \tag{2.6}$$

The first term is what we want: we just need to bound the second term on the right-hand side of (2.6). Call this term $\mathbf{E}$. Changing the order of summation and defining $N_y := \{x \in V : y \in L_x\}$, we have

$$\mathbf{E} := \sum_{x\in V}\sum_{y\in L_x} |B_y| = \sum_{y\in Y}\sum_{x: y\in L_x} |B_y| = \sum_{y\in Y} |N_y| \cdot |B_y|. \tag{2.7}$$

So it now suffices to give an upper bound on $|N_y| \cdot |B_y|$ for every net point $y \in Y$.

Now we change our perspective to a single net point $y \in Y$. Let N'_y be an r-net of N_y. Since all points in N_y are at distance at most $4r$ from y, it follows that $|N'_y| \le \lambda^3$. Moreover, $x \in N_y$ implies that $|\mathbf{B}(x,r)| < |B_y|$. Also, we have $N_y \subseteq \cup_{x\in N'_y} \mathbf{B}(x,r)$. It follows that $|N_y| \le \lambda^3 |B_y|$. Plugging this into (2.7), we get

$$\mathbf{E} \le \sum_{y\in Y} \lambda^3 |B_y|^2. \tag{2.8}$$

For any $z \in B_y$, note that $B_y = \mathbf{B}(y, \frac{r}{2}) \subseteq \mathbf{B}(z,r)$. Observe that $|B_y| = \sum_{z\in B_y} 1$, and hence $|B_y|^2 \le \sum_{z\in B_y} |\mathbf{B}(z,r)|$. This implies that

$$\mathbf{E} \le \lambda^3 \sum_{y\in Y}\sum_{z\in B_y} |\mathbf{B}(z,r)| = \lambda^3 \sum_{z\in V}\sum_{y\in Y: z\in B_y} |\mathbf{B}(z,r)|. \tag{2.9}$$

The second equality is a change in the order of summation. Hence, to show that the expression on the right is at most $\lambda^4 \sum_{x\in V} |\mathbf{B}(x,r)|$, it suffices to show that $|\{y \in Y \mid z \in B_y\}| \le \lambda$. Define $M_z := \{y \in Y \mid z \in B_y\}$: note that we want to show $|M_z| \le \lambda$. Note that M_z is contained in a ball of radius $\frac{r}{2}$ centered at z and that any two distinct points in M_z are more than $\frac{r}{2}$ apart. From the doubling property of V, M_z contains at most λ points. Combining this with (2.6) and (2.9), we have

$$\sum_{x\in V} |\mathbf{B}(x,2r)| \le 2\lambda^4 \sum_{x\in V} |\mathbf{B}(x,r)|,$$

completing the proof. ∎

2.2.1. The converse is false. The converse of Theorem 1.1 is not true: a metric with bounded correlation dimension does not necessarily have bounded doubling dimension. To see this, consider the metric obtained by attaching a path with $n-\sqrt{n}$ nodes to a clique of size $\sqrt{n}$ (Figure 1.1(a)). The doubling dimension of this metric is at least $\log_2 \sqrt{n} = \frac{1}{2}\log n$. However, note that the quantity $\sum_{x\in V} |\mathbf{B}(x,r)|$ starts off at n (for $r=0$) and is about $\Theta(nr)$ for arbitrary $r \le n$. Moreover, this also holds true for any ε-net N, with $\sum_{x\in N} |\mathbf{B}_N(x,r)|$ being $|N|$ for $r \le \varepsilon$ and being $\Theta(|N|r/\varepsilon)$ for general $r \ge \varepsilon$. Hence the correlation dimension of this metric is $O(1)$.

2.3. Why require closure under nets? Suppose we define a metric to have *weak correlation dimension* k if

$$\sum_{x\in V} |\mathbf{B}(x,2r)| \le 2^k \cdot \sum_{x\in V} |\mathbf{B}(x,r)| \tag{2.10}$$

holds only for the original metric. Note that here we do not require (2.1) to hold for all nets N, as for correlation dimension. We can show that this weaker definition is too inclusive.

Proposition 2.3. *Given any metric $M=(V,d)$, one can find a metric $M'=(V\cup V', d')$ with the restriction $d'|_V = d$, the number of new points $|V'|=|V|$, and the* weak *correlation dimension of M' is* 2.

Proof. Without loss of generality, let the minimum nonzero interpoint distance in V be at least 1. Let $\varepsilon > 0$ be small enough such that $\varepsilon n \ll 1$. Let V' be a path on n new vertices, with edge-lengths on the path being ε, and place the path at distance 1 to some point in V. If we view the original metric as a complete graph on V, the distances given by metric d' are the shortest-path distances in the new graph formed by adding this “tail.”

We now verify that inequality (2.10) holds with some constant k for any $r>0$. The first case is when $r \ge \frac{1}{2}$. Consider the right-hand side of (2.10). Each point in the tail sees at least $\frac{n}{2}$ points for this range of r. Hence, the right-hand side is at least $\frac{n^2}{2}$. The left-hand side is at most $(2n)^2$. Hence, the inequality holds with $k=3$, which completes the proof for the case when $r \ge 1/2$. For the other case, let $0 < r < \frac{1}{2}$. For each original point $x \in V$, $|\mathbf{B}(x,2r)| = 1 = |\mathbf{B}(x,r)|$. For each point $x \in V'$ in the tail, $|\mathbf{B}(x,2r)| \le 2|\mathbf{B}(x,r)|$. Hence, in this case, the inequality (2.10) holds with $k=1$. ∎

Hence, if we do not require the closure under taking subnets, we can realize *any metric* as a submetric of a (slightly larger) low-dimensional metric, making the definition fairly uninteresting; this motivates why we need to restrict the definition further. It is an interesting question whether one can obtain an interesting definition with weaker or different restrictions.

3. Hardness of approximating the correlation dimension. We now show that it is hard to approximate the correlation dimension of a metric to a factor better than $O(\log n)$; since the correlation dimension always lies in the interval $[1, \log n]$, this proves that only trivial approximation guarantees are possible unless $NP \subseteq ZPP$.

Theorem 3.1. *Given a metric $M=(V,d)$ with n points, it is hard to distinguish between the cases* $\dim_C(M) = O(1)$ *and* $\dim_C(M) = \Omega(\log n)$*, unless* $NP \subseteq ZPP$.

Proof. The proof is by reduction from the hardness of approximation of the independent set problem. Let $G=(V,E)$ be an instance of independent set, namely a graph on n vertices, and let $\alpha(G)$ be the size of a maximum independent set in G. We will construct a metric M such that if $\alpha(G) \le n^{k_1}$ then $\dim_C(M) = O(1)$, and if $\alpha(G) \ge n^{k_2}$, then $\alpha(G) = \Omega(\log n)$. The following result then implies the hardness.

THEOREM 3.2 (see [Hås96]). *There exist constants $0 < k_1 < k_2 < 1$ such that given a graph with n vertices, it is hard to distinguish whether the size of a maximum independent set is smaller than n^{k_1} or larger than n^{k_2}, unless $NP \subseteq ZPP$.*

Define M_G to be a metric on n points, each corresponding to a vertex in G, with unit distance between two points if there is an edge between the corresponding vertices in G, and distance 2 otherwise. Hence M_G is a metric of diameter 2; note that any ε-net for M_G with $\varepsilon > 1$ is an independent set in G, and this is going to be useful for the hardness.

Let us define a parameter $l = 2(1 - k_1)$, where k_1 is the smaller constant in the hardness result for the independent set problem, and let $K = n^l$; note that $1 < K \leq n^2$; this is going to be a size parameter. Define $R = 2n^2$; this is going to be a distance parameter. We now define a metric $M = (X, d)$, with $|X| = 2nK + n^2K$. This metric M consists of the following three "components"; points in different components are at distance $10n^2KR$ from each other:

1. *Superclique*. This component consists of K copies of the metric M_G. Two points lying in different copies of M_G are at distance R from each other.
2. *Chain-of-clusters*. This component consists of a chain of K "clusters," with each cluster being a uniform metric on n points (hence unit interpoint distance). The distance between points from adjacent clusters is 2, and hence the distance between points in the ith and jth clusters is $2|i - j|$.
3. *Tail*. This component consists of a line metric with Kn^2 points, with adjacent points at distance R from each other.

We now examine the correlation dimension of this metric M. Note that bounding the correlation dimension involves analyzing the quantity $F_N(r) = \sum_{x \in N} |B_N(x, r)|$ as a function of r, starting from $r = 0$ and checking whether or not there is a sudden increase as r doubles. The first claim shows that the only interesting ϵ-nets are those with $1 \leq \epsilon < 2$.

LEMMA 3.3. *If N is an ϵ-net for the metric M where $\epsilon < 1$ or $\epsilon \geq 2$, then $\sum_{x \in N} |B_N(x, 2r)| \leq O(1) \sum_{x \in N} |B_N(x, r)|$ for any $r > 0$.*

Proof. Let us consider ϵ-nets for $\epsilon < 1$. Since the smallest distance in M is 1, the net N consists of the entire set X. For $r < 1$, since each point sees only itself, $F_N(r) = |N| = \Theta(n^2K)$. As r increases past 1 and reaches 2, all the points within each copy of M_G in the superclique, or within each cluster in the chain-of-clusters, can see one another. This gives a contribution of $2K \times n^2 = \Theta(Kn^2)$ to $F_N(r)$, but since $F_N(0) = \Theta(Kn^2)$ to begin with, the increase is not large. As r increases from 2 to R, the quantity $F_N(r)$ also increases to $\Theta(n^2K^2)$ due to the chain-of-clusters. Hence, when r reaches R, the sudden contribution of $\Theta(n^2K^2)$ due to the superclique also does not cause any sudden jumps in $F_N(r)$. Finally, as r increases beyond R, nothing interesting happens.

For $\epsilon \geq 2$, each copy of M_G can contain at most one point in N. Observe there are at most K such points, and these are at distance R from one another. On the other hand, in the chain-of-clusters, each uniform metric can have at most one point in N; hence, this component becomes a line metric. Finally, the tail is a line metric with Kn^2 points with adjacent distance R (the same as the distance between different M_G's), and hence the tail, after taking ϵ-net, counteracts any dense local clustering effect caused by the M_G's. It follows that in this case $F_N(2r) = O(1)\ F_N(r)$ for all $r > 0$. ∎

Hence it suffices to consider ϵ-nets N where $1 \leq \epsilon < 2$. For these values of ϵ, the net N can contain only one point from each cluster in the chain-of-clusters; moreover, for each copy of M_G in the superclique, the points that remain in N correspond to an independent set in the graph G. As r increases to R, the chain-of-clusters can give only a contribution of $\Theta(K^2) = o(n^2K)$; hence, if there is a large contribution to $F_N(r)$ due to the superclique as the r reaches R, there would be a sudden increase in $F_N(r)$. Thus the number of net points in each copy of M_G in the superclique (i.e., the size of the independent sets in G) becomes crucial to the ratio $F_N(2r)/F_N(r)$ for $R/2 \leq r < R$. The following two lemmas make this intuition formal.

Lemma 3.4. *Suppose a maximum independent set of G has size $\alpha(G) \leq n^{k_1}$. Then, for $1 \leq \epsilon < 2$, for any ϵ-net N of M, $F_N(2r) = O(1)F(r)$ for any $r > 0$.*

Proof. As before, the interesting action takes place when $R/2 \leq r < R$. Observe that $F_N(r) \geq n^2K = n^{2+l}$. Since the net points in each M_G correspond to an independent set in G, the contribution to $F_N(2r)$ due to the superclique is at most $(n^{k_1}K)^2 = n^{2k_1+2l} = n^{2+l}$. Hence, $F_N(2r) = O(1)F_N(r)$. ∎

Lemma 3.5. *Suppose $\alpha(G) \geq n^{k_2}$. Then, for some $1 \leq \epsilon < 2$, there exists an ϵ-net N and $R/2 \leq r < R$ such that $F_N(2r) \geq \Omega(n^{2(k_2-k_1)})F_N(r)$.*

Proof. Let $\epsilon = 1.5$ and $r = R/2$. Since G contains an independent set of size at least n^{k_2}, for each copy M_G, we can pick at least n^{k_2} net points to be in N. It follows as before that $F_N(r) \leq O(n^2K) = O(n^{2+l})$. Observe that the superclique contributes at least $(n^{k_2}K)^2 = n^{2k_2+2l}$. Hence, the ratio $F_N(2r)/F_N(r) \geq \Omega(n^{2k_2+2l-2-l}) = \Omega(n^{2(k_2-k_1)})$. ∎

Combining the lemmas completes the proof of Theorem 3.1. ∎

Corollary 3.6 (distortion does not preserve correlation dimension). *The correlation dimension of a metric can change by a factor of $\Omega(\log n)$ when it is modified with distortion 2.*

Proof. Observe that for any two graphs G and G', the corresponding metrics M_G and $M_{G'}$ as defined in the proof of Theorem 3.1 have distortion 2 from each other. It follows that the resulting metrics M and M' also have distortion 2 from each other. However, from the proof of Theorem 3.1, it is possible for one to have correlation dimension at most $O(1)$, while the other has correlation dimension at least $\Omega(\log n)$. ∎

Note that the above hardness result does not rule out using correlation dimension for algorithm design. In particular, the algorithms in this paper do not require us to know the correlation dimension of the input metric. For example, the algorithm for finding spanners is completely oblivious of the correlation dimension. On the other hand, the TSP approximation algorithm of section 6 seems to require this information at first glance, but this issue can be resolved using standard "guess-and-double" ideas, as is explained in that section.

4. Sparse spanners. We begin our study of metrics with small correlation dimension with a simple construction of sparse spanners; this also serves to introduce the reader to some basic concepts we use later in the paper. Given a metric $M = (V, d)$, a *c-spanner* is a graph $G = (V, E)$ such that the shortest-path distances in G (denoted by d_G) approximate the distances in d up to a factor of c: i.e., for all $x, y \in V$, $d(x, y) \leq d_G(x, y) \leq c \cdot d(x, y)$. This factor c is called the *stretch* of the spanner, and the two quantities of interest for a spanner are the stretch and the number of edges $|E|$ in the spanner. For motivations behind spanners, see [PS89, ADD$^+$93, CDNS95].

In this section, we show that metrics with bounded correlation dimension admit $(1+\varepsilon)$-stretch spanners with $O_\varepsilon(\min\{n^{1.5}, n\log\Delta\})$ edges, where Δ is the aspect ratio of the metric. This result should be contrasted with a trivial lower bound for general metrics: any spanner with stretch less than three for the shortest path metric for $K_{n,n}$ requires $\Omega(n^2)$ edges.

4.1. Upper bound for spanners. We have the following upper bound.

THEOREM 4.1 (sparse spanner theorem). *Given a metric $M = (V, d)$ with $\dim_C(M) \le k$, and given $\varepsilon > 0$, there exists a polynomial time algorithm that outputs a $(1+\varepsilon)$-spanner with $(2+\frac{1}{\varepsilon})^{O(k)} \min\{n^{1.5}, n \log \Delta\}$ edges.*

The algorithm for constructing sparse spanners for metrics with bounded correlation dimension is the same as that for doubling metrics in [CGMZ05],[2] though the proofs are different. For completeness, we briefly describe the algorithm from that paper. Given a metric (V, d) with intervertex distance at least 1 and a parameter $\varepsilon > 0$, define two parameters, $\gamma := 4 + \frac{32}{\varepsilon}$ and $p := \lceil \log_2 \gamma \rceil + 1$. Define $Y_{-p} := V$. For $i > -p$, let Y_i be a 2^i-net of Y_{i-1}; hence these nets are nested. (Note that since the intervertex distance is at least 1, $Y_i = V$ for $-p \le i < 0$.) For each net Y_i in the sequence, add edges between vertices which belong to the net Y_i and are "close together." In particular, for $i \ge -p$, define the edges at level i to be $E_i = \{(u,v) \in Y_i \times Y_i \mid \gamma \cdot 2^{i-1} < d(u,v) \le \gamma \cdot 2^i\}$. The union of all these edge sets $\widehat{E} = \cup_i E_i$ is the spanner returned by the construction. The following lemma (proved in [CGMZ05]) states that the spanner $(V, \widehat{E})$ preserves distances well.

LEMMA 4.2 (low stretch). *The graph $(V, \widehat{E})$ forms a $(1+\varepsilon)$-spanner for (V, d).*

LEMMA 4.3. *If the metric (V, d) has correlation dimension at most k, the number of edges in E_i is at most $2^{kp}|Y_i|$. Hence $|\widehat{E}| \le (2+\frac{1}{\varepsilon})^{O(k)} n \log \Delta$.*

Proof. Observe that $|E_i| \le \sum_{v \in Y_i} |\mathbf{B}_{Y_i}(v, \gamma \cdot 2^i)|$. By repeatedly using Definition 2.1 for correlation dimension, and the fact that $p = \lceil \log_2 \gamma \rceil + 1$, it follows that the sum is bounded by $2^{kp}|Y_i|$. Now, since each $|Y_i| \le n$ and $2^p = O(\varepsilon^{-1})$, summing this bound over all i implies that $\widehat{E}$ has at most $n \log \Delta \cdot \varepsilon^{-O(k)}$ edges, where Δ is the aspect ratio of the metric. ∎

This proves half of Theorem 4.1; we now need to prove the other half of the bound. The following lemma shows that if there are many edges in E_i, then a large number of points in the net Y_i would no longer belong to the net Y_{i+p}.

LEMMA 4.4. *Let $U := Y_i \setminus Y_{i+p}$ be the points in Y_i that do not belong to the net Y_{i+p}. Then, the number of edges $|E_i| \le \frac{1}{2}|U|(|U|+1)$.*

Proof. By the construction of the edge set E_i, note that if $(u,v) \in E_i$, then $d(u,v) \le \gamma \cdot 2^i$. Since $2^p > \gamma$, at most one of the two vertices $\{u, v\}$ can still be in 2^{i+p}-net Y_{i+p}, and hence any edge in E_i must have *at least* one endpoint in U. Now consider any $u \in U$. If both (x,u) and (y,u) are in E_i, then $d(x,y) \le \gamma \cdot 2^{i+1}$. Hence, at most one of $\{x, y\}$ can survive in Y_{i+p}. Thus, for each node $u \in U$, there can be *at most* one edge in E_i connecting to a point outside U; all other edges in E_i having u as one endpoint must have some other vertex in U as their other endpoint. It follows that $|E_i| \le \binom{|U|}{2} + |U|$, which completes the proof. ∎

LEMMA 4.5. *For any $r \in \{0, 1, \ldots, p-1\}$, the number of edges in all the E_i's with $i \equiv r \pmod{p}$ is $\sum_j |E_{jp+r}| \le O(2^{kp/2} n^{1.5})$.*

Proof. We want to find a function $F(\cdot)$ such that for any j_0, if $|Y_{j_0 p+r}| = a$, then $\sum_{j \ge j_0} |E_{jp+r}| \le F(a)$; we want to find the sharpest upper bound function $F(\cdot)$ possible. Lemma 4.4 implies that if $|U| = b$, then $F(a) \le \max_b\{\frac{1}{2}b(b+1) + F(a-b)\}$. Note that the right-hand side is maximized when b is maximized; however, the value of $b = |U|$ cannot be too large, since by Lemma 4.3 we have $|E_{j_0 p+r}| \le 2^{kp} a$. Putting these together forces $F(a) \le 2^{kp} a + F(a - 2^{kp/2}\sqrt{a})$ and implies that $F(a) = O(2^{kp/2} a^{1.5})$. Since any $|Y_i| \le n$, the result follows. ∎

[2]Using well-separated pair decomposition from [Tal04], a similar spanner construction can be obtained.

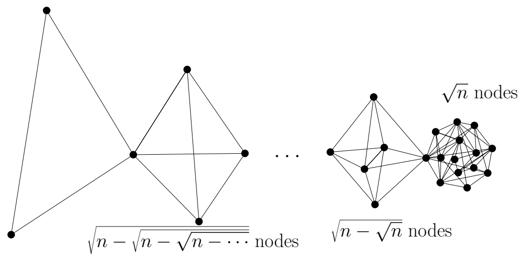


FIG. 4.1. *Lower bound example for sparse spanners.*

Applying Lemma 4.5 for each $r \in [p]$ and summing up the resulting bounds gives us $|\widehat{E}| \leq O(2^{kp/2}pn^{1.5}) \leq (2+\frac{1}{\varepsilon})^{O(k)}n^{1.5}$, proving the second part of Theorem 4.1.

We contrast this with the result that metrics with doubling dimension k admit $(1+\varepsilon)$-spanners with $O(n\varepsilon^{-O(k)})$ edges [CGMZ05, HPM05]. In the following section, we show that such a result is not possible with bounded correlation dimension and that the upper bound in Theorem 4.1 is indeed tight.

4.2. Lower bound for spanners. We have the following lower bound.

THEOREM 4.6 (lower bound on sparsity). *For every n, there exists a metric with bounded correlation dimension such that any 1.5-stretch spanner has $\Omega(n^{1.5})$ edges.*

The metric in the lower bound is supported on a graph such as that given in Figure 4.1. Before we give the details of the construction, let us note that it is essential that this lower bound metric have superpolynomial aspect ratio Δ, for we can obtain such a spanner with $O(n \log \Delta)$ edges from Theorem 4.1. Now we give the formal construction: let $\beta \geq 4$ be a parameter which specifies the difference in distance scales in different levels of the recursive construction. We define the construction via an algorithm that takes an integer n, the number of points in the metric, and a positive real $\alpha > 0$, the minimum distance in the metric. We denote the corresponding metric by $M(n, \alpha)$. For ease of exposition, we omit all ceilings or floors from the description, and moreover, each $M(n, \alpha)$ has a special node u.

Construction for $M(n, \alpha)$.

1. If n is less than some threshold n_0 (say, 10), then return a uniform metric of n points with interpoint distance α; set u to be any point.
2. Otherwise, recursively construct $M' := M(n-\sqrt{n}, \alpha\beta)$, together with the special point u'. Replace u' with a uniform metric U with $\sqrt{n}+1$ points with interpoint distance α. Each point in U has as distance to any other point the same as that from u'. Set the special point u to be any point in U.

LEMMA 4.7. *For all $n \geq 1$, the metric $M(n, 1)$ has correlation dimension $O(1)$.*

Proof. Let N be an R-net of $M(n, 1)$, where $\beta^{i-1} \leq R < \beta^i$. Note that by our construction, we have $N = M(n_i, \beta^i)$ for some n_i. Let u_i be the corresponding special point. We can assume $u_i \in N$. Consider $r \geq R/2$. There are four simple cases:

(1) If $2r < \beta^i$, then trivially we have

$$\sum_{x \in N} |\mathbf{B}_N(x, 2r)| = n_i = \sum_{x \in N} |\mathbf{B}_N(x, r)|.$$

(2) If $2r \geq \beta^i > r$, then we have

$$\sum_{x \in N} |\mathbf{B}_N(x, 2r)| = (\sqrt{n_i}+1)^2 + (n_i - \sqrt{n_i} - 1) \leq 3 \sum_{x \in N} |\mathbf{B}_N(x, r)|.$$

(3) Consider $2r \geq \beta^{\widehat{i}} > r$, where $\widehat{i} > i$. Define the quantities $p := |\mathbf{B}_N(u_i, r)|$ and $q := |\mathbf{B}_N(u_i, 2r) \backslash \mathbf{B}_N(u_i, r)|$. Note $p \geq \sqrt{n_i}$ and $q \leq \sqrt{n_i}$. Hence,

$$\sum_{x \in N} |\mathbf{B}_N(x, 2r)| = (p+q)^2 + (n_i - p - q) \leq 2(p^2 + q^2) + (n_i - p - q)$$
$$\leq 3(p^2 + n_i - p) = 3 \sum_{x \in N} |\mathbf{B}_N(x, r)|.$$

(4) Consider $\beta^{\widehat{i}+1} > 2r > r \geq \beta^{\widehat{i}}$, where $\widehat{i} \geq i$. Then, $p := |\mathbf{B}_N(u_i, 2r)| = |\mathbf{B}_N(u_i, r)|$. Hence,

$$\sum_{x \in N} |\mathbf{B}_N(x, 2r)| = p^2 + n_i - p = \sum_{x \in N} |\mathbf{B}_N(x, r)|.$$

Hence, any net of the metric $M(n, 1)$ satisfies (2.1). ▯

THEOREM 4.8. *Any 1.5-spanner for $M(n, 1)$ must have at least $\Omega(n^{1.5})$ edges.*

Proof. Let $h(n)$ be the size of a sparsest 1.5-spanner H for $M(n, 1)$. Observe that $M(n, 1)$ contains a uniform metric U of size $\sqrt{n} + 1$. Hence, there must be an edge in H between any two vertices in U. Suppose we contract U to a single point in H. Then, the resulting graph is a 1.5-spanner for $M(n - \sqrt{n}, \beta)$, and hence it contains at least $h(n - \sqrt{n})$ edges. Hence, we have $h(n) \geq (\sqrt{n} + 1)^2 + h(n - \sqrt{n})$. Solving the recurrence, we have $h(n) \geq \Omega(n^{1.5})$. ▯

5. Algorithms for metrics with bounded correlation dimension. Given an $\varepsilon \leq 1$, we consider randomized $(1 + \varepsilon)$-approximation algorithms for TSP on a metric $M = (V, d)$ on n points and $\dim_C = k$. Let OPT be the cost of the optimal TSP.

Our first algorithm returns a tour with expected cost at most $(1 + \varepsilon)\mathsf{OPT}$, in time $2^{\sqrt{n} \cdot (\varepsilon^{-1} \log n)^{O(\dim_C)}}$. As a by-product of our techniques, we also get Theorem 1.2: an embedding of the original metric into distribution of graphs with treewidth $\sqrt{n} \cdot (\varepsilon^{-1} \log n)^{O(\dim_C)}$. To prove these results, we extend the ideas of Arora [Aro02] and Talwar [Tal04] for TSP on doubling metrics. The main conceptual hurdle we have to overcome is the use of so-called "$O(1)$-separating decompositions" in these previous proofs; these are randomized decompositions of the metric into portions of diameter D such that two points at distance d lie in different clusters with probability $O(d/D)$. Metrics with small $\dim_C$ do not necessarily admit such separating decompositions, so instead we give a new decomposition procedure for such metrics.

For the second algorithm, which appears in section 6, we show how to improve our decomposition procedure, and use an improved global charging scheme to get our main result Theorem 1.3: an algorithm for TSP that returns a tour of expected cost $(1 + \epsilon)\mathsf{OPT}$ and runs in subexponential time.

5.1. An algorithm for TSP in time $\exp(\tilde{O}(\sqrt{n}))$. We assume that the aspect ratio of the input metric is $O(n/\varepsilon)$; this can be achieved by the following simple preprocessing step similar to that used in [Aro02, Tal04]. Suppose Δ is the diameter of the metric M. Let V_a be an $\varepsilon_a \Delta/n$-net of M. Suppose OPT_a is the length of an optimal tour just for the points in V_a. Then, it follows that $\mathsf{OPT}_a \leq \mathsf{OPT}$. From an optimal tour for the points in V_a, we can construct a tour for all points in V, with extra length at most $n \cdot 2\varepsilon_a \Delta/n = \varepsilon_a \cdot 2\Delta \leq \varepsilon_a \mathsf{OPT}$. Hence, we will assume that $V_a = V$ and that our metric has an aspect ratio of at most n/ε. We will also assume that the closest pair of points is at unit distance, and hence the aspect ratio equals the largest distance, which will be denoted by Δ. Moreover, we assume that $\varepsilon > 1/n$, or else we can solve TSP exactly in $2^{O(1/\varepsilon)}$-time.

The (Q)PTASs for TSP [Aro02, Tal04] are based on the following general conceptual steps. (a) These algorithms first find a good probabilistic hierarchical decomposition into clusters with geometrically decreasing diameters. (b) They choose a small set of *portals* in each cluster in this decomposition by taking a suitably fine net of the cluster, and force the tour to enter and leave the cluster using only these portals—such a tour is called a *portal-respecting* tour. The main structure lemma of these papers shows that the expected cost of the best portal-respecting tour is at most $(1+\varepsilon)$ times its original cost. (c) Finally, the algorithms find the best portal-respecting tour using dynamic programming. Specifically, for a cluster C, if there are only B portals among all its child clusters, the time to build the table for C is at most $B^{O(B)} = 2^{O(B \log B)}$. (See, e.g., section 5.3.) Since the total number of clusters is poly(n), total run-time is poly$(n)2^{O(B \log B)}$. For doubling metrics, each cluster has only $2^{O(\dim_D)}$ child clusters, each with $O(\varepsilon^{-1} \log n)^{O(\dim_D)}$ portals; the run-time is quasi-polynomial [Tal04].

Extending this approach to metrics with small correlation dimension, we face two immediate problems. First, metrics with low correlation dimension do not admit $O(1)$-separating decompositions which are traditionally used in step (a) to obtain probabilistic hierarchical decompositions. Second, while we can ensure that the number of portals in any single cluster is bounded by $O(\sqrt{n})$ using Lemma 2.2, each cluster may have as many as $\sqrt{n}$ child clusters, and hence the number of portals for all the child clusters may be close to $\Theta(n)$. To take care of these problems, we give a new partitioning and portaling scheme such that the union of the portals in each cluster and all its child clusters is only $\tilde{O}(\sqrt{n})$; this requires the partitioning and portal-creation steps to happen in a dependent fashion.

We will need several more definitions for the following algorithms.

Hierarchical decomposition. We consider sequences of partitions of the metric $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_L$, where $L := \lceil \log_H(n/\varepsilon) \rceil$. These partitions have the following properties: they define $D_L \geq \Delta$ and $D_{i-1} := D_i/H$, where $H \geq 4$ is some parameter that can possibly depend on n. For each i, $\mathcal{P}_i$ will be a partition of the point set V into *clusters* such that each cluster in the partition has diameter at most D_i, and such a cluster is said to have *height* i. Note that $\mathcal{P}_L$ consists of just one cluster containing all points in V, and the partition $\mathcal{P}_0$ has each point in V in its own separate cluster. The sequence of partitions $\{\mathcal{P}_i\}_i$ is a *hierarchical decomposition* if each height-i cluster is contained in some height-$(i+1)$ cluster. Given a partition $\mathcal{P}$ and a point z, we use $\mathcal{P}(z)$ to denote the cluster in $\mathcal{P}$ that contains z.

Portal assignment. Given a hierarchical decomposition, for each $0 \leq i \leq L$ and each height-i cluster C, we associate cluster C with a subset $U(C)$ of points within C known as *portals*. We impose the condition that a portal for a height-i cluster is also a portal for some height-j cluster, for all $j \leq i$. A *child portal* for a cluster $C \in \mathcal{P}_i$ is a portal in one of its child clusters, i.e., a portal in some cluster $C' \in \mathcal{P}_{i-1}$ such that $C' \subseteq C$. An assignment of portals to a set of hierarchical clusters is called *β-proper* if for all i, for all height-i clusters C, the portal set $U(C)$ is a βD_i-cover for the set of child portals of C.

A path or tour satisfies the *portal condition* (or is *portal-respecting*) if it only enters or leaves a cluster through its portals.

α-padded decomposition. Given a metric (V, d) and a diameter bound D, an α-padded decomposition is a randomized polynomial-time algorithm that outputs a random partition with each cluster in this partition having diameter at most D,

and moreover, for every $u, v \in V$ and every $r \geq 0$, the probability that the union $\mathbf{B}(u,r) \cup \mathbf{B}(v,r)$ is not contained within a single cluster is at most $\alpha \cdot \frac{d(u,v)+2r}{D}$. It is known that algorithms due to Bartal [Bar96] and Fakcharoenphol, Rao, and Talwar [FRT04] allow us to sample from α-padded decompositions with $\alpha = O(\log n)$. In particular, we will assume the existence of an algorithm Bartal(V, D) with the following property.

Fact 5.1. *There exists a universal constant $t > 0$ such that given any n point metric space (V, d) and a parameter $D > 0$, the algorithm* Bartal *outputs a random decomposition into clusters with diameter at most D, such that for all points u, v, and $r > 0$, the probability that $B(u,r) \cup B(v,r)$ is separated is at most*

$$t \log n \cdot \frac{d(u,v) + 2r}{D}.$$

A closely related notion is that of an *α-separating decomposition*, where we require only that for any $u, v \in V$, the probability that u, v do not lie in the same cluster is at most $\alpha \frac{d(u,v)}{D}$. Note that this corresponds to setting $r = 0$ in the definition of an α-padded decomposition; hence, every padded decomposition is also a separating decomposition.

α-separating hierarchical decomposition. An *α-separating hierarchical decomposition* is a randomized polynomial-time algorithm that outputs a random hierarchical decomposition $\{P_i\}_i$ of the metric such that the height-i decomposition $\mathcal{P}_i$ is an α-separating decomposition simultaneously for each value $i \in [L]$. The aforementioned algorithms of [Bar96, FRT04] can be used to obtain α-separating hierarchical decompositions with $\alpha = O(\log n)$.

Given these definitions, we can now state a general lemma indicating the utility of good separating hierarchical decompositions and portal assignments.

Lemma 5.2. *Suppose $\{\mathcal{P}_i\}$ is drawn from an α-separating hierarchical decomposition of (V, d) with an associated β-proper portal assignment. Then, for any $u, v \in V$, the expected length of the shortest portal-respecting path between u and v is at most $(1 + 6L\alpha\beta) \cdot d(u,v)$, where the expectation is over the randomness of the hierarchical decomposition.*

Proof. Consider the event that u and v are separated in $\mathcal{P}_i$ but not separated in $\mathcal{P}_{i+1}$. The probability of this event is at most $\alpha \cdot d(u,v)/D_i$. We show in the following that under this event, the shortest path from u to v satisfying the portal condition is at most $6\beta D_i + d(u,v)$; i.e., the extra distance traveled from u to v is at most $6\beta D_i$.

For node u, we define a series of portals in the following way. Let $u_0 := u$. For $j \geq 1$, let u_j be the closest height-j portal to u_{j-1} in the height-j cluster containing u_{j-1}. Since we assume that the portal assignment is proper, it is possible to go directly from u_{j-1} to u_j, and $d(u_{j-1}, u_j) \leq \beta D_j$. By the triangle inequality, $d(u, u_i) \leq \sum_{j=1}^{i} \beta D_j \leq \frac{4}{3}\beta D_i$.

We define v_j's similarly for v, and we also have $d(v, v_i) \leq \frac{4}{3}\beta D_i$.

Observe that we have a portal-respecting tour from u to v in the following way. We start from $u = u_0, u_1, \ldots, u_i$ and then go across to $v_i, v_{i-1}, \ldots, v_1, v_0 = v$. Observe that $d(u_i, v_i) \leq 2 \cdot \frac{4}{3}\beta D_i + d(u,v)$. The total distance is at most $d(u,v) + 4 \cdot \frac{4}{3}\beta D_i \leq d(u,v) + 6\beta D_i$.

Next, summing over all heights i, we can show that the expected extra distance traveled between u and v is at most $\sum_{i=0}^{L} \alpha \cdot \frac{d(u,v)}{D_i} \cdot 6\beta D_i \leq 6L\alpha\beta \cdot d(u,v)$. ∎

Hence, using Lemma 5.2, we can show that even if we require the tour to satisfy the portal condition, the length of the resulting optimal tour does not increase too much.

5.2. A partitioning and portaling algorithm. In order to apply Lemma 5.2, we need to construct an α-separating hierarchical decomposition and an associated β-proper portaling scheme such that both α and β are small; moreover, to use dynamic programming to find the best portal-respecting tour, we want the number of child portals B for each cluster to be small too. In this section, we show how to achieve this.

To start off, observe that if the child portals of each cluster form a packing, then Lemma 2.2 implies that B is small for each cluster. However, using a standard hierarchical decomposition (e.g., those due to [Bar96, FRT04]) and choosing an arbitrary net for each cluster as the set of portal does not imply this property, because the portals near the boundary of two different clusters might be too close together. We resolve this by using Bartal's decomposition [Bar96] twice: after obtaining a standard decomposition, we apply the decomposition technique again to make minor adjustments to the boundaries of clusters. Here is the main result listing the properties of our hierarchical decomposition and portaling scheme.

THEOREM 5.3. *Given a metric (V, d) and a parameter $\beta \leq 1$, there is a randomized polynomial-time algorithm to generate hierarchical decompositions of the metric such that*

(A1) *the diameter of a height-i cluster is guaranteed to be at most $(1 + 2\beta)D_i$, where $D_i = 4^i$, and*

(A2) *for every $u, v \in V$, the probability that u, v fall in different clusters at height-i but fall in the same height-$(i + 1)$ cluster is at most $O(\log^2 n) \times \frac{d(u,v)}{D_i}$.*

Moreover, each cluster C is assigned portals *$U(C)$ such that*

(B1) *(β-proper) for each nonroot cluster C at height-i (i.e., $i < L$), the set of portals $U(C)$ forms a βD_i-cover of C (and hence also a βD_i-cover for its child portals).*

(B2) *Moreover, the set of portals in C and in all its children (which are height-$(i - 1)$ clusters) forms a $(\beta/4) D_{i-1}$-packing.*

The proof of Theorem 5.3 appears in section 5.2.1, the TSP algorithm obtained using this theorem appears in section 5.3, and the application of this theorem to embedding metrics with bounded correlation dimension into small treewidth graphs appears in section 5.4.

5.2.1. The proof of Theorem 5.3. Let us first present the algorithm claimed in Theorem 5.3. Recall that the metric (V, d) has unit minimum distance and aspect ratio $\Delta = O(n/\varepsilon)$. Define $H := 4$ and $L := \lceil \log_H \Delta \rceil$. Set $D_L := \Delta$ and $D_{i-1} := D_i/4$. As mentioned above, the hierarchical decomposition is constructed in a top-down fashion. Each parent cluster is partitioned into child clusters using Bartal's decomposition twice; the details appear in Algorithm 1.

Let us first show that the portal assignment satisfies the properties required in Theorem 5.3(B1)–(B2).

LEMMA 5.4 (covering/packing properties). *For $i < L$, for any height-$(i + 1)$ cluster C produced by the decomposition algorithm, the following two properties hold:*

(1) *For any child cluster C' of C, the set $U(C')$ is a βD_i-cover of C'; and*

(2) *the union of $U(C')$'s, over all the child clusters C' of C is a $\frac{1}{4}\beta D_i$-packing.*

Proof. We show that if the portal set $U(C)$ is a $\frac{1}{4}\beta D_{i+1}$-packing for a height-$(i+1)$ cluster C, then for any child cluster C' of C, the portal set $U(C')$ is a βD_i-cover of C', and the union of $U(C')$'s over all the child clusters C' of C is a $\frac{1}{4}\beta D_i$-packing. The lemma then follows by induction on i, the base case following from the fact that

Algorithm 1 An algorithm for hierarchical decompositions and portal assignment.

1: **let** $\mathcal{P}_L \leftarrow \{V\}$ and $U(V) \leftarrow \emptyset$.
2: **for** $i = L-1$ down to 0 **do**
3:   **for** each height-$(i+1)$ cluster $C \in \mathcal{P}_{i+1}$ **do**
4:     $\widetilde{\mathcal{P}}_i \leftarrow$ Bartal(points C, diameter bound D_i)
    {**note:** we inductively ensure $U(C)$ is a $\frac{1}{4}\beta D_{i+1} = \beta D_i$-packing.}
5:     augment $U(C)$ to obtain a βD_i-net $\widehat{U}(C)$ of C
6:     **let** $Z \leftarrow \{z \in C \mid d(z, \widehat{U}(C) \cap \widetilde{\mathcal{P}}_i(z)) > \beta D_i\}$ be the points that are far from the points in $\widehat{U}(C)$ within their height-i cluster
7:     **let** $\mathcal{L} \leftarrow Z$, $G \leftarrow C$, and $\overline{U}(C) \leftarrow \emptyset$
    [Boundary Adjustment]
8:     **while** $\mathcal{L} \neq \emptyset$ **do**
9:       **let** $u \leftarrow$ an arbitrary point in $\mathcal{L}$, $r \leftarrow \beta D_i / 4 \ln n$.
10:       **pick** $z \in [0, \beta D_i/4]$ randomly with probability density function $p(z) := \frac{n}{n-1} \cdot \frac{1}{r} e^{-z/r}$.
11:       **let** $\mathbf{B}_u \leftarrow \mathbf{B}(u, \beta D_i/4 + z)$.
12:       **if** $\mathbf{B}_u$ contains some point c in $\widehat{U}(C)$ **then**
13:         **move** all points in $\mathbf{B}_u \cap G$ to the height-i cluster currently containing c
14:       **else**
15:         **let** $\overline{U}(C) \leftarrow \overline{U}(C) \cup \{u\}$
16:         **move** all points in $\mathbf{B}_u \cap G$ to the height-i cluster currently containing u
17:       **end if**
18:       **let** $G \leftarrow G \setminus \mathbf{B}_u$, $\mathcal{L} \leftarrow \mathcal{L} \setminus \mathbf{B}_u$
19:     **end while**
20:     **let** the new partition on C be $\mathcal{P}_i$.
21:     **for** each new height-i cluster C' **do**
22:       **let** $U(C') := C' \cap (\widehat{U}(C) \cup \overline{U}(C))$
23:     **end for**
24:   **end for**
25: **end for**

the only height-L cluster is the entire point set V, and the empty set $U(V)$ is a trivial $\frac{1}{4}\beta D_L$-packing.

Suppose C is a height-$(i+1)$ cluster returned by the algorithm and the corresponding $U(C)$ is a $\frac{1}{4}\beta D_{i+1}$-packing. We first show the covering property for each child cluster C' of C. Since $D_{i+1} = HD_i = 4D_i$, the subset $U(C)$ is a βD_i-packing; it can indeed be augmented to a βD_i-net $\widehat{U}(C)$ for C in step 5. Observe that points in $\widehat{U}(C)$ are not reassigned to different height-i clusters in the boundary adjustment steps (the **while** loop in steps 8–19).

Let x be a point in C. We claim there is a point in $\widehat{U}(C) \cup \overline{U}(C)$ that is in the same height-i cluster induced by $\mathcal{P}_i$ and also within distance βD_i of x.

There are two cases: The first is that x is not in the set Z (which was defined in step 6). Then, there is a point $v \in \widehat{U}(C) \cap \widetilde{\mathcal{P}}_i(x)$ such that $d(x, v) \leq \beta D_i$. Note that points in $\widehat{U}(C)$ are not reassigned—so if point x is also not reassigned to another height-i cluster, it will still be covered by the point v after boundary adjustment. On the other hand, suppose point x is in some ball $\mathbf{B}_u$ (created at step 11) with diameter at most βD_i, which contains a point in $\widehat{U}(C) \cup \overline{U}(C)$. At step 16, all points in $\mathbf{B}_u$ will be removed from G and stay in the same height-i cluster throughout the boundary adjustment process.

The second case is that x is in Z: then eventually x must be removed from list $\mathcal{L}$. Then, by a similar argument, at some moment x must be in some ball $\mathbf{B}_u$ with diameter at most βD_i, which contains a point in $\widehat{U}(C) \cup \overline{U}(C)$. The same argument follows.

To complete the proof, we next show that the union of $U(C')$'s, which is the same as $\widehat{U}(C) \cup \overline{U}(C)$, is a $\frac{1}{4}\beta D_i$-packing. First, observe that $\widehat{U}(C)$ is a βD_i-net and so is trivially also a $\frac{1}{4}\beta D_i$-packing. Next, observe that whenever a new point u is added to $\overline{U}(C)$, it must be at a distance of more than $\frac{1}{4}\beta D_i$ from $\widehat{U}(C)$ and existing points in $\overline{U}(C)$. Hence, the packing property follows. $\square$

Now we begin to give an upper bound on the separation probability from Theorem 5.3(A2). We use the following fact, which follows immediately using techniques in [Bar96, section 3].

LEMMA 5.5. *Suppose $x, y \in V$. Then the probability that exactly one of x and y lies in some $\mathbf{B}_u$ during some execution of the* ***while*** *loop (in steps* 8–19*) is at most $O(\log n) \cdot \frac{d(x,y)}{\beta D_i}$, and this is independent of the randomness in step* 4.

Proof. We use the algorithm in [Bar96, section 3]. Observe that in the **while** loop (in steps 8–19), we grow balls of radii at most $\frac{\beta D_i}{4}$ with centers from Z.

Hence, by the same analysis as Bartal's, the probability that the points x and y are separated by some ball $\mathbf{B}_u$ is at most $O(\log |Z|) \cdot \frac{d(x,y)}{\beta D_i}$, from which the result follows because $|Z| \le n$. $\square$

LEMMA 5.6 (separation probability). *For each level i,* $\Pr[(x, y)$ *separated by* $\mathcal{P}_i \mid (x, y)$ *not separated by* $\mathcal{P}_{i+1}] \le O(\log^2 n)\frac{d(x,y)}{D_i}$.

Observe that the probability that (x, y) is *first* separated at level i is upper bounded by this conditional probability, and hence this proves Theorem 5.3(A2).

Proof. Consider $x, y \in V$. Let B_x and B_y be the balls centered at x and y, respectively, with radius βD_i. We claim that if $B_x \cup B_y$ is not separated by $\widetilde{\mathcal{P}_i}$, then x and y cannot be separated by $\mathcal{P}_i$. The reason is as follows: since B_x is contained in one cluster in $\widetilde{\mathcal{P}_i}$, the point x has a net point in $\widehat{U}(C)$ in the same cluster within distance βD_i. Moreover, x is at a distance of at least βD_i from any point from another cluster. Hence, it follows that the point x will remain in the same cluster during the boundary adjustment process. By the same argument for y, it follows that x and y cannot be separated by $\mathcal{P}_i$.

First, consider the case when $d(x, y) \ge \beta D_i$. Hence, by the observation above, the probability that $\mathcal{P}_i$ separates x and y is upper bounded by that of the event that $\widetilde{\mathcal{P}_i}$ separates $B_x \cup B_y$, which is at most $O(\log n) \cdot \frac{d(u,v)+2\beta D_i}{D_i}$, by Fact 5.1. By the assumption that $d(x, y) \ge \beta D_i$, the probability is at most $O(\log n) \cdot \frac{d(x,y)}{D_i}$.

Next consider the case when $d(x, y) < \beta D_i$. By the same argument as before, if x and y are separated by $\mathcal{P}_i$, then $B_x \cup B_y$ must be separated by $\widetilde{\mathcal{P}}_i$. The event that x and y are separated by $\mathcal{P}_i$ implies one of the following two events:

(1) x and y are separated by $\widetilde{\mathcal{P}}_i$, or

(2) x and y are not separated by $\widetilde{\mathcal{P}}_i$ but are separated in the boundary adjustment process in the **while** loop.

The probability of event (1) is at most $O(\log n) \cdot \frac{d(x,y)}{D_i}$. Note that event (2) implies that the union of B_x and B_y is separated by $\widetilde{\mathcal{P}}_i$, and also that x and y are separated in the boundary adjustment step. Since these are independent, we can bound the first probability using Fact 5.1, and the second using Lemma 5.5, and hence we can upper bound the probability of event (2) by the following expression:

$$O(\log n) \cdot \frac{d(x,y) + 2\beta D_i}{D_i} \cdot O(\log n)\frac{d(x,y)}{\beta D_i} \le O(\log^2 n)\frac{d(x,y)}{D_i}.$$

This completes the proof. ∎

Lemma 5.7 (proof of Theorem 5.3(A1)). *The diameter of each cluster in $\mathcal{P}_i$ is at most $(1+2\beta)D_i$.*

Proof. In step 4, the diameter of each cluster in $\widetilde{\mathcal{P}}_i$ is at most D_i; in the boundary adjustment step, balls of diameter at most βD_i can be added to a cluster. Hence, the diameter of a cluster in $\mathcal{P}_i$ is at most $(1+2\beta)D_i$. ∎

Lemmas 5.4, 5.6, and 5.7 complete the proof of Theorem 5.3. Finally, let us record a useful corollary of Theorem 5.3.

Corollary 5.8. *The hierarchical decomposition produced by Theorem* 5.3 *is α-separating with $\alpha = (1+2\beta)\cdot O(\log^2 n)$, and the portal assignment is β-proper.*

Proof. Note that the diameter of the height-i clusters is bounded by $D'_i = D_i(1+2\beta)$. Hence, we would like to find a value α such that the probability of u, v being separated at height-i is bounded by $\alpha\frac{d(u,v)}{D'_i}$. Theorem 5.3(A2) bounds the probability that a pair of points is *first* separated at height i. However, if the pair is separated at height i, the pair must be first separated at some height j for some value $j \ge i$. Now a trivial union bound implies that the probability that a pair (u,v) is separated at height i is at most $\sum_{j\ge i} O(\log^2 n) \times \frac{d(u,v)}{D_j} \le O(\log^2 n) \times \frac{d(u,v)}{D_i}$, observing that the D_i's form a geometric series. Finally, since $D_i = D'_i/(1+2\beta)$, we get that the hierarchical decomposition is $O(\log^2 n)/(1+2\beta)$-separating. To end, we note that the portal assignment is a βD_i-cover, which also implies that it is a $\beta D'_i$-cover, since $D_i \le D'_i$. ∎

5.3. The first TSP algorithm. We now proceed to give our first algorithm for TSP. We will use Corollary 5.8, which gives us the properties of the hierarchical decomposition and the portal assignment. Moreover, since the child portals of every cluster form a packing, we can bound their number, as the following corollary shows.

Corollary 5.9. *Suppose the metric space (V,d) has correlation dimension at most k. For any cluster C created by Algorithm* 1*, the union of $U(C')$ over all child clusters C' of C has size at most $(16/\beta + 4)^{k/2}\sqrt{n}$.*

Proof. Suppose cluster C is at height $i+1$. By Theorem 5.3(B2), the union S of $U(C')$ over all child clusters C' of C is a $\frac{1}{4}\beta D_i$-packing. Hence, it can be extended to a $\frac{1}{4}\beta D_i$-net N for the whole space V. Observe that C has diameter at most $3D_{i+1}$. Since N is a $4^i\beta$-net, C is contained in a ball of radius at most $4^{i+1} + 4^i\beta$ centered at some net point $u \in N$. Hence, $S \subseteq \mathbf{B}_N(u, 4^{i+1} + 4^i\beta)$, which by Lemma 2.2 has size at most $(16/\beta+4)^{k/2}\sqrt{|N|} \le (16/\beta+4)^{k/2}\sqrt{n}$. ∎

5.3.1. The dynamic programming framework for solving TSP. We briefly outline a dynamic program to solve TSP given a hierarchical decomposition and its corresponding portals for each cluster. The basic idea is similar to the constructions used by Arnbourg and Proskurowski [AP89] and Arora [Aro02], and we give the details here for completeness.

We describe the dynamic programming table. For each cluster C with its portals $U(C)$, there are entries indexed by (J, I), where J is a set of unordered distinct pairs of portals from $U(C)$, and I is a subset of $U(C)$. Any portal that appears in a pair in J does not appear in I. Note that if $r = |U(C)|$, then there are at most $r!r^2$ such entries.

An entry index by (J, I) represents the scenario in which a tour visits the nonportals of cluster C using entry and exit portals described by pairs in J. Each point in I is visited while the tour remains in cluster C; hence, in this case, the points in I are not behaving as portals in the sense that for each point in I, the two points adjacent to it in the tour are in the cluster C. For each portal x in $U(C)$ that does not appear in J or I, the adjacent points in the tour are both not in $U(C)$; i.e., the tour enters the cluster through that portal x and leaves immediately afterward. We keep track of the length of the portion of the tour that is within the cluster C. More precisely, we count only the part of the tour that is between u and v for some pair $\{u, v\}$ in J. The entry indexed by (J, I) keeps the length of the shortest possible internal segments for tours consistent with the scenario imposed by (J, I). Note that if we have to construct the tour, under each entry we have to store the internal segments of the tour as well.

Finally, there are special entries, each of which is indexed by only a single portal $x \in U(C)$. This corresponds to the (suboptimal) case when we enter the cluster C through x, perform a tour visiting all points in C, and leave through x. The value of such an entry corresponds to the length of a tour for points in cluster C.

As shown in [AP89], if the number of child portals for a cluster is at most B, then the time to complete all entries for that cluster is $2^{O(B \log B)}$. Hence, if this holds for all clusters in the decomposition, the total running time is at most $nL \cdot 2^{O(B \log B)}$, though typically nL is absorbed in the exponential term.

5.3.2. The algorithm. We now use the framework along with Lemma 5.2 and Theorem 5.3 to give our first algorithm.

THEOREM 5.10. *For any constant* $0 < \varepsilon < 1$*, there is a randomized algorithm for metric TSP, which, for a metric* $M = (V, d)$ *with* $\dim_C(M) = k$*, returns a tour of expected length at most* $(1 + \varepsilon)\mathsf{OPT}$ *in time* $2^{((\log n)/\varepsilon)^{O(k)}\sqrt{n}}$.

Proof. By Corollary 5.8, the hierarchical decomposition is $O(\log^2 n)$-separating, and the portal assignment is β-proper. Also, the height of the decomposition is $L = O(\log \frac{n}{\varepsilon})$. Now, if we set $\beta := \frac{\varepsilon}{6L\alpha}$, Lemma 5.2 implies that the expected length of the optimal portal-respecting tour is at most $(1 + 6L\alpha\beta)\mathsf{OPT} = (1 + \varepsilon)\mathsf{OPT}$.

Finally, we need to bound the running time of the dynamic program: recall that an upper bound B for the number of portals in each cluster and its children would imply a $B^{O(B)}$ run-time. By Corollary 5.9, it follows that $B \leq (16/\beta + 4)^{k/2}\sqrt{n}$. Hence, the running time of the algorithm is $nL \cdot 2^{O(B \log B)} = \exp\{(\varepsilon^{-1} \log n)^{O(k)}\sqrt{n}\}$, as required. ∎

5.4. Embedding into small treewidth graphs. Observe that our probabilistic hierarchical decomposition procedure actually gives an embedding into a distribution of low treewidth graphs (see [Die00] for the definition of treewidth). Suppose we are given a particular hierarchical decomposition together with the portals for each cluster. We start with the complete weighted graph consistent with the metric and delete any edge that is going out of a cluster, but not via a portal. If the number of child portals for each cluster is at most B, then the treewidth of the resulting graph is at most B. From Lemma 5.2, the expected distortion of the distance between any pair of points is small. Using the same parameters as in the proof of Theorem 5.10, we have the following theorem.

THEOREM 5.11 (embedding into small treewidth graphs). *Given any constant* $0 < \varepsilon < 1$ *and* k*, a metric space of size* n *with correlation dimension at most* k *can be* *embedded with distortion* $1+\varepsilon$ *into a distribution of dominating graphs*[3] *with treewidth* $((\log n)/\varepsilon)^{O(k)}\sqrt{n})$.

6. A subexponential time $(1+\varepsilon)$-approximation for TSP. In the previous section, we saw how to get a $(1+\varepsilon)$-approximation algorithm for TSP on metrics with bounded correlation dimension, essentially using the idea of random embeddings into small treewidth graphs. This approach gives approximations for problems on metric spaces which can be solved for small-treewidth graphs. However, similar to the example of $\sqrt{n}$-lollipop graph, a metric with bounded correlation dimension can contain an arbitrary submetric of size $\sqrt{n}$. Hence, from the result of Carroll and Goel [CG08], we can conclude that in general randomly $(1+\varepsilon)$-approximating metrics with bounded correlation dimension require the use of graphs with treewidth of order $\sqrt{n}$.

In this section, we get an improved approximation for TSP using a different observation inspired by the examples in Figure 1.1: in those instances, the contribution to the optimal TSP due to the dense substructure is much smaller than that from the low-dimensional ambient structure. For example, for the subgrid with an $(1,2)$-TSP instance attached to it (Figure 1.1(b)), we can obtain a $(1+\varepsilon)$-approximation to TSP on the grid (which contributes about $\Theta(n)$ to OPT) and stitch it together with a naïve 2-approximation to the hard instance (which contributes only $\Theta(\sqrt{n})$ to OPT). Metrics with small correlation dimension do not admit such trivial clusterings in general, and hence our algorithm must do some kind of clustering for all instances. Moreover, this indicates that we need a global accounting scheme: the poor approximation obtained for the hard subproblem needs to be charged to the entire OPT and not just the optimal tour on the subproblem. Following are some of the issues we need to address, along with high-level descriptions of how we handle them.

Avoiding large tables. The immediate hurdle to a better run-time is that some cluster may have $\Theta(\sqrt{n})$ child portals, and we have to spend $\exp(\sqrt{n}\log\sqrt{n})$ time to compute the tables. Our idea here is to set a threshold B_0 such that in the dynamic program, if a cluster has $B > B_0$ portals among its children, we compute, in linear time, a tour on C that *only enters and leaves C once*. This "patching" incurs an increase in length of a constant times the weight of a minimum spanning tree on the portals. We call this extra length the "MST-loss" and bound this trivially by $B\cdot\operatorname{diam}(C)$, the number of portals times the diameter of the cluster. This step implies that we need only spend $\max\{O(B), 2^{O(B_0\log B_0)}\}$ time on any table computation, at the expense of these MST-losses. The patching procedure is reminiscent of the patching from [Aro96] and is described in section 6.2.

Paying for the MST-loss. In contrast to previous works, the "MST-loss" due to patching cannot be charged locally, and hence we need to charge this to the cost of the global OPT. Moreover, we may need to account for the MST-loss occurring in many clusters. Hence we need to show that OPT is large enough, and the MST-loss is incurred infrequently enough, so that we can charge all the MST-losses over the entire run of the algorithm to $\varepsilon\mathsf{OPT}$. To charge MST-losses in a global manner, consider the hierarchical decomposition. The extra length incurred for patching height-i clusters is proportional to the number of child portals of the clusters for which patchings are applied. Now, if the union of *all of* the height-$(i-1)$ portals in the decomposition formed a good packing, Lemma 2.2 would bound their number, and hence also bound the total MST-loss at height-i of the decomposition tree. However, the techniques in the previous section can ensure only that the child portals of a *single cluster* form a

[3] A weighted graph dominates a metric space (with the same point set) if the shortest-path distance in the graph between every pair of points is at least that given by the metric space.

packing, and hence we need to give a new partitioning scheme and portal assignment algorithm.

A new partitioning and portaling procedure. The algorithm from Theorem 5.3 first decomposed a cluster C at height-$(i+1)$, and then adjusted the boundaries of the child clusters created at height-i so that their portals formed a packing. However, the union of the portals in the grandchildren of C (i.e., all the clusters at height-$(i-1)$ below C) may not form a packing: hence we have to readjust the boundaries created at height-i yet again. In general, when clusters at a certain level are created, the boundaries for clusters in all higher levels have to be readjusted. This increases the probability that a pair of points is separated at each level. To control this, we require that cluster diameters fall by logarithmic factors instead of by constants. The details are given in section 6.1.

Avoiding computation of correlation dimension. Finally, in the following, we assume that we know the value of the correlation dimension. Though Theorem 3.1 shows a hardness result, we can use a "guess-and-verify" strategy: we start with a small value as our guess for k, and for each net encountered during the run of the algorithm, we verify in polynomial time the bounded average growth rate property in (2.1). Whenever property (2.1) is violated for some net, we know the current estimation of the correlation dimension is too small, so we double the current guess for k and restart the algorithm. Since the correlation dimension is at most $O(\log n)$ and the running time is doubly exponential in k, the extra time incurred for trying out smaller values of k does not affect the asymptotic running time.

We formalize the ideas sketched above in the following. The general framework described in Lemma 5.2 of using a good hierarchical decomposition and portal assignment to approximate TSP still applies here. We give a more sophisticated partitioning and portaling algorithm in section 6.1 and analyze the MST-loss incurred from patching in section 6.2.2.

6.1. A modified partitioning and portaling algorithm. The algorithm in this section is similar to that in section 5.2: the main difference is that when a height-i partition is performed, all height-j partitions with $j < i$ are modified in order to ensure that all height-i portals form a packing. Again, let $H \geq 4$ be a parameter (possibly depending on n) to be determined later. Let $L := \lceil \log_H(n/\varepsilon) \rceil$. Set $D_L := \Delta$, the diameter of (V,d), and set $D_{i-1} := D_i/H$.

We now give a hierarchical decomposition of (V,d) in Algorithm 2; moreover, we associate a single set of portals U_i for each height i such that U_i is a $\frac{1}{4}\beta D_i$-packing, and moreover, for each height-i cluster C, the set $U(C) := U_i \cap C$ is a βD_i-cover of $U_{i-1} \cap C$, the child portals of C. We ensure that once a U_i is formed, it is not modified, and moreover, the height-i partition of U_i remains invariant throughout the remaining execution of the algorithm.

Analogous to Theorem 5.3, the hierarchical partition given by Algorithm 2 has the following properties.

THEOREM 6.1. *Given a metric (V,d) and a parameter $\beta \leq 1$, Algorithm 2 is a randomized polynomial-time algorithm that generates a hierarchical decomposition of the metric such that*

(A1) *the diameter of each height-i cluster is guaranteed to be at most $4D_i$, and*
(A2) *for every $u, v \in V$, the probability that u, v fall in different clusters at height-i is at most $O(\log n)^L \cdot \frac{d(u,v)}{D_i}$.*

Moreover, for each i, a set U_i of nodes is designated as height-i portals such that

Algorithm 2 A modified algorithm for hierarchical decompositions and portal assignment.

let $\mathcal{P}_L \leftarrow \{V\}$ and $U_L \leftarrow \emptyset$.
for $i = L-1$ down to 0 **do**
  for each height-$(i+1)$ cluster $C \in \mathcal{P}_{i+1}$ **do**
    $\mathcal{P}_i \leftarrow$ BARTAL(points C, diameter bound D_i)
  end for
  {**note:** we inductively ensure U_{i+1} is a $\frac{1}{4}\beta D_{i+1} = \frac{1}{4}\beta H D_i \geq \beta D_i$-packing.}
  augment U_{i+1} to obtain a βD_i-net $\widehat{U}_i$ of V
  let $Z \leftarrow \{z \in V \mid d(z, \widehat{U}_i \cap \mathcal{P}_i(z)) > \beta D_i\}$ be the points that are far from the points in $\widehat{U}_i$ within their height-i cluster
  let $\mathcal{L} \leftarrow Z$, $G \leftarrow V$, and $\overline{U}_i \leftarrow \emptyset$
  [Boundary Adjustment]
  while $\mathcal{L} \neq \emptyset$ **do**
    let $u \leftarrow$ an arbitrary point in $\mathcal{L}$, $r \leftarrow \beta D_i / 4 \ln n$.
    pick $z \in [0, \beta D_i/4]$ randomly with probability density function $p(z) := \frac{n}{n-1} \cdot \frac{1}{r} e^{-z/r}$.
    let $\mathbf{B}_u \leftarrow \mathbf{B}(u, \beta D_i/4 + z)$.
    if $\mathbf{B}_u$ contains some point c in $\widehat{U}_i$ **then**
      for each $j \geq i$ **do**
        Modify $\mathcal{P}_j$ by moving all points in $\mathbf{B}_u \cap G$ to the height-j cluster currently containing c
      end for
    else
      let $\overline{U}_i \leftarrow \overline{U}_i \cup \{u\}$
      for each $j \geq i$ **do**
        Modify $\mathcal{P}_j$ by moving all points in $\mathbf{B}_u \cap G$ to the height-j cluster currently containing u
      end for
    end if
    let $G \leftarrow G \setminus \mathbf{B}_u$, $\mathcal{L} \leftarrow \mathcal{L} \setminus \mathbf{B}_u$
  end while
  let $U_i \leftarrow \widehat{U}_i \cup \overline{U}_i$
  for each $j \geq i$ **do**
    let $\mathcal{P}_j^{(i)} \leftarrow \mathcal{P}_j$
  end for
  {**note:** the $\mathcal{P}_j^{(i)}$'s are only for analysis and are not necessary for the algorithm.}
end for
return $\{(\mathcal{P}_i, U_i)\}_{i \geq 0}$

(B1) *(β-proper) for each nonroot cluster C at height-i (i.e., $0 < i < L$), the set of portals in $U_i \cap C$ forms a $\beta\, D_i$-cover of $U_{i-1} \cap C$, the child portals of C.*

(B2) *Moreover, the set of portals U_i is a $\frac{1}{4}\beta D_i$-packing.*

The proof of the theorem follows from the subsequent lemmas: statement (A1) follows from Lemma 6.2, statement (A2) follows from Lemma 6.5, and statements (B1) and (B2) follow from Lemma 6.3.

LEMMA 6.2. *Each cluster in $\mathcal{P}_i$ has diameter at most $4D_i$.*

Proof. Using the same argument as in Lemma 5.7, each cluster in $\mathcal{P}_i$ has diameter at most $(1 + 2\beta)D_i$. After that, for each boundary adjustment step at height $j < i$, the diameter of a height-i cluster can increase by at most $2\beta D_j$. Observing that D_j decreases geometrically by a factor of at least 4 gives the result. ∎

LEMMA 6.3. *For each height i, the set U_i of height-i portals is a $\frac{1}{4}\beta D_i$-packing. Moreover, for each height-i cluster C in $\mathcal{P}_i$, the set $U_i \cap C$ of portals is a βD_i-cover of C.*

Proof. Using exactly the same argument as in Lemma 5.4, we obtain the following statement for $\mathcal{P}_i^{(i)}$: the set U_i is a $\frac{1}{4}\beta D_i$-packing, and for each height-i cluster C in $\mathcal{P}_i^{(i)}$, the set $U_i \cap C$ of portals is a βD_i-cover for C. Since U_i is not modified once it is created, it is still a $\frac{1}{4}\beta D_i$-packing after boundary adjustment steps for lower heights.

We next prove that after $\mathcal{P}_{i-1}^{(i-1)}$ is formed, for each height-i cluster C and each child cluster C' of C, $U_i(C) = C \cap U_i$ is a βD_i-cover for $U_{i-1}(C') = U_{i-1} \cap C'$. To do this, consider how U_{i-1} is formed. At step 6, $\widehat{U}_{i-1}$ is formed. At this moment, for each height-i cluster C, $U_i \cap C$ of portals is a βD_i-cover for C, and hence also a βD_i-cover for $C \cap \widehat{U}_{i-1}$. Observe that if $u_i \in U_i$ and $u_{i-1} \in \widehat{U}_{i-1}$ are in the same height-i cluster at this moment, then they will always remain in the same height-i cluster for the remaining execution of the algorithm.

Hence, it suffices to analyze points in $\overline{U}_{i-1}$. Observe that when a point u is added to $\overline{U}_{i-1}$, it remains in the same height-$(i-1)$ cluster in $\tilde{\mathcal{P}}_{i-1}$, and hence also in the same height-i cluster in $\mathcal{P}_i^{(i)}$. It follows that u is at a distance of at most βD_i from some node in $U_i \cap \mathcal{P}_i^{(i)}(u)$. Since $U_{i-1} := \widehat{U}_{i-1} \cup \overline{U}_{i-1}$, it follows that for each height-i cluster, $U_i(C)$ is a βD_i-cover for $C \cap U_{i-1}$. ∎

Next, we analyze the probability that a pair of points u, v is separated in $\mathcal{P}_i^i$. Since the decomposition procedure is quite involved, the analysis requires more care. Throughout this section, the parameter t refers to the one that comes from Fact 5.1. We first prove the following lemma, which is analogous to Lemma 5.6. Recall that $D_{i+1} := HD_i$; for technical reasons, we assume that $H \geq 4t \log n$.

LEMMA 6.4. *Suppose that $u, v \in V$ and that B_u and B_v are balls of radius r centered at u and v, respectively. The probability that $B_u \cup B_v$ is separated by $\mathcal{P}_i^{(i)}$ is at most $4t^2 \log^2 n \cdot \frac{d(u,v)+2r}{D_i}$.*

Proof. We prove this by induction on i. For $i = L$, the statement is trivial because $\mathcal{P}_L = \{V\}$ and no points are separated from one another. Now consider $i < L$. Let $\delta := d(u, v) + 2r$ and $r' := r + \beta D_i$. Observe that if $\mathcal{P}_i^{(i)}$ separates $B_u \cup B_v$, then at least one of the following two events happens: event $\mathcal{E}_1$ is that the partition $\mathcal{P}_{i+1}^{(i+1)}$ separates $B(u, r') \cup B(v, r')$, and event $\mathcal{E}_2$ is that the partition $\tilde{\mathcal{P}}_i$ separates $B(u, r') \cup B(v, r')$.

The probability of event $\mathcal{E}_1$ is, by the induction hypothesis, at most $4t^2 \log^2 n \cdot \frac{\delta+\beta D_i}{D_{i+1}}$. The probability of event $\mathcal{E}_2 \setminus \mathcal{E}_1$ is at most $t \log n \cdot \frac{\delta+\beta D_i}{D_i}$, by Fact 5.1. Hence, observing that $D_{i+1} = HD_i \geq 4tD_i \log n$, the probability of the event $\mathcal{E}_1 \cup \mathcal{E}_2$ is at most

$$2t \log n \cdot \frac{\delta + \beta D_i}{D_i}.$$

Now we consider two cases to bound this expression: if $\delta \geq \beta D_i$, then the expression is at most $4t \log n \cdot \frac{\delta}{D_i}$, and we are done. Else, suppose $\delta < \beta D_i$. Observe that in order for $\mathcal{P}_i^{(i)}$ to separate $B_u \cup B_v$, in addition to the event $\mathcal{E}_1 \cup \mathcal{E}_2$, the event $\mathcal{E}_3$ that $B_u \cup B_v$ is separated during the boundary adjustment step must also occur.

Note that the probability of $\mathcal{E}_3$ given the event $\mathcal{E}_1 \cup \mathcal{E}_2$ is at most $t \log n \cdot \frac{\delta}{\beta D_i}$. Hence, it follows that the required probability is at most

$$2t \log n \cdot \frac{\delta + \beta D_i}{D_i} \cdot t \log n \cdot \frac{\delta}{\beta D_i} \leq 4t^2 \log^2 n \frac{\delta}{D_i},$$

which completes the inductive step, and hence the proof. ∎

LEMMA 6.5. *The probability that a pair (u, v) of points is separated by the final $\mathcal{P}_i$ is at most $(4t \log n)^L \cdot \frac{d(u,v)}{D_i} = O(\log n)^L \cdot \frac{d(u,v)}{D_i}$.*

Proof. Observe that if the final $\mathcal{P}_i$ separates u and v, then for some $j \leq i$, the partition $\mathcal{P}_i^{(j)}$ separates u and v. Let this event be $\mathcal{E}_j$. We consider the probability of such event $\mathcal{E}_j$. Observe that in order for this to happen, for each $j \leq l < i$, the partition $\mathcal{P}_l^{(l)}$ has to separate $B(u, \beta D_{l-1}) \cup B(v, \beta D_{l-1})$, due to boundary adjustment at height l. Let k be the integer such that $2\beta D_k \leq d(u, v) < 2\beta D_{k+1}$, and $\bar{i} := \max\{k + 1, j\}$. Hence, the probability of the event $\mathcal{E}_j$ is at most

$$\begin{aligned} &4t^2 \log^2 n \cdot \frac{d(u,v) + 2\beta D_{i-1}}{D_i} \cdot \left(\prod_{l=\bar{i}+1}^{i-1} t \log n \cdot \frac{d(u,v) + 2\beta D_{l-1}}{D_l} \right) \cdot t \log n \cdot \frac{2d(u,v)}{\beta D_{\bar{i}}} \\ &\quad \leq \frac{1}{2} \cdot (4t \log n)^{i-j+2} \cdot \frac{d(u,v)}{D_i}, \end{aligned}$$

where the first term comes from Lemma 6.4, and each subsequent terms comes from Fact 5.1 applied to each boundary adjustment step. Now, summing $\Pr[\mathcal{E}_j]$ over $j \leq i$ shows that the probability that (u, v) is cut by the final $\mathcal{P}_i$ is at most $(4t \log n)^L \cdot \frac{d(u,v)}{D_i}$. ∎

6.2. Handling large portal sets via patching. As we mentioned earlier, we want to avoid a situation where a cluster C has many child portals, since computing the standard TSP table for the cluster would require time exponential in the number of portals. To avoid this, we do two kinds of patching as we explain below.

6.2.1. Patching a single cluster. The first idea is simple: if we are willing to pay an extra additive term of $O(BD)$ in the length (where B is the number of portals and D is the diameter of the cluster), we show that there exists a tour that enters and leaves cluster C at a single portal. Indeed, we can alter the tour to enter cluster C through some portal x, perform a traveling salesperson tour on points in cluster C, and leave cluster C through x. We describe this patching process in more detail in the following lemmas.

LEMMA 6.6. *Suppose cluster C has diameter D and that there are at most B portals in the cluster C. Then, given any portal-respecting tour on the vertices V, the tour can be modified such that it enters and leaves the cluster C through a single portal and has additional length at most $2BD$.*

Proof. Suppose the height-i cluster C has diameter D. Suppose a given portal-respecting tour enters and leaves C at the portal pairs $\{(x_j, y_j)\}_{j=1}^r$. Let us define the *height* of a portal x to be the maximum value h such that x is a portal of some height-h cluster. Without loss of generality, suppose x_1 is a portal of greatest height among $\cup_j \{x_j, y_j\}$. For each $j \in \{1, 2, \ldots, r\}$, use P_j to denote the part of the tour within C between x_j and y_j, and use Q_j to denote the part of the tour outside C between y_j and x_{j+1} (where we define $x_{r+1} = x_1$).

We modify the given tour in the following way. We start at x_1 and follow P_1 to y_1. Instead of leaving C, we go from y_1 to x_2 inside cluster C, and then follow P_2

inside C to y_2. This goes on until y_r is reached, and then we go directly from y_r to x_1. Now, all points in cluster C have been visited, and we leave cluster C through x_1. Note that at this stage, we have added distances due to $(y_1, x_2), (y_2, x_3), \ldots, (y_r, x_1)$. Observe that all the x_j's and y_j's are portals of height-i clusters and hence are also portals of lower height clusters. Therefore, the patching procedure does not violate the portal-respecting property for lower height clusters.

Now, we continue to patch the tour from outside. For the time being, imagine that each portal x_j or y_j has a twin copy (x'_j or y'_j) colocated with it but considered to be "outside" C. Resuming the tour from x_1, we go to x'_1 and on to y'_1, follow Q_1 to x'_2, go from x'_2 to y'_2, follow Q_2 to x'_3, and so on, until we go from y'_r via Q_r back to x'_1. Note that at this stage, we added distances due to $(x'_1, y'_1), (x'_2, y'_2), \ldots, (x'_r, y'_r)$. Hence, with at most B portals and the diameter of C being at most D, the total distance added so far is at most $2rD \leq BD$.

Now, we have to resolve the issue of going from x'_j to y'_j, because in reality these points do not exist. Let p_j be the last point in Q_{j-1} just before reaching x'_j, and let q_j be the first point in Q_j just after y'_j. Instead of going as $p_j \to x'_j \to y'_j \to q_j$, we want to go from p_j directly to q_j. If this is possible and does not violate the portal condition, then by the triangle inequality, we cannot increase the tour length by doing so. But suppose going from p_j directly to q_j would violate the portal condition. Observe that both x_j and y_j are portals for the height-i cluster C. Hence, if p_j can go directly to x_j, and y_j can go directly to q_j in the original tour, this means both p_j and q_j are portals of height at least i. Since x_1 is a portal of highest height, it is portal respecting to go from p_j to x_1 and then to q_j. This further adds a distance of D for each such p_j and q_j.

Observe that although the point x_1 can be visited a multiple number of times in the whole tour, it is used only once to enter and exit cluster C. Hence, the total increase in distance is at most $2BD$. ∎

The above lemma promises the existence of a good tour that enters and leaves a cluster C via a single portal, but we must specify how to find the subtour within the cluster C efficiently. Indeed, if cluster C has too many child portals, it may be too expensive to perform dynamic programming to find the best tour possible from the information in C's child clusters. Hence, we may need a second patching step.

LEMMA 6.7. *Consider the dynamic program in section* 5.3*, and look at a cluster C with diameter D and having B child portals. Suppose l is the length of the shortest tour for the points in C that is computable from the entries in the child clusters of C (possibly in $2^{\Omega(B \log B)}$ extra time). Then, it is possible to obtain a tour for cluster C, again from the entries in the child clusters of C, that has length at most $l + BD$. Moreover, with the entries in the child clusters already computed, it takes further $O(B)$ computation time to obtain this tour.*

Proof. From each child cluster C_λ of C, pick the entry such that the length l_λ of its partial segments is smallest. Each partial segment corresponds to the part of the tour within the corresponding child cluster C_λ. Note that the length l of the optimal tour on C is at least $\sum_\lambda l_\lambda$. Since there are at most B child portals and the diameter of C is D, it takes an extra length of BD to join the partial segments returned by each child cluster to form a tour on C. ∎

Observe that any portal of C is also a child portal of C. Hence, using Lemmas 6.6 and 6.7, for any cluster C with diameter D and B child portals, we can do the patching procedure in time $O(B)$ from the entries of its child clusters. After the procedure, each entry of cluster C is indexed by a single portal and has a value corresponding to the length of some tour on cluster C. The resulting increase in length for the overall tour is at most $3BD$.

6.2.2. Applying a patching technique in the dynamic program. We analyze the increase in tour lengths when we apply the patching procedure described in section 6.2.1. Since the dynamic program is performed in a bottom-up fashion, patching is also performed starting from lower height clusters.

Let OPT_0 be the length of the optimal tour returned by the dynamic program (without patching) described in section 5.3. Suppose patching is performed for all the clusters with heights $j \leq i$ that had more than B_0 child portals, and no patching is performed for clusters at heights higher than i. Let the length of the best tour obtainable with such a partial patching be OPT_i. We want to bound OPT_L, which is the length of the tour returned by the dynamic program when patching is performed whenever appropriate. We do this by the recurrence in the following lemma, which bounds the extra length incurred by patching all clusters at one level. (Recall that k is the correlation dimension of the metric.)

LEMMA 6.8. *For $0 \leq i < L$, $\mathsf{OPT}_{i+1} \leq \mathsf{OPT}_i + \frac{3}{2B_0}(\frac{8H}{\beta})^{k+1}\mathsf{OPT}$.*

Proof. Suppose $\{C_\lambda : \lambda \in \Lambda\}$ is the set of height-$(i+1)$ clusters such that each one has $B_\lambda > B_0$ child portals. Observe that the set of height-i portals is a $\frac{1}{4}\beta D_i$-packing. Hence, we can extend it to a $\frac{1}{4}\beta D_i$-net N_i for V. From section 6.2.1, it follows that the extra length to patch up all appropriate height-$(i+1)$ clusters is at most $3\sum_\lambda B_\lambda D_{i+1}$. Now, from the definition of correlation dimension, we have for all integers t,

$$\sum_{x \in N_i} |\mathbf{B}_{N_i}(x, D_{i+1})| \leq 2^{kt} \sum_{x \in N_i} |\mathbf{B}_{N_i}(x, 2^{-t} \cdot D_{i+1})|.$$

By setting $t := \lceil \log_2(4D_{i+1}/\beta D_i) \rceil$ and recalling $D_{i+1} = HD_i$, we have

$$\sum_\lambda B_\lambda^2 \leq \sum_{x \in N_i} |\mathbf{B}_{N_i}(x, D_{i+1})| \leq \left(\frac{8H}{\beta}\right)^k |N_i|. \tag{6.1}$$

Observing that each $B_\lambda > B_0$, we have

$$|\Lambda| \leq \frac{1}{B_0^2} \left(\frac{8H}{\beta}\right)^k |N_i|. \tag{6.2}$$

Using the Cauchy–Schwarz inequality and inequalities (6.1) and (6.2), we have

$$\sum_\lambda B_\lambda \leq \sqrt{|\Lambda| \cdot \sum_\lambda B_\lambda^2}. \tag{6.3}$$

By substituting (6.1) and (6.2) into (6.3), we have

$$\sum_\lambda B_\lambda \leq \frac{1}{B_0} \left(\frac{8H}{\beta}\right)^k |N_i|.$$

Finally, observing that $\mathsf{OPT} \geq \frac{1}{4}\beta D_i |N_i|$, we conclude that the extra length incurred by patching all appropriate height-$(i+1)$ clusters is at most

$$3\sum_\lambda B_\lambda D_{i+1} \leq \frac{3}{2B_0} \left(\frac{8H}{\beta}\right)^{k+1} \mathsf{OPT}. \qquad \square$$

Applying Lemma 6.8, we get that by patching whenever applicable, the total extra length incurred is small:

$$\mathsf{OPT}_L \le \mathsf{OPT}_0 + \frac{3L}{2B_0}\left(\frac{8H}{\beta}\right)^{k+1}\mathsf{OPT}. \tag{6.4}$$

6.3. The subexponential time TSP algorithm. We improve on the TSP algorithm from section 5.3 using the enhanced partitioning, portaling, and patching ideas to get the following result.

THEOREM 6.9 (subexponential time algorithm for TSP). *For any constant k (independent of n), there is a randomized $(1+\varepsilon)$-approximation algorithm for TSP on metrics of size n with correlation dimension at most k that runs in time $\exp\{(\varepsilon^{-1}2^{\sqrt{\log n \log\log n}})^{4k}\} = 2^{O(n^\delta \varepsilon^{-O(k)})}$ for any $\delta > 0$.*

Proof. We create a probabilistic hierarchical decomposition, where the diameter at height-i is $D_i = H^i$ for some parameter $H = \Omega(\log n)$. Hence the depth of the tree is $L := \Theta(\log_H(n/\varepsilon))$. As indicated above (and proved in Lemma 6.5), the probability that (u,v) are separated at level-i is at most $\alpha\frac{d(u,v)}{D_i}$, with $\alpha = O(\log n)^L$. Moreover, portals in clusters of diameter D_i form a βD_i-cover, and since there are L levels, the total increase in the TSP length is $O(\alpha\,\beta\,L)\,\mathsf{OPT}$. To make this at most $\varepsilon/2$, we set $\beta = O(\varepsilon/L\,\alpha)$.

Finally, from an analysis in section 6.2.2, the length increase from patching is $\frac{3L}{2B_0}(\frac{8H}{\beta})^{k+1}\mathsf{OPT}$. To make this at most $\varepsilon/2$ as well, we pick B_0 such that $\frac{3L}{2B_0}(\frac{8H}{\beta})^{k+1} = \varepsilon/2$.

The only parameter left to be chosen is H. Observe that the running time depends on B_0, and so H is chosen to minimize B_0. Note that

$$B_0 = \left(\frac{L}{\varepsilon}\right)^{k+2} O(H\alpha)^{k+1}.$$

Observe that $H\alpha$ is the dominating term and also that as H increases, α decreases. It happens that in this case the best value is attained when $H = \alpha$. This is satisfied when $\log H = \sqrt{\log\frac{n}{\varepsilon}\log\log n}$.

It follows that it suffices to set the threshold $B_0 = \varepsilon^{-(k+2)}2^{2(k+1)\sqrt{\log\frac{n}{\varepsilon}\log\log n}} \le (\varepsilon^{-1}\cdot 2^{\sqrt{\log n\log\log n}})^{3k}$, recalling $\varepsilon > \frac{1}{n}$. Hence, we obtain a tour with expected length $(1+\varepsilon)$ times that of the optimal tour in time

$$nL\cdot 2^{O(B_0\log B_0)} = \exp\{(\varepsilon^{-1}\cdot 2^{\sqrt{\log n\log\log n}})^{4k}\} = 2^{O(n^\delta\varepsilon^{-O(k)})}$$

for any $\delta > 0$. ∎

7. Summary and conclusions. In this paper, we considered a global notion of dimension, which tries to capture the "average" complexity of metrics: our notion of correlation dimension captures metrics that potentially contain small near-uniform metrics (of size $O(\sqrt{n})$) but still have small average growth-rate. We showed that metrics with a low correlation dimension do indeed admit efficient algorithms for a variety of problems.

Many questions remain open: can we improve the running time of our algorithm for TSP? A more open-ended question is how to define other notions of dimension for metric spaces: it is unlikely that one notion can capture the complexity of metrics, and it seems reasonable to consider several definitions whose properties can then be exploited under the appropriate circumstances.

REFERENCES

[ADD$^+$93] 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.

[AGGM06] I. Abraham, C. Gavoille, A. Goldberg, and D. Malkhi, *Routing in networks with low doubling dimension*, in Proceedings of the 26th IEEE International Conference on Distributed Computing Systems (ICDCS), IEEE Computer Society Press, Los Alamitos, CA, 2006, p. 75.

[AM05] I. Abraham and D. Malkhi, *Name independent routing for growth bounded networks*, in Proceedings of the 17th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA), ACM, New York, 2005, pp. 49–55.

[AP89] S. Arnborg and A. Proskurowski, *Linear time algorithms for NP-hard problems restricted to partial k-trees*, Discrete Appl. Math., 23 (1989), pp. 11–24.

[Aro96] S. Arora, *Polynomial time approximation schemes for Euclidean TSP and other geometric problems*, in Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science (Burlington, VT, 1996), IEEE Computer Society Press, Los Alamitos, CA, 1996, pp. 2–11.

[Aro98] S. Arora, *Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems*, J. ACM, 45 (1998), pp. 753–782.

[Aro02] S. Arora, *Approximation algorithms for geometric TSP*, in The Traveling Salesman Problem and Its Variations, Comb. Optim. 12, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2002, pp. 207–221.

[ARR99] S. Arora, P. Raghavan, and S. Rao, *Approximation schemes for Euclidean k-medians and related problems*, in Proceedings of the 13th Annual ACM Symposium on Theory of Computing (STOC '98), ACM, New York, 1999, pp. 106–113.

[Ass83] P. Assouad, *Plongements lipschitziens dans* $\mathbf{R}^n$, Bull. Soc. Math. France, 111 (1983), pp. 429–448.

[Bar96] Y. Bartal, *Probabilistic approximations of metric spaces and its algorithmic applications*, in Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society Press, Los Alamitos, CA, 1996, pp. 184–193.

[BF95] A. Belussi and C. Faloutsos, *Estimating the selectivity of spatial queries using the 'correlation' fractal dimension*, in Proceedings of the 21st International Conference on Very Large Data Bases (VLDB), Morgan Kaufmann, San Francisco, 1995, pp. 299–310.

[BKL06] A. Beygelzimer, S. Kakade, and J. Langford, *Cover trees for nearest neighbor*, in Proceedings of the 23rd International Conference on Machine Learning, ACM, New York, 2006, pp. 97–104.

[CDNS95] B. Chandra, G. Das, G. Narasimhan, and J. Soares, *New sparseness results on graph spanners*, Internat. J. Comput. Geom. Appl., 5 (1995), pp. 125–144.

[CG06] 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 the Theory of Computing (STOC), ACM, New York, 2006, pp. 574–583.

[CG08] D. E. Carroll and A. Goel, *Lower Bounds for Embedding into Distributions over Excluded Minor Graph Families*, preprint, arXiv:0807.4582v1, 2008.

[CGMZ05] H. T.-H. Chan, A. Gupta, B. M. Maggs, and S. Zhou, *On hierarchical routing in doubling metrics*, in Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), ACM, New York, SIAM, Philadelphia, 2005, pp. 762–771.

[CL98] A. Czumaj and A. Lingas, *A polynomial time approximation scheme for Euclidean minimum cost k-connectivity*, in Automata, Languages and Programming (Aalborg, 1998), Lecture Notes in Comput. Sci. 1443, Springer, Berlin, 1998, pp. 682–694.

[Cla99] K. L. Clarkson, *Nearest neighbor queries in metric spaces*, Discrete Comput. Geom., 22 (1999), pp. 63–93.

[Cla06] K. L. Clarkson, *Nearest-neighbor searching and metric space dimensions*, in Nearest-Neighbor Methods for Learning and Vision: Theory and Practice, G. Shakhnarovich, T. Darrell, and P. Indyk, eds., MIT Press, Cambridge, MA, 2006, pp. 15–59.

[CLZ02] A. Czumaj, A. Lingas, and H. Zhao, *Polynomial-time approximation schemes for the Euclidean survivable network design problem*, in Automata, Languages and Programming, Lecture Notes in Comput. Sci. 2380, Springer, Berlin, 2002, pp. 973–984.

[Die00] R. Diestel, *Graph Theory*, 2nd ed., Grad. Texts in Math. 173, Springer-Verlag, New York, 2000.

[FK94] C. Faloutsos and I. Kamel, *Beyond uniformity and independence: Analysis of r-trees using the concept of fractal dimension*, in Proceedings of the 13th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, ACM, New York, 1994, pp. 4–13.

[FRT04] J. Fakcharoenphol, S. Rao, and K. Talwar, *A tight bound on approximating arbitrary metrics by tree metrics*, J. Comput. System Sci., 69 (2004), pp. 485–497.

[GJ79] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman, San Francisco, 1979.

[GKL03] A. Gupta, R. Krauthgamer, and J. R. Lee, *Bounded geometries, fractals, and low–distortion embeddings*, in Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society Press, Los Alamitos, CA, 2003, pp. 534–543.

[GP83] P. Grassberger and I. Procaccia, *Measuring the strangeness of strange attractors*, Phys. D, 9 (1983), pp. 189–208.

[Gri00] M. Grigni, *Approximate tsp in graphs with forbidden minors*, in Proceedings of the 27th International Colloquium on Automata, Languages and Programming (ICALP), Springer-Verlag, London, 2000, pp. 869–877.

[Hås96] J. Håstad, *Clique is hard to approximate within* $n^{1-\epsilon}$, in Proceedings of the 37th Annual IEEE Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society Press, Los Alamitos, CA, 1996, pp. 627–636.

[HKRZ02] K. Hildrum, J. D. Kubiatowicz, S. Rao, and B. Y. Zhao, *Distributed object location in a dynamic network*, in Proceedings of the 14th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA), ACM, New York, 2002, pp. 41–52.

[HPM05] S. Har-Peled and M. Mendel, *Fast constructions of nets in low dimensional metrics, and their applications*, in Proceedings of the 21st ACM Symposium on Computational Geometry (SOCG), ACM, New York, 2005, pp. 150–158.

[IM04] P. Indyk and J. Matoušek, *Low-distortion embeddings of finite metric spaces*, in Handbook of Discrete and Computational Geometry, 2nd ed., J. E. Goodman and J. O'Rourke, eds., Chapman & Hall/CRC, Boca Raton, FL, 2004, pp. 177–196.

[KK77] L. Kleinrock and F. Kamoun, *Hierarchical routing for large networks. Performance evaluation and optimization*, Comput. Networks, 1 (1976/77), pp. 155–174.

[KL04] R. Krauthgamer and J. R. Lee, *Navigating nets: Simple algorithms for proximity search*, in Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), ACM, New York, SIAM, Philadelphia, 2004, pp. 798–807.

[KL05] R. Krauthgamer and J. R. Lee, *The black-box complexity of nearest-neighbor search*, Theoret. Comput. Sci., 348 (2005), pp. 262–276.

[KL06] R. Krauthgamer and J. R. Lee, *Algorithms on negatively curved spaces*, in Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society Press, Los Alamitos, CA, 2006, pp. 119–132.

[Kle05] P. N. Klein, *A linear-time approximation scheme for planar weighted tsp*, in Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society Press, Los Alamitos, CA, 2005, pp. 647–657.

[KMW05] F. Kuhn, T. Moscibroda, and R. Wattenhofer, *On the locality of bounded growth*, in Proceedings of the 24th Annual ACM Symposium on Principles of Distributed Computing (PODC), ACM, New York, 2005, pp. 60–68.

[KR99] S. G. Kolliopoulos and S. Rao, *A nearly linear-time approximation scheme for the Euclidean k-median problem*, in Algorithms—ESA '99 (Prague), Lecture Notes in Comput. Sci. 1643, Springer, Berlin, 1999, pp. 378–389.

[KR02] D. R. Karger and M. Ruhl, *Finding nearest neighbors in growth-restricted metrics*, in Proceedings of the 34th Annual ACM Symposium on the Theory of Computing (STOC), ACM, New York, 2002, pp. 63–66.

[KRX06a] G. Konjevod, A. W. Richa, and D. Xia, *On sampling in higher-dimensional peer-to-peer systems*, in LATIN 2006: Theoretical Informatics, Lecture Notes in Comput. Sci. 3887, Springer, Berlin, 2006, pp. 641–652.

[KRX06b] G. Konjevod, A. W. Richa, and D. Xia, *Optimal-stretch name-independent compact routing in doubling metrics*, in Proceedings of the 25th Annual ACM Symposium on Principles of Distributed Computing (PODC), ACM, New York, 2006, pp. 198–207.

[KSW04] J. M. Kleinberg, A. Slivkins, and T. Wexler, *Triangulation and embedding using small sets of beacons*, in Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS), IEEE Computer Society Press, Los Alamitos, CA, 2004, pp. 444–453.

[Mat02] J. Matoušek, *Lectures on Discrete Geometry*, Grad. Texts in Math. 212, Springer-Verlag, New York, 2002.

[Mit99] J. S. B. Mitchell, *Guillotine subdivisions approximate polygonal subdivisions: A simple polynomial-time approximation scheme for geometric TSP, k-MST, and related problems*, SIAM J. Comput., 28 (1999), pp. 1298–1309.

[PKF00] B.-U. Pagel, F. Korn, and C. Faloutsos, *Deflating the dimensionality curse using multiple fractal dimensions*, in Proceedings of the 16th International Conference on Data Engineering, IEEE Computer Society Press, Los Alamitos, CA, 2000, pp. 589–598.

[PRR99] C. G. Plaxton, R. Rajaraman, and A. W. Richa, *Accessing nearby copies of replicated objects in a distributed environment*, Theory Comput. Syst., 32 (1999), pp. 241–280.

[PS89] D. Peleg and A. A. Schäffer, *Graph spanners*, J. Graph Theory, 13 (1989), pp. 99–116.

[PY93] C. H. Papadimitriou and M. Yannakakis, *The traveling salesman problem with distances one and two*, Math. Oper. Res., 18 (1993), pp. 1–11.

[RS99] S. B. Rao and W. D. Smith, *Approximating geometrical graphs via "spanners" and "banyans,"* in Proceedings of the Annual ACM Symposium on Theory of Computing, ACM, New York, 1999, pp. 540–550.

[Sli05] A. Slivkins, *Distance estimation and object location via rings of neighbors*, in Proceedings of the 24th Annual ACM Symposium on Principles of Distributed Computing (PODC), ACM, New York, 2005, pp. 41–50.

[Tal04] K. Talwar, *Bypassing the embedding: Algorithms for low-dimensional metrics*, in Proceedings of the 36th Annual ACM Symposium on Theory of Computing (STOC), ACM, New York, 2004, pp. 281–290.

[Tre00] L. Trevisan, *When Hamming meets Euclid: The approximability of geometric TSP and Steiner tree*, SIAM J. Comput., 30 (2000), pp. 475–485.