Approximation Algorithms for Bounded Dimensional Metric Spaces

T.-H. Hubert Chan

CMU-CS-07-157

September 2007

School of Computer Science
Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Anupam Gupta, Chair
Avrim Blum
R. Ravi
Kenneth L. Clarkson, IBM Research, Almaden

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Keywords: Approximation algorithms, bounded dimension, metric spaces, embeddings, spanners

Abstract

The study of finite metrics is an important area of research, because of its wide applications to many different problems. The input of many problems (for instance clustering, near-neighbor queries and network routing) naturally involves a set of points on which a distance function has been defined. Hence, one would be motivated to store and process metrics in an efficient manner. The central idea in metric embedding is to represent a metric space by a "simpler" one so that the properties of the original metric space are well preserved.

More formally, given a *target class* $\mathcal{C}$ of metrics, an *embedding* of a finite metric space $M = (V, d)$ into the class $\mathcal{C}$ is a new metric space $M' = (V', d')$ such that $M' \in \mathcal{C}$. Most of the work on embeddings has used *distortion* as the fundamental measure of quality; the distortion of an embedding is the worst multiplicative factor by which distances are increased by the embedding. In the theoretical community, the popularity of the notion of distortion has been driven by its applicability to approximation algorithms: if the embedding $\varphi : (V, d) \to (V', d')$ has a distortion of D, then the costs of solutions to some optimization problems on (V, d) and those on (V', d') can only differ by some function of D; this idea has led to numerous approximation algorithms. Seminal results include the $O(\log n)$ distortion embeddings of arbitrary metrics into Euclidean spaces with $O(\log n)$ dimensions, and the fact that any metric admits an $O(\log n)$ stretch spanner with $O(n)$ edges.

The theoretical results mentioned above are optimal. However, they are pessimistic in the sense that such guarantees hold for any arbitrary metric. It is conceivable that better results can be obtained if the input metrics are "simple". The main theme of this work is to investigate notions of complexity for an abstract metric space and theoretical guarantees for problems in terms of the complexity of the input metric.

One popular notion for measuring the complexity of a metric is the doubling dimension, which restricts the local growth rate of a metric. We show that the results on spanners and embeddings can be improved if the given metrics have bounded doubling dimension. For instance, we give a construction for constant stretch spanners with a linear number of edges. Moreover, such metrics can be embedded into Euclidean space with $O(\log \log n)$ dimensions and $o(\log n)$ distortion.

We also study a new notion of dimension that captures the global growth rate of a metric. Such a notion strictly generalizes doubling dimension in the sense that it places weaker restrictions on a given metric than those posed by doubling dimension. However, we can still obtain good guarantees for problems in which the objective depends on the global nature of the metric, an example of which is the Traveling Salesperson Problem (TSP). In particular, we give a sub-exponential time algorithm to solve TSP with approximation ratio arbitrarily close to 1 for such metrics.

Contents

1 Introduction **1**

1.1 Outline of the Thesis . . . 3

1.1.1 Sparse Spanners for Doubling Metrics . . . 3

1.1.2 Ultra-Low Dimensional Embeddings for Doubling Metrics . . . 4

1.1.3 Approximating TSP on Metrics with Bounded Global Growth . . . 4

1.2 Definitions and Notations . . . 5

2 Sparse spanners for doubling metrics **8**

2.1 Introduction . . . 8

2.1.1 Notation and Preliminaries . . . 10

2.2 Basic Construction of Sparse $(1 + \epsilon)$-Spanners for Doubling Metrics . . . 10

2.3 Construction of $((1 + \epsilon)$-Spanners with Bounded Degree . . . 13

2.4 Sparse Spanners with Small Hop-diameter . . . 15

2.4.1 A Warm-up: Obtaining $O(\log n)$ Hop-diameter . . . 15

2.4.2 The General Upper Bound for Hop-diameter . . . 16

2.4.3 The Lower Bound on Hop-diameter . . . 18

3 Ultra-Low Dimensional Embeddings for Doubling Metrics **25**

3.1 Introduction . . . 25

3.1.1 Related Work . . . 28

3.1.2 Notation and Preliminaries . . . 28

3.2 The Basic Embedding . . . 29

3.2.1 Basic Embedding: Defining The Embedding . . . 29

3.2.2 Basic Embedding: Bounding Contraction . . . 31

3.2.3 Basic Embedding: Bounding Expansion . . . 34

3.2.4 Resolving Dependency among Random Variables 36
3.3 A Better Embedding via Uniform Padded Decompositions 38
3.3.1 Uniform Padded Decompositions 38
3.3.2 The Better Embedding: Defining the Embedding 42
3.3.3 The Better Embedding: Bounding Contraction for Nearby Net Points . . 42
3.3.4 The Better Embedding: Bounding the Expansion 44
3.3.5 The Better Embedding: Bounding Contraction for All Pairs 45

4 Approximating TSP on Metrics with Bounded Global Growth 47
4.1 Introduction 47
4.1.1 Related Work 50
4.2 Correlation Dimension: Definition and Motivation 50
4.3 Relating Doubling and Correlation Dimensions 52
4.3.1 Correlation Dimension Generalizes Doubling 52
4.3.2 The Converse is False 54
4.4 Hardness of Approximating Correlation Dimension 54
4.5 Sparse Spanners 56
4.5.1 Sparse Spanners: Upper Bound 57
4.5.2 Sparse Spanners: Lower Bound 58
4.6 Algorithms for Metrics with Bounded Correlation Dimension 59
4.6.1 An Algorithm for TSP in Time $2^{\tilde{O}(\sqrt{n})}$ 61
4.6.2 Hierarchical Decomposition and Portal-Respecting Tour 61
4.6.3 A Partitioning and Portaling Algorithm 63
4.6.4 Dynamic Programming for Solving TSP 66
4.6.5 The First TSP Algorithm 67
4.6.6 Embedding into Small Treewidth Graphs 67
4.7 A Sub-Exponential Time $(1 + \epsilon)$-Approximation for TSP 67
4.7.1 The Modified Partitioning and Portaling Algorithm 69
4.7.2 Handling Large Portal Sets via Patching 71
4.7.3 The Second TSP Algorithm 73
4.8 Summary and Conclusions 74

5 Conclusion **75**
5.1 Spanners for Doubling Metrics 75
5.2 Low Dimensional Embeddings for Doubling Metrics 76
5.3 Global Notion of Dimension 76

Bibliography **78**

Chapter 1

Introduction

The study of finite metrics[1] is an important area of research, because of its wide applications to many different problems. The input of many problems (for instance clustering, near-neighbor query and network routing) naturally involves a set of points on which a distance function has been defined. Hence, one would be motivated to store and process metrics in an efficient manner. The central idea in metric embedding is to represent a metric space with a "simpler" one so that the properties of the original metric space are well preserved.

More formally, given a *target class* $\mathcal{C}$ of metrics, an *embedding* of a finite metric space $M = (V, d)$ into the class $\mathcal{C}$ is a new metric space $M' = (V', d')$ such that $M' \in \mathcal{C}$. Most of the work on embeddings has used *distortion* as the fundamental measure of quality; the distortion of an embedding is the worst multiplicative factor by which distances are increased by the embedding[2]. Given the metric $M = (V, d)$ and the class $\mathcal{C}$, one natural goal is to find an embedding $\varphi((V, d)) = (V', d') \in \mathcal{C}$ such that the distortion of the map φ is minimized.

This notion of metric embedding is general in the sense that it captures several embedding frameworks. For example, when the class $\mathcal{C}$ is the class of all Euclidean metrics, or the class of all ℓ_1 metrics, we have the familiar notion of embeddings of metric spaces into geometric spaces. On the other hand, if the class $\mathcal{C}$ is the class of metrics generated by sparse (weighted) graphs, such embeddings give rise to *sparse spanners*. Note that the concept of distortion is often called "stretch" in the spanners literature. Moreover, we have the notion of embeddings into a distribution over tree metrics, where $\mathcal{C}$ is the class of convex combinations of tree metrics.

In the theoretical community, the popularity of the notion of distortion/stretch has been driven by its applicability to approximation algorithms: if the embedding $\varphi : (V, d) \to (V', d')$ has a distortion of D, then the costs of solutions to some optimization problems on (V, d) and those on (V', d') can only differ by some function of D; this idea has led to numerous approximation algorithms [Ind01]. Seminal results include the $O(\log n)$ distortion embeddings of arbitrary metrics into ℓ_p spaces [Bou85], the fact that any metric admits an $O(\log n)$ stretch spanner with $O(n)$ edges [ADD$^+$93], and that any metric can be embedded into a distribution of trees with

[1]A list of formal definitions of concepts appearing frequently in this work is found in Section 1.2.

[2]Formally, for an embedding $\varphi : (V, d) \to (V', d')$, the *distortion* is the smallest D so that $\exists K > 0$ such that $d(x, y) \le d'(\varphi(x), \varphi(y))/K \le D\, d(x, y)$ for all pairs $(x, y) \in V \times V$.

distortion $O(\log n)$ [FRT04], where n is the size of V in all cases. All the above three results are known to be tight.

In parallel to the theoretical work on embeddings, there has been much recent interest within more applied communities in embeddings (and more generally, but also somewhat vaguely, on problems on finding "simpler representations" of distance spaces). One example arises in the networking community [NZ02, DCKM04], which is interested in taking the point-to-point latencies between nodes in a network, treating it as a metric space $M = (V, d)$ satisfying the triangle inequality,[3] and then finding some simpler representation $M' = (V', d')$ of this resulting metric so that distances between nodes can be quickly and accurately computed in this "simpler" metric M'. (E.g., they are interested in assigning each node a short *label* so that the distance between two nodes can be approximately inferred merely by looking at their labels.)

The theoretical results mentioned above, although being optimal, are pessimistic in the sense that such guarantees hold for any arbitrary metric. Simply using the size of a given metric to quantify the performance of algorithms is unsatisfactory, for it is conceivable that better results can be obtained if the input metrics are "simple". The main theme of this work is to investigate notions of complexity for an abstract metric space and theoretical guarantees for problems in terms of the complexity of the input metric (and its size).

One popular notion for measuring the complexity of a metric is the doubling dimension, which restricts the local growth rate of a metric. The *doubling dimension* of a metric $M = (V, d)$ is the minimum value k such that every ball B in the metric can be covered by 2^k balls of half the radius of B. This can be seen as a generalization of Euclidean dimension to arbitrary metric spaces; indeed, it is not difficult to see that $\mathbb{R}^k$ equipped with any of the ℓ_p norms has doubling dimension $\Theta(k)$. Apart from being a generalization of the ℓ_p notion of dimension, designing algorithms that only use the doubling properties (instead of the geometry of $\mathbb{R}^k$) has other advantages: the notion of doubling dimension is fairly resistant to small perturbations in the distances: for instance, if one takes a distance matrix of a set of points in ℓ_p^k and slightly changes some of the entries, then the doubling dimension does not change by much, but the metric may not remain isometrically embeddable in ℓ_p (into any number of dimensions).

The notion of doubling dimension was introduced by Larman [Lar67] and Assouad [Ass83], and first used in nearest-neighbor searching by Clarkson [Cla99]. The properties of doubling metrics have since been studied extensively, and various algorithms have been generalized to adapt gracefully to the doubling dimension of the input metric; for examples, see [GKL03, KL03, KL04, Tal04, HPM05, BKL06, CG06b, IN, KRX06, KRX07].

Continuing research in this direction, we show that the results on spanners and embeddings can be improved if the given metrics have bounded doubling dimension. For instance, we give a construction for constant stretch spanners with a linear number of edges. Moreover, such metrics can be embedded into Euclidean space with $O(\log \log n)$ dimensions and $o(\log n)$ distortion.

On the other hand, although doubling dimension is defined for any metric and is indeed an extension of the ℓ_p notion of dimension, it is still a stringent notion in the sense that the doubling

[3]While the triangle inequality can be violated by network latencies, empirical evidence [LGS07] suggests that these violations are small and/or infrequent enough to make metric methods a useful approach.

property has to be satisfied everywhere, namely every ball can be covered by a small number of balls of half its radius. Observe that if the metric contains a uniform metric of size $\Omega(\sqrt{n})$, then the doubling dimension of the metric is at least $\Omega(\log n)$. Intuitively, one would like to get a notion of dimension such that if a metric behaves "nicely" in general except for small localized regions, then the metric has small dimension. Hence, for problems in which the objective depends globally on the metric, one would expect such "nice" metrics to be easy instances of the problem.

In this work, we study a new notion of dimension, which we call "correlation dimension"[4], that captures the global growth rate of a metric. Such a notion strictly generalizes doubling dimension in the sense that any metric with bounded doubling dimension also has bounded global dimension. Intuitively, we should be able to obtain good guarantees for problems in which the objective depends on the global nature of the metric. In particular, we consider $(1 + \epsilon)$-approximation algorithm for Traveling Salesman Problem (TSP), in the context of metrics with bounded global dimension. Indeed, we give a sub-exponential time algorithm to solve TSP with approximation ratio arbitrarily close to 1 for such metrics.

1.1 Outline of the Thesis

We present this work in three chapters: (1) Sparse spanners for doubling metrics; (2) Ultra-low dimensional embeddings for doubling metrics; and (3) Approximating TSP on metrics with bounded global growth. The chapter on spanners contains results which have appeared in the papers [CGMZ05, CG06a], while the results on the other two chapters will appear in the proceeding of the upcoming SODA in 2008. The work on low dimensional embeddings is a collaboration with Anupam Gupta and Kunal Talwar, while that on global dimension is done with Anupam Gupta. Each chapter is self-contained and can be read separately from the others. We summarize the results of each chapter in the following.

1.1.1 Sparse Spanners for Doubling Metrics

In Chapter 2, we give good constructions of spanners for doubling metrics. A t-spanner for a metric is a weighted subgraph whose shortest path distance preserves the original metric within a multiplicative factor of t. Given a metric with doubling dimension $\dim$, we show how to construct $(1 + \epsilon)$-spanners with $n(1 + 1/\epsilon)^{O(\dim)}$ edges. Observe that a $(1 + \epsilon)$-spanner for an arbitrary metric can have at least $\Omega(n^2)$ edges. From this basic sparse spanner construction, we can obtain a sparse spanner that has either (1) bounded degree ($(1+1/\epsilon)^{O(\dim)}$) or (2) small *hop-diameter*. Observe that it is not possible to achieve both, as that would imply the total number of points is too small.

A t-spanner has hop-diameter D if every pair $u, v \in V$ are connected by some short path in G having length at most $t\,d(u, v)$, and there are at most D edges on this path. In particular, we show one can find a $(1+\epsilon)$-spanner for the metric with a nearly linear number of edges (i.e., only

[4]There is a previously defined notion of "correlation dimension" that inspires our definition. Perhaps the name "net correlation dimension" is more suitable for us. However, for brevity, we still use the term "correlation dimension" in later discussion.

$O(n \log^* n + n\epsilon^{-O(\dim)}$ edges) and *constant* hop diameter; we can also obtain a $(1+\epsilon)$-spanner with a linear number of edges (i.e., only $n\epsilon^{-O(\dim)}$ edges) that achieves a hop diameter that grows like the functional inverse of the Ackermann's function. Moreover, we prove that such tradeoffs between the number of edges and the hop-diameter are asymptotically optimal.

1.1.2 Ultra-Low Dimensional Embeddings for Doubling Metrics

In Chapter 3, we consider the problem of embedding a metric into low-dimensional Euclidean space. The classical theorems of Bourgain [Bou85], and of Johnson and Lindenstrauss [JL84] say that any metric on n points embeds into an $O(\log n)$-dimensional Euclidean space with $O(\log n)$ distortion. Moreover, a simple "volume" argument shows that this bound is nearly tight: a uniform metric on n points requires a nearly logarithmic number of dimensions to embed with logarithmic distortion. It is natural to ask whether such a volume restriction is the only hurdle to low-dimensional embeddings. In other words, do *doubling* metrics, that do not have large uniform submetrics, and thus no volume hurdles to low dimensional embeddings, embed in low dimensional Euclidean spaces with small distortion?

We give a positive answer to this question. We show how to embed any doubling metric into $O(\log \log n)$-dimensional Euclidean space with $o(\log n)$ distortion. This is the first embedding for doubling metrics into fewer than logarithmic number of dimensions, even allowing for logarithmic distortion.

This result is one extreme point of our general trade-off between distortion and dimension: given an n-point metric (V, d) with doubling dimension $\dim_D$, and any target dimension T in the range $\Omega(\dim_D \log \log n) \leq T \leq O(\log n)$, we show that the metric embeds into Euclidean space $\mathbb{R}^T$ with $O(\log n \sqrt{\dim_D / T})$ distortion.

1.1.3 Approximating TSP on Metrics with Bounded Global Growth

In Chapter 4, we approximate the Traveling Salesperson Problem (TSP) for a class of metrics broader than doubling metrics. Observe that TSP is a canonical NP-complete problem which is known to be MAX-SNP hard even on Euclidean metrics (of high dimensions) [Tre00]. In order to circumvent this hardness, researchers have been developing approximation schemes for "simpler" instances of the problem. For instance, Arora [Aro98] and Talwar [Tal04] showed how to approximate TSP on low-dimensional metrics (for different notions of dimension). This has been part of a larger effort to quantify "simple metrics" (say, with respect to some problem such as TSP). In particular, can we define the "dimension" of metric spaces so that the performance of algorithms on a given metric space can be quantified meaningfully in terms of the dimension of the metric? Many proposed notions of dimension have been shown to have good algorithmic properties (see, e.g., [PRR99, KR02, Cla99, GKL03, KL06]).

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, and such strong properties might not be satisfied in real-life metrics. *What if our metric looks a bit more realistic: it has a few "dense" regions, but is "well-behaved on the average"?* How do we even begin to formalize this idea? We give a global notion of dimension: the *correlation dimension* ($\dim_C$). Loosely speaking, a metric has constant correlation dimension if the number of node-pairs in the metric within distance r of each other only increases by a constant factor if we go from $r \to 2r$ (i.e., if their range-of-sight doubles).

We show that this global notion of dimension generalizes the popular notion of doubling dimension: the class of metrics with $\dim_C = O(1)$ contains not only all doubling metrics, but also some metrics containing cliques of size $\sqrt{n}$ (but no larger cliques). We first show that we can solve TSP (and other optimization problems) on these metrics in time $2^{O(\sqrt{n})}$; then we take advantage of the global nature of TSP (and the global nature of our definition) to give a $(1+\epsilon)$-approximation algorithm that runs in *sub-exponential time*: i.e., in $2^{O(n^{\delta}\epsilon^{-4 \dim_C})}$-time for every constant $0 < \delta < 1$. For this new algorithm, we have to develop new techniques beyond those used for earlier PTASs for TSP: since metrics with bounded $\dim_C$ may contain hard metrics of size $O(\sqrt{n})$, we show that beating the $\exp(\sqrt{n})$ running time requires finding $O(1)$-approximations to some portions of the tour, and $(1+\varepsilon)$-approximations for other portions, and stitching them together; these new ingredients are potentially of independent interest.

1.2 Definitions and Notations

We end this chapter by defining precisely the terminology we frequently use, which might have different meanings in another context. We consider finite metric spaces, and we use (V, d) to denote a finite metric space; unless otherwise stated, we denote the size of the metric by $n = |V|$. We make precise what we mean by a metric space in the following definition.

Definition 1.2.1 (Metric space) *A* metric space (V, d) *consists of a point set* V *and a distance function* $d : V \times V \to [0, \infty)$*, also called a* metric*, such that the following properties are satisfied.*

1. $d(u, v) = 0$ iff $u = v$.
2. *Symmetry: for all* $u, v \in V$*,* $d(u, v) = d(v, u)$*.*
3. *Triangle inequality: for all* $u, v, w \in V$*,* $d(u, w) \le d(u, v) + d(v, w)$*.*

The symmetry and the triangle inequality are the important properties of metric spaces that we use. If the first condition is replaced with "$d(u, v) = 0$ if $u = v$", then the distance function d is called a *semi-metric*. However, observe that a semi-metric d always induces a metric on the equivalence classes obtained from the equivalence relation $u \sim v$ *iff* $d(u, v) = 0$.

One way to measure the complexity of a metric space is its dimension. Intuitively, a metric space consisting of points on the real line is simple, while a metric whose points lie in high dimensional Euclidean space is complex. Since not all metrics have a valid Euclidean dimension, a popular notion of dimension is used for general metric spaces: the doubling dimension.

Definition 1.2.2 (Doubling dimension) *A metric space* (V, d) *has* doubling dimension *at most* k *if for all* $R > 0$*, any ball of radius* R *is contained in the union of at most* 2^k *balls of radius* $R/2$*. A ball of radius* R *consists of all the points that are at distance at most* R *from some center point.*

Observe that any finite number of points in ℓ_p^k induce a metric space with doubling dimension at most $O(k)$. A *doubling metric* is a metric which has bounded doubling dimension. The concept of a net is useful for doubling metrics.

Definition 1.2.3 (Net) *Let S be a set of points in a metric space (V, d), and $r > 0$. A subset N of S is an* r-net *for S if the following conditions hold.*

1. *For all $x \in S$, there exists some $y \in N$ such that $d(x, y) \leq r$.*
2. *For all $y, z \in N$ such that $y \neq z$, $d(y, z) > r$.*

The following fact states that for a doubling metric, one cannot pack too many points in some fixed ball such that the points are far away from one another.

Fact 1.2.4 *Suppose S is a set of points in a metric space with doubling dimension at most k. If S is contained in some ball of radius R and for all $y, z \in S$ such that $y \neq z$, $d(y, z) > r$, then $|S| \leq (4R/r)^k$.*

A spanner is a structure that preserves the distance function of a metric space. The precise definition is given in the following.

Definition 1.2.5 (Spanner for a metric) *A spanner H for a metric space (V, d) is a weighted undirected graph on the vertex set V with the edge set H such that the weight of an edge $\{u, v\} \in H$ is $d(u, v)$. For $t \geq 1$, the spanner H is a t-spanner if for all $u, v \in V$, the shortest path distance $d_H(u, v)$ between u and v in the graph H satisfies $d_H(u, v) \leq t d(u, v)$, in which case we say the* stretch *or* distortion *of the spanner H is at most t.*

Note that in the literature, the term *spanner* usually means a subgraph of an unweighted graph. What we define as a spanner is referred to as an *emulator*, whose definition differs slightly in the sense that it is only required that $d_H(u, v) \geq d(u, v)$. However, since we are interested in minimizing the stretch, without loss of generality, one can assume that every edge in H has weight given by the metric d.

Using spanners is only one way a metric space can be tranformed. In general, an *embedding* is a mapping $\varphi : (V, d) \to (V', d')$ from the metric space (V, d) to the metric space (V', d'). The quality of an embedding is measured by how much distances are distorted, which is quantified by distortion.

Definition 1.2.6 (Distortion) *The* distortion *of an embedding $\varphi : (V, d) \to (V', d')$ is the smallest D so that $\exists\, K > 0$ such that for all pairs $(x, y) \in V \times V$, $d(x, y) \leq d'(\varphi(x), \varphi(y))/K \leq D\, d(x, y)$, if φ is an injection; if φ is not an injection, the distortion is infinity.*

Note. Observe that the role of K in the above definition is that if we scale all distances of an embedding by the same multiplicative factor, then its distortion does not change. However, if we only consider embeddings that do not contract distances, then it is enough to have for all pairs (x, y) satisfy $d(x, y) \leq d'(\varphi(x), \varphi(y)) \leq D\, d(x, y)$.

We recall a useful technique that gives a probabilistic decomposition of a metric such that each component has small diameter and the probability that two points are separated is proportional to their distance.

Definition 1.2.7 (Padded Decompositions [GKL03, KLMN05]) *Given a finite metric space (V, d), a positive parameter $\Delta > 0$ and $\alpha > 1$, a Δ-bounded* α-padded decomposition *is a distribution Π over partitions of V such that the following conditions hold.*

(a) For each partition P in the support of Π, the diameter of every cluster in P is at most Δ.

(b) Suppose $S \subseteq V$ is a set with diameter d. If P is sampled from Π, then the set S is partitioned by P with probability at most $\alpha \cdot \frac{d}{4D}$.

For simplicity, say that a metric *admits α-padded decompositions* if for every $\Delta > 0$ it admits a Δ-bounded α-padded decomposition. It is known that any finite metric space admits an $O(\log n)$-padded decomposition [Bar96]. Moreover, metrics of doubling dimension $\dim_V$ admit $O(\dim_V)$-padded decompositions [GKL03]; furthermore, if a graph G excludes K_r-minors (e.g., if it has treewidth $\leq r$), then its shortest-path metric admits $O(r^2)$-padded decompositions [KPR93, Rao99, FT03].

Chapter 2

Sparse spanners for doubling metrics

2.1 Introduction

In this chapter, we give constructions for obtaining sparse representations of metrics: these are called *spanners*, and they have been studied extensively both for general and Euclidean metrics. Formally, a t-spanner for a metric $M = (V, d)$ is a weighted undirected graph $G = (V, E)$ such that the distances according to d_G (the shortest-path metric of G) are close to the distances in d: i.e., $d(u, v) \leq d_G(u, v) \leq t\, d(u, v)$.[1] In this case, we also say that the spanner has *stretch* at most t. Clearly, one can take a complete graph and obtain $t = 1$, and hence the quality of the spanner is typically measured by how few edges G can contain while maintaining a stretch of at most t. The notion of spanners has been widely studied for general metrics (see, e.g. [PS89, ADD$^+$93, CDNS95]), and for geometric distances (see, e.g., [CK95, Sal91, Vai91, ADM$^+$95]). Here, we are particularly interested in the case when the input metric has bounded doubling dimension and the spanner we want to construct has small stretch, i.e. $t = 1 + \epsilon$, for small $\epsilon > 0$. We show that for fixed ϵ and metrics with bounded doubling dimension, it is possible to construct linear sized $(1 + \epsilon)$-spanners. Observe that any 1.5-spanner for a uniform metric on n points must be the complete graph. Hence, without any restriction on the input metric, it is not possible to construct a $(1 + \epsilon)$-spanner with a linear number of edges.

We also show how to construct sparse spanners with small hop diameter. A t-spanner has hop-diameter D if every pair $u, v \in V$ are connected by some short path in G having length at most $t\, d(u, v)$, and there are at most D edges on this path.

Main Results. We first give a basic construction of sparse spanners for doubling metrics.

Theorem 2.1.1 (Basic Spanner Construction) *Given a metric (V, d) with doubling dimension $\dim$, there exists a $(1 + \epsilon)$-spanner with $(2 + \frac{1}{\epsilon})^{O(\dim)} n$ edges.*

We can modify the edges in this basic sparse spanner construction to obtain a spanner that has bounded degree.

[1] Note that the first inequality implies that an edge (u, v) in G has weight at least $d(u, v)$.

Theorem 2.1.2 (Constant Degree Spanners) *Given a metric* (V, d) *with doubling dimension* $\dim$*, there exists a* $(1+\epsilon)$*-spanner such that the degree of every vertex is at most* $(2+\frac{1}{\epsilon})^{O(\dim)}$.

On the other hand, we can add extra edges to the spanner from Theorem 2.1.1 to obtain one with small hop-diameter. Observe that the constant degree spanner obtained in Theorem 2.1.2 must have a hop diameter of $\Omega(\log \Delta)$. We prove upper bounds on hop-diameter as well as essentially matching lower bounds.

Theorem 2.1.3 (Upper Bound on Hop-diameter) *Given a metric* $M = (V, d)$ *with doubling dimension* $\dim$ *and* $n = |V|$*, there exists a* $(1+\epsilon)$*-spanner with with* $m + (2+\frac{1}{\epsilon})^{O(\dim)} n$ *edges and hop diameter* $O(\alpha(m, n))$*, where* α *is the inverse of Ackermann's function. Such a spanner can be constructed in* $2^{O(\dim)} n \log n$ *time.*

Note that the result above allows us to trade off the number of edges in the spanner with the hop-diameter: if we desire only a linear number of edges, then the hop-diameter goes as $\alpha(n)$, and as we increase the number of edges, the hop-diameter decreases. After proving this result (which turns out to be fairly straight-forward given known techniques), we then turn to the lower bound and show that the trade-off in Theorem 2.1.3 is essentially tight.

Theorem 2.1.4 (Lower Bound on Hop-diameter) *For any* $\epsilon > 0$*, there are infinitely many integers* n *such that there exists a metric* M *induced by* n *points on the real line, for which any* $(1+\epsilon)$*-spanner for* M *with at most* m *edges has hop diameter at least* $\Omega(\alpha(m, n))$.

Our Techniques and Related Work. Independent of our work, Har-Peled and Mendel [HPM05] also use a similar construction to obtain $(1+\epsilon)$-spanners with $n(1+1/\epsilon)^{O(\dim)}$ edges. However, the spanners obtained have a hop-diameter of $\Omega(\log \Delta)$, where Δ is the aspect ratio of the metric.

The upper bound in Theorem 2.1.3 generalizes a result of Arya et al. [ADM$^+$95] for Euclidean spaces. Indeed, the proof of our result is not difficult given previously known techniques. The basic idea is to first construct a *net-tree* representing a sequence of nested nets of the metric space: this is fairly standard, and has been used earlier, e.g., in [CGMZ05, KL04, Tal04]. A nearly-linear-time construction of net-trees is given by Har-Peled and Mendel [HPM05]. A second phase then adds some more edges in order to "short-cut" paths in this net tree which have too many hops. The techniques we use are based on those originally used by Yao [Yao82] for range queries on the line, and on the extensions to trees due to Chazelle [Cha87]. As pointed out by Arya et al. [ADM$^+$95], a similar construction was given by Alon and Schieber [AS87].

To the best of our knowledge, there are no previously known lower bounds which show metrics with low doubling (or Euclidean) dimension that require many edges in order to get low hop-diameter $(1+\epsilon)$-spanners. We first consider lower bounds for binary "hierarchically well-separated" trees (HSTs), where the length of an edge from each node to its child node is much smaller than that to its parent node: this well-separation ensures that low-stretch paths must be "well-behaved": i.e., the low-stretch path between vertices in any subtree cannot escape the subtree, thus allowing us to reason about them. Our lower bound result for line metrics then follows from the fact that binary HSTs with large separation embed into the real line with small distortion. We note that the lower bounds for the range-query problem given by Yao [Yao82], and Alon and Scheiber [AS87], while inspiring our work, directly apply to our problem only for the case $\epsilon = 0$; i.e., for the case where we are not allowed to introduce any further stretch in the second, "short-cutting" phase. Thus Theorem 2.1.4 can be seen as generalizing Yao's lower bound proof to all $\epsilon > 0$.

Other Related Work. Abraham et al. [IA04] study compact routing on Euclidean metrics, and their construction also essentially gives a $(1+\epsilon)$-spanner with $O_\epsilon(n)$ edges that has hop diameter $O(\log \Delta)$ with high probability.

Low-stretch spanners with small hop-diameter are potentially useful in network routing protocols. For example, many wireless ad-hoc networks find paths that minimize hop count [PC97, PBR99, PB94]. Our results may be useful in such situations to build sparse networks admitting paths having few hops and low stretch simultaneously.

2.1.1 Notation and Preliminaries

We recall the definitions of some frequently encountered concepts. We consider a finite metric $M = (V, d)$ where $|V| = n$. A metric has *doubling dimension* [GKL03] at most k if for every $R > 0$, every ball of radius R can be covered by 2^k balls of radius $R/2$.

Definition 2.1.5 ($(1+\epsilon)$-spanner) *Let (V, d) be a finite metric space. Suppose $G = (V, E)$ is an undirected graph such that each edge $\{u, v\} \in E$ has weight $d(u, v)$, and $d_G(u, v)$ is the length of the shortest path between vertices u and v in G. The graph G, or equivalently, the set E of edges, is a $(1+\epsilon)$-*spanner *for (V, d) if for all pairs u and v, $d_G(u, v)/d(u, v) \leq 1 + \epsilon$.*

A $(1+\epsilon)$*-path* in the metric $M = (V, d)$ between u and v is one with length at most $(1+\epsilon)d(u, v)$. Thus a $(1 + \epsilon)$-spanner is a subgraph $G = (V, E)$ that contains a $(1 + \epsilon)$ path for each pair of nodes in V.

Definition 2.1.6 (Hop Diameter) *A $(1+\epsilon)$-spanner is said to have* hop diameter *at most D if for every pair of nodes, there exists a $(1+\epsilon)$-path in the spanner between them having at most D edges or* hops.

2.2 Basic Construction of Sparse $(1+\epsilon)$-Spanners for Doubling Metrics

In this section, we show the existence of sparse spanners by giving an explicit construction. In particular, we have the following result.

Theorem 2.2.1 *Given a metric (V, d) with doubling dimension k, there exists a $(1+\epsilon)$-spanner $\widehat{E}$ that has $(2 + \frac{1}{\epsilon})^{O(k)} n$ edges.*

The basic idea is to first construct a *net-tree* representing a sequence of nested nets of the metric space: this is fairly standard, and has been used earlier, e.g., in [Tal04, KL04, CGMZ05]. A nearly-linear-time construction of net-trees is given by Har-Peled and Mendel [HPM05].

Net trees are formally defined in the following.

Definition 2.2.2 (Hierarchical Tree) *A* hierarchical tree *for a set V is a pair (T, φ), where T is a rooted tree, and φ is a labeling function $\varphi : T \to V$ that labels each node of T with an element in V, such that the following conditions hold.*

1. *Every leaf is at the same depth from the root.*
2. *The function φ restricted to the leaves of T is a bijection into V.*
3. *If u is an internal node of T , then there exists a child v of u such that $\varphi(v) = \varphi(u)$. This implies that the nodes mapped by φ to any $x \in V$ form a connected subtree of T.*

Definition 2.2.3 (Net-Tree) *A* net tree *for a metric (V, d) is a hierarchical tree (T, φ) for the set V such that the following conditions hold.*

1. *Let N_i be the set of nodes of T that have height i. (The leaves have height 0.) Suppose δ is the minimum pairwise distance in (V, d). Let $0 < r_0 < \delta/2$, and $r_{i+1} = 2r_i$, for $i \geq 0$. (Hence, $r_i = 2^i r_0$.) Then, for $i \geq 0$, $\varphi(N_{i+1})$ is an r_{i+1}-net for $\varphi(N_i)$.*
2. *Let node $u \in N_i$, and its parent node be p_u. Then, $d(\varphi(u), \varphi(p_u)) \leq r_{i+1}$.*

In order to construct the spanner, we include an edge if the end points are from the same net in some scale and "reasonably close" to each other with respect to that scale. Using this idea, one can obtain the following theorem.

Theorem 2.2.4 *Given a finite metric $M = (V, d)$ with doubling dimension bounded by* $\dim$. *Let $\epsilon > 0$ and (T, φ) be any net tree for M. For each $i \geq 0$, let*

$$E_i := \{\{u, v\} \mid u, v \in \varphi(N_i), d(u, v) \leq (4 + \tfrac{32}{\epsilon}) \cdot r_i\} \setminus E_{i-1},$$

where E_{-1} is the empty set. (Here the parameters N_i, r_i are as in Definition 2.2.3.) Then $\widehat{E} := \cup_i E_i$ forms a $(1 + \epsilon)$-spanner for (V, d), with the number of edges being $|\widehat{E}| \leq (2 + \frac{1}{\epsilon})^{O(\dim)}|V|$.

We prove Theorem 2.2.4 through Lemmas 2.2.5 and 2.2.8.

Lemma 2.2.5 *The graph $(V, \widehat{E})$ is a $(1 + \epsilon)$-spanner for (V, d).*

Proof: Let $\widehat{d}$ be the distance function induced by $(V, \widehat{E})$. Let $\gamma := 4 + \frac{32}{\epsilon}$. We first show that each point in V is close to some point in $\varphi(N_i)$ under the metric $\widehat{d}$.

Claim 2.2.6 *For all $x \in V$, for all i, there exists $y \in \varphi(N_i)$ such that $\widehat{d}(x, y) \leq 2r_i$.*

Proof: We shall prove this by induction on i. For $i = 0$, $\varphi(N_0) = V$. Hence, the result holds trivially.

Suppose $i \geq 1$. By the induction hypothesis, there exists $y' \in \varphi(N_{i-1})$ such that $\widehat{d}(x, y') \leq 2r_{i-1}$. Since $\varphi(N_i)$ is an r_i-net of $\varphi(N_{i-1})$, there exists $y \in \varphi(N_i) \subseteq \varphi(N_{i-1})$ such that $d(y', y) \leq r_i = 2r_{i-1} \leq \gamma \cdot r_{i-1}$. Hence,, $(y', y) \in E_i \subseteq \widehat{E}$ and $\widehat{d}(y', y) = d(y', y)$, which is at most r_i.

Finally, by the triangle inequality, $\widehat{d}(x, y) \leq \widehat{d}(x, y') + \widehat{d}(y', y) \leq 2r_{i-1} + r_i = 2r_i$. ■

We next show that for any pair of vertices $x, y \in V$, $\widehat{d}(x, y) \leq (1 + \epsilon)d(x, y)$. Suppose $r_i \leq d(x, y) < r_{i+1}$.

Suppose q is the integer such that $\frac{8}{2^q} \leq \epsilon < \frac{16}{2^q}$, i.e. $q := \lceil \log_2 \frac{8}{\epsilon} \rceil$.

We first consider the simple case when $i \leq q-1$. Then, $d(x,y) < 2^{i+1} r_0 \leq 2^q r_0 \leq \frac{16}{\epsilon} \cdot r_0 \leq \gamma \cdot r_0$. Since $x, y \in \varphi(N_0)$, it follows that $(x,y) \in \widehat{E}$ and $\widehat{d}(x,y) = d(x,y)$.

Next we consider the case when $i \geq q$. Let $j := i - q \geq 0$.

By Claim 2.2.6, there exist vertices $x', y' \in \varphi(N_j)$ such that $\widehat{d}(x,x') \leq 2r_j$ and $\widehat{d}(y,y') \leq 2r_j$.

We next show that $(x', y') \in \widehat{E}$. It suffices to show that $d(x', y') \leq \gamma \cdot r_j$.

$$
\begin{array}{rcll}
d(x',y') & \leq & d(x',x) + d(x,y) + d(y,y') & \text{(Triangle inequality)} \\
& \leq & 2r_j + r_{i+1} + 2r_j & \text{(Choice of } x', y' \text{ and } i\text{)} \\
& = & r_j(4 + 2 \cdot 2^q) & (i = j + q) \\
& \leq & r_j(4 + \frac{32}{\epsilon}) & (2^q < \frac{16}{\epsilon}) \\
& = & \gamma \cdot r_j &
\end{array}
$$

Hence, we have $\widehat{d}(x', y') = d(x', y')$. Note that by the triangle inequality,

$$
d(x',y') \leq d(x',x) + d(x,y) + d(y,y') \leq 4 \cdot r_j + d(x,y). \tag{2.1}
$$

Finally, we obtain the desired upper bound for $\widehat{d}(x,y)$.

$$
\begin{array}{rcll}
\widehat{d}(x,y) & \leq & \widehat{d}(x,x') + \widehat{d}(x',y') + \widehat{d}(y',y) & \text{(Triangle inequality)} \\
& \leq & 8 \cdot r_j + d(x,y) & \text{(Choice of } x', y' \text{ and (2.1))} \\
& = & \frac{8}{2^q} \cdot r_i + d(x,y) & (j = i - q) \\
& \leq & (1 + \frac{8}{2^q}) d(x,y) & (r_i \leq d(x,y)) \\
& \leq & (1 + \epsilon) d(x,y) & (\frac{8}{2^q} \leq \epsilon)
\end{array}
$$

■

Observe that we have not used the definition of doubling dimension so far. We next proceed to show that the spanner $(V, \widehat{E})$ is sparse, by using the fact that the metric is doubling. We first show that for each vertex u, for each i, the number of edges in E_i incident on u is small.

Claim 2.2.7 *Define* $\Gamma_i(u) := \{v \in V : \{u,v\} \in E_i\}$. *Then,* $|\Gamma_i(u)| \leq (4\gamma)^k$.

Proof: Observe that $\Gamma_i(u)$ is contained in a ball of radius at most $\gamma \cdot r_i$ centered at u. Moreover, since $S \subseteq \varphi(N_i)$, any two points in S must be more than r_i apart. Hence, from Fact 1.2.4, it follows that $|\Gamma_i(u)| \leq (4\gamma)^k$. ■

Lemma 2.2.8 *The number of edges in* $\widehat{E}$ *is at most* $(2 + \frac{1}{\epsilon})^{O(k)} n$.

Proof: It suffices to show that the edges of $\widehat{E}$ can be directed such that each vertex has out-degree bounded by $(2 + \frac{1}{\epsilon})^{O(k)}$.

For each $v \in V$, define $i^*(v) := \max\{i \,|\, v \in \varphi(N_i)\}$. For each edge $(u,v) \in \widehat{E}$, we direct the edge from u to v if $i^*(u) < i^*(v)$. If $i^*(u) = i^*(v)$, the edge can be directed arbitrarily. By *arc* (u,v), we mean an edge that is directed from vertex u to vertex v.

We now bound the out-degree of vertex u. Suppose there exists an arc $(u, v) \in E_i$.

By definition of E_i, $d(u, v) \leq \gamma \cdot r_i$. Set $p = \lceil \log_2 \gamma \rceil$. Hence, it is not possible for both u and v to be contained in $\varphi(N_{i+p})$. Since $i^*(u) \leq i^*(v)$, it follows that $i^*(u) \leq i + p$. On the other hand, $u \in \varphi(N_i)$ and so $i^*(u) \geq i$. So, $i^*(u) - p \leq i \leq i^*(u)$.

There are at most $p + 1 = O(\log \gamma)$ values of i such that E_i contains an edge directed out of u. By Claim 2.2.7, for each i, the number of edges in E_i incident on u is at most $(4\gamma)^k$.

Hence, the total number of edges in $\widehat{E}$ directed out of u is $(4\gamma)^k \cdot O(\log \gamma) = (2 + \frac{1}{\epsilon})^{O(k)}$.

■

Observe that in the proof of Lemma 2.2.5, we have actually shown that for any points x and y, there is a short path of a particular form. This property will be useful when we construct spanners with small hop-diameter.

Theorem 2.2.9 *Consider the construction in Theorem 2.2.4. For any x, y in V, the spanner $\widehat{E}$ contains a $(1+\epsilon)$-path of the following form. If x_0 and y_0 are the leaf nodes in T with $\varphi(x_0) = x$ and $\varphi(y_0) = y$, and x_i and y_i are the ancestors of x_0 and y_0 at height $i \geq 1$, then there exists $\bar{i}$ such that the path*

$$x = \varphi(x_0), \varphi(x_1), \ldots, \varphi(x_{\bar{i}}), \varphi(y_{\bar{i}}), \ldots, \varphi(y_1), \varphi(y_0) = y$$

is a $(1 + \epsilon)$-path (after removing repeated vertices).

2.3 Construction of ($(1 + \epsilon)$-Spanners with Bounded Degree

We have shown that the edges in $\widehat{E}$ can be directed such that the out-degree of every vertex is bounded. We next describe how to modify $\widehat{E}$ to get another set of edges $\tilde{E}$ that has size at most that of $\widehat{E}$, but the resulting undirected graph $(V, \tilde{E})$ has bounded degree (Lemma 2.3.1). Moreover, we show in Lemma 2.3.2 that the modification preserves distances between vertices.

We form the new graph $(V, \tilde{E})$ by modifying the directed graph $(V, \widehat{E})$ in the following way.

> **Modification Procedure.** Let l be the smallest positive integer such that $\frac{1}{2^{l-1}} \leq \epsilon$. Then, $l = O(\log \frac{1}{\epsilon})$.
>
> For each i and point u, define $M_i(u)$ to be the set of vertices w such that $w \in \Gamma_i(u)$ and (w, u) is directed into u in $\widehat{E}$.
>
> Let $I_u := \{i \mid \exists v \in M_i(u)\}$. Suppose the elements of I_u are listed in increasing order $i_1 < i_2 < \cdots$. To avoid double subscripts, we write $M_j^u := M_{i_j}(u)$.
>
> We next modify arcs going into each vertex u in the following manner. For $1 \leq j \leq l$, we keep the arcs directed from M_j^u to u. For $j > l$, we pick an arbitrary vertex $w \in M_{j-l}^u$ and for each point $v \in M_j^u$, replace the arc (v, u) by the arc (v, w).
>
> Observe that since M_j^u is defined with respect to the directed graph $(V, \hat{E})$, the ordering of the u's for which the modification is carried out is not important.

Let $(V, \tilde{E})$ be the resulting undirected graph. Since every edge in $\widehat{E}$ is either kept or replaced by another edge (which might be already in $\widehat{E}$), $|\tilde{E}| \leq |\widehat{E}|$.

Lemma 2.3.1 *Every vertex in* $(V, \tilde{E})$ *has degree bounded by* $(2 + \frac{1}{\epsilon})^{O(k)}$.

Proof: Let α be an upper bound for the out-degree of the graph $(V, \widehat{E})$. From Lemma 2.2.8, we have $\alpha = (2+\frac{1}{\epsilon})^{O(k)}$. Let β be an upper bound for $|M_i(u)|$. We have $\beta \leq |\Gamma_i(u)| = (2+\frac{1}{\epsilon})^{O(k)}$.

We next bound the maximum degree of a vertex in $(V, \tilde{E})$. Consider a vertex $u \in V$. The edges incident on u can be grouped as follows.

1. There are at most α edges directed out of u in $\widehat{E}$.
2. Out of the edges in $\widehat{E}$ directed into u, at most βl remain in $\tilde{E}$.
3. New edges can be attached to u in $(V, \tilde{E})$. For each arc (u, v) directed out of u in $\widehat{E}$, there can be at most β new edges attaching to u in $\tilde{E}$. The reason is (u, v) can be in exactly one E_i and so there exists unique j such that $u \in M_j^v$. Hence, there could be potentially only at most $|M_{j+l}^v|$ new arcs directed into u because of the arc (u, v) in $\widehat{E}$.

Hence, the number of edges incident on u in $(V, \tilde{E})$ is bounded by $\alpha + \beta l + \alpha\beta = (2 + \frac{1}{\epsilon})^{O(k)}$.

■

We next show that the modification from $(V, \widehat{E})$ to $(V, \tilde{E})$ does not increase the distance between any pair of vertices too much.

Lemma 2.3.2 *Suppose* $\tilde{d}$ *is the metric induced by* $(V, \tilde{E})$. *Then,* $\tilde{d} \leq (1 + 4\epsilon)\widehat{d}$.

Proof: It suffices to show that for each edge $(v, u) \in \widehat{E}$ removed, $\tilde{d}(v, u) \leq (1 + 4\epsilon)d(v, u)$.

Suppose (v, u) in $\widehat{E}$ is directed into u . Then, by construction, $v \in M_j^u$ for some $j > l$.

Let $v_0 = v$. Then, from our construction, for $0 \leq s \leq s_j := \left\lfloor \frac{j-1}{l} \right\rfloor$, there exists $v_s \in M_{j-sl}^u$ such that for $0 \leq s < s_j$, $(v_s, v_{s+1}) \in \tilde{E}$, and $(v_{s_j}, u) \in \tilde{E}$. Then, there is a path in $(V, \tilde{E})$ going from v to u traversing vertices in the following order: $v = v_0, v_1, \ldots, v_{s_j}, u$. By the triangle inequality, the quantity $\tilde{d}(v, u)$ is at most the length of this path, which we show is comparable to $d(v, u)$.

Claim 2.3.3 *For* $0 \leq s < s_j$, $d(u, v_{s+1}) \leq \epsilon d(u, v_s)$.

Proof: Note that $v_{s+1} \in M_i(u)$ and $v_s \in M_j(u)$ for some i and j. From step 3 of our construction, $j - i \geq l$.

Since $d(v_s, u) \geq \gamma \cdot r_{j-1}$ and $d(v_{s+1}, u) \leq \gamma \cdot r_i$, it follows that $d(v_{s+1}, u) \leq \frac{2}{2^l} d(v_s, u) \leq \epsilon d(v_s, u)$.

■

Claim 2.3.4 *For* $0 \leq s \leq s_j$, $d(v_s, u) \leq \epsilon^s d(v_0, u)$.

Proof: The claim can be proved by induction on s and using Claim 2.3.3. ■

From the triangle inequality and Claims 2.3.3 and 2.3.4, we have

$$d(v_s, v_{s+1}) \leq d(v_s, u) + d(u, v_{s+1}) \leq (1 + \epsilon)d(v_s, u) \leq (1 + \epsilon)\epsilon^s d(v_0, u) \tag{2.2}$$

Finally, we have

$$\begin{array}{rcll}\tilde{d}(v,u) & \leq & \sum_{s=0}^{s_j-1} d(v_s, v_{s+1}) + d(v_{s_j}, u) & \text{(Triangle inequality)} \\ & \leq & \sum_{s=0}^{s_j-1} (1+\epsilon)\epsilon^s d(v_0,u) + \epsilon^{s_j} d(v_0,u) & \text{((2.2) and Claim 2.3.4)} \\ & \leq & \frac{1+\epsilon}{1-\epsilon} d(v_0,u) & \\ & \leq & (1+4\epsilon) d(v,u) & \end{array}$$

The last inequality follows from the fact that for $0 < \epsilon < \frac{1}{2}$, $\frac{1+\epsilon}{1-\epsilon} \leq 1 + 4\epsilon$. ■

Finally, we show that $(V, \tilde{E})$ is the desired spanner.

Theorem 2.3.5 *Given a metric (V, d) with doubling dimension k, there exists a $(1+\epsilon)$-spanner such that the degree of every vertex is at most $(2 + \frac{1}{\epsilon})^{O(k)}$.*

Proof: We show that $\tilde{E}$ gives the desired spanner. Lemma 2.3.1 gives the bound on its degree. From Lemmas 2.2.5 and 2.3.2, we have $\tilde{d} \leq (1 + 4\epsilon)\widehat{d} \leq (1 + 4\epsilon)(1 + \epsilon)d \leq (1 + 7\epsilon)d$, for $0 < \epsilon \leq \frac{1}{2}$. Substituting $\epsilon := \frac{\epsilon'}{7}$ gives the required result. ■

2.4 Sparse Spanners with Small Hop-diameter

Observe that our spanner in Theorem 2.2.4 has $(2+\frac{1}{\epsilon})^{O(\mathrm{dim})} \cdot n$ edges, and hence is optimal (with respect to n) in terms of the sparsity achieved while preserving shortest path distance. It is easy to check that the number of hops in a $(1+\epsilon)$-path obtained in Theorem 2.2.9 is $\Theta(\log \Delta)$, where Δ is the *aspect ratio* of the metric (V, d) (i.e., the ratio of the maximum to the minimum pairwise distances). Indeed, the net tree (T, φ) has a height of $\Theta(\log \Delta)$, and in general, a $(1 + \epsilon)$-path can have $\Omega(\log \Delta)$ hops.

Before we begin in earnest to investigate how many extra edges are required in order to achieve small hop-diameter, let us make a simple observation. For each node u in the tree T, let L_u be the set of leaves under u. For each node u, suppose we add an edge between $\varphi(u)$ and every point in $\varphi(L_u)$. Since the tree has $O(\log \Delta)$ levels, the number of extra edges added is $O(n \log \Delta)$, while the hop-diameter of the augmented spanner is at most 3. In the next section, we will build on this idea to show how one can reduce the number of additional edges to $O(n \log n)$ (independent of the aspect ratio Δ) and achieve the same hop-diameter.

2.4.1 A Warm-up: Obtaining $O(\log n)$ Hop-diameter

Notice that Theorem 2.2.4 holds for any net tree (T, φ). Hence, by choosing a net tree more carefully, we could possibly improve the trade-off between the hop-diameter of the spanner and its size. Indeed, we show in the next theorem that we can improve the parameter $\log \Delta$ to $\log n$ in both cases. (Note that if a metric has constant doubling dimension, $\log \Delta = \Omega(\log n)$.)

Theorem 2.4.1 *Suppose (V, d) is a finite metric, where $|V| = n$. Then, there exists a net tree (T, φ) from which the spanner $\widehat{E}$ constructed in the manner described in Theorem 2.2.4 has the following properties.*

1. *The hop-diameter of the spanner $\widehat{E}$ is $O(\log n)$.*
2. *It is possible to add $n(\lfloor \log_2 n \rfloor - 1)$ extra edges such that for all leaves $u \in N_0$ in T and any ancestor v of u, there is an edge between $\varphi(u)$ and $\varphi(v)$. (Hence, the hop-diameter of the spanner can be reduced to 3.)*

Proof: We describe a way to construct a net tree (T, φ). Let N_0 be the set of leaves for which there is a one-one correspondence φ onto V.

Suppose we have obtained the set N_i of nodes of height i. We would be done if $|N_i| = 1$. Otherwise, we would obtain an r_{i+1}-net for $\varphi(N_i)$ in the following way. We show a way to greedily construct a net for a set. Start with a list L initially containing all the nodes in N_i, ordered such that a node containing more leaves in its subtree would appear earlier.

As long as the list L is not empty, we repeat the following process. Remove the first node u in the remaining list, form a new node $v \in N_{i+1}$ such that $\varphi(v) := \varphi(u)$ and set the parent of u to be v. For each node w in the remaining list L such that $d(\varphi(w), \varphi(v)) \leq r_{i+1}$, remove w from list L and set the parent of w to be v.

Claim 2.4.2 *For each $x \in N_0$, let A_x be the set of its ancestors in T. Then, $|\varphi(A_z)| \leq \lfloor \log_2 n \rfloor + 1$. In particular, $|\varphi(A_z) \setminus \{\varphi(z)\}| \leq \lfloor \log_2 n \rfloor$.*

Proof: Let a_i be the ancestor of z in N_i. Suppose there exists i such that $\varphi(a_i) \neq \varphi(a_{i+1})$. It follows that the node a_i must have a sibling c, for which $\varphi(c) = \varphi(a_{i+1})$, whose subtree contains at least as many leaves as the subtree at a_i does. Hence, the subtree at a_{i+1} contains at least twice as many leaves as a_i does. Thus there can be at most $\lfloor \log_2 n \rfloor$ values of i for which $\varphi(a_i) \neq \varphi(a_{i+1})$. ■

For the first part of the theorem, it follows that the $(1+\epsilon)$-path guaranteed in Theorem 2.2.4 has at most $2\lfloor \log_2 n \rfloor + 1$ hops.

For the second part of the theorem, for every $z \in N_0$, we add an edge between $\varphi(z)$ and every point in $\varphi(A_z) \setminus \{\varphi(z)\}$. Note that $|\varphi(A_z) \setminus \{\varphi(z)\}| \leq \lfloor \log_2 n \rfloor$. Suppose y is the lowest ancestor of z such that $\varphi(z) \neq \varphi(y)$, and suppose x is the ancestor of z that is also the child of y. Then, observe that the spanner $\widehat{E}$ already includes the edge between $\varphi(y)$ and $\varphi(x) = \varphi(z)$. Hence, for each vertex z, we actually only need to add at most $\lfloor \log_2 n \rfloor - 1$ extra edges. The $(1+\epsilon)$-path in Theorem 2.2.4 can be reduced to $x = \varphi(x_0), \varphi(x_{\bar{i}}), \varphi(y_{\bar{i}}), \varphi(y_0) = y$, which has 3 hops. ■

In the following section, we will investigate the tradeoff between the hop-diameter of a spanner and the number of edges, this time using any given net tree instead.

2.4.2 The General Upper Bound for Hop-diameter

In this section, we assume that the given metric (V, d) has doubling dimension bounded by k. Given a net tree (T, φ) for the metric, suppose E_T is the spanner obtained in Theorem 2.2.4. Note that E_T is dependent on the stretch parameter ϵ. However, for ease of notation, we would leave out the dependency on ϵ throughout this section.

The approach we use is similar to that used by Arya et al. [ADM$^+$95] for Euclidean metrics, which is a subclass of doubling metrics. Instead of using net trees, they worked with "dumbbell trees", which have similar properties. Applying a construction from [Cha87, AS87] to "shortcut" edges in the net-tree, we can show that one can add few extra edges to E_T in order to achieve small hop-diameter. Moreover, as shown in [AS87], this can be done in $O(n \log n)$ time.

We first consider how to add extra edges to a tree such that every pair of nodes has a path with a small number of hops between them.

Definition 2.4.3 *Define $g(m, n)$ to be the minimum i such that for any tree metric with with vertex set V, where $|V| = n$, there exists a spanner P with m edges that preserves all pairwise distances exactly, and for any pair of points, there is a shortest path in P with i hops.*

Lemma 2.4.4 *Suppose a metric (V, d) with n points has a net tree (T, φ), and suppose E_T is the $(1 + \epsilon)$-spanner obtained in Theorem 2.2.4. Then, it is possible to add m extra edges to E_T such that the hop-diameter of the new spanner is at most $2g(m, n) + 1$.*

Proof: Suppose u is an internal node of T that has a child v such that $\varphi(u) = \varphi(v)$. We contract the edge $\{u, v\}$ by merging the two nodes u and v, and renaming the new node v' such that $\varphi(v') = \varphi(v)$. We repeat the process to obtain the resulting tree (T', φ). Note that (T', φ) is a tree with V as its vertex set, and is no longer a net tree or a hierarchical tree. However, observe that if u is an ancestor of v in T, then $\varphi(u)$ is an ancestor of $\varphi(v)$ in T'.

Consider the tree T' with unit weights on its edges. By the definition of g, there is a spanner F on T' that preserves all pairwise distances such that for every pair of nodes, there is a shortest path with at most $g(m, n)$ hops. We add the following set of edges to the spanner E_T.

$$E_F := \{\{\varphi(a), \varphi(b)\} \, : \, \{a, b\} \in F\}.$$

Suppose x and y points in V, x_0 and y_0 are the leaf nodes in T such that $\varphi(x_0) = x$ and $\varphi(y_0) = y$, and x_i and y_i are the ancestors in T at height i for x_0 and y_0 respectively. By Theorem 2.2.9, there exists $\bar{i}$ such that the following points form a $(1 + \epsilon)$-path P_0, after removing repeated points.

$$x = \varphi(x_0), \varphi(x_1), \ldots, \varphi(x_{\bar{i}}), \varphi(y_{\bar{i}}), \ldots, \varphi(y_1), \varphi(y_0) = y$$

Suppose $x_{\bar{i}}$ and $y_{\bar{i}}$ are contracted to $\widehat{x}$ and $\widehat{y}$ respectively in T'. By the choice of F, there exist at most $g(m, n) - 1$ intermediate vertices $\{v_i\}_{i=1}^k$ on the path from x_0 to $\widehat{x}$ in T' such that $\{x_0, v_1\}$, $\{v_i, v_{i+1}\}$ $(1 \leq i < k)$ and $\{v_k, \widehat{x}\}$ are in F. Hence, we have a path with at most $g(m, n)$ hops from x to $\varphi(\widehat{x})$: $x = \varphi(x_0), \varphi(v_1), \varphi(v_2), \ldots, \varphi(v_k), \varphi(\widehat{x})$. Since this sequence of points is a subsequence of $\varphi(x_0), \varphi(x_1), \ldots, \varphi(x_{\bar{i}})$, it follows this length of this path is at most that of the sub-path from $\varphi(x_0)$ to $\varphi(x_{\bar{i}})$ in P_0.

Similarly, there is a path with at most $g(m, n)$ hops from $\varphi(\widehat{y})$ to y whose length is at most that of the corresponding sub-path in P_0. Hence, there is a $(1 + \epsilon)$-path with at most $2g(m, n) + 1$ hops from x to y in the spanner $E_T \cup E_F$.

■

Theorem 2.4.5 (Chazelle [Cha87]) *For $m \geq 2n$, $g(m,n) = O(\alpha(m,n))$, where α is the functional inverse of Ackermann's function.*

Definition 2.4.6 (Ackermann's function [Tar75]) *Let $A(i,j)$ be a function defined for integers $i, j \geq 0$ as the following.*

$$\begin{array}{ll} A(0,j) = 2j & \text{for } j \geq 0 \\ A(i,0) = 0, A(i,1) = 2 & \text{for } i \geq 1 \\ A(i,j) = A(i-1, A(i,j-1)) & \text{for } i \geq 1, j \geq 2 \end{array}$$

Define the function α as $\alpha(m,n) = \min\{i \mid i \geq 1,\ A(i, 4\lceil m/n \rceil) > \log_2 n\}$.

From Lemma 2.4.4 and Theorem 2.4.5, we obtain the following theorem.

Theorem 2.4.7 *Suppose a metric (V,d) with n points has a net tree (T,φ), and suppose E_T is the $(1+\epsilon)$-spanner obtained in Theorem 2.2.4. Then, it is possible to add m extra edges to E_T such that the hop-diameter of the new spanner is at most $O(\alpha(m,n))$.*

Observing that $A(2, 4\log^* n) > \log_2 n$, we have the following corollary.

Corollary 2.4.8 *Suppose a metric (V,d) with n points has a net tree (T,φ), and suppose E_T is the $(1+\epsilon)$-spanner obtained in Theorem 2.2.4. Then, it is possible to add $n\log^* n$ extra edges to E_T such that the hop-diameter of the new spanner is $O(1)$.*

2.4.3 The Lower Bound on Hop-diameter

We now show that the trade-off between the size of the spanner and its hop-diameter obtained in Theorem 2.1.3 is essentially optimal.

Theorem 2.4.9 *For any $\epsilon > 0$, for infinitely many integers n, there exists a metric M induced by n points on the real line such that any $(1+\epsilon)$-spanner with m edges on the metric M has hop-diameter $\Omega(\alpha(m,n))$.*

Our general approach first consider a family of metrics, each of which induced by some binary "hierarchically well-separated tree" (HST). We define a function $G(i,j)$ that is a variant of the Ackermann's function such that if a metric from the family contains $n \geq G(i,j)$ points, then any spanner on the metric with hop-diameter bounded by $i+1$ must have more than $\Omega(jn)$ edges. The relationship between $G(i,j)$ and the Ackermann's function is used to obtain the lower bound for HSTs. The proof technique we used is an extension of that used in Yao's paper [Yao82]. Our lower bound result for line metrics then follows from the fact that binary HSTs with large separation embed into the real line with small distortion.

Remark 2.4.1 *For technical reasons, we assume that a spanner contains a self-loop for every point. Since any spanner must contain a linear number of edges, this assumption does not affect the asymptotic lower bound.*

Construction of the family of HST metrics. For $k \geq 0$, let M_k be the metric induced by the 2^k leaves of the weighted complete binary tree T_k defined as follows. Let $\beta > 0$ be the separation parameter for the HST. The tree T_k is a binary tree containing 2^k leaves such that for each internal node u at height $h \geq 1$, the distance from u to any of the leaves in the subtree rooted at u is β^{h-1}.

The following proposition follows from the construction of the metrics M_k.

Proposition 2.4.10 *Let the HST metric M_k be defined as above.*

(a) Suppose M_k is constructed with separation $\beta \geq 100(1+\epsilon)$. Let U be the subset of points corresponding to the leaves of T_k which are the descendants of some internal node. Then, any $(1+\epsilon)$-path between points in U cannot contain any point outside U.

(b) Consider T_k and suppose $h \leq k$. Suppose T' is the tree obtained from T_k by replacing each subtree rooted at an internal node of height h by a leaf whose distance from the root is the same as before, i.e., β^{k-1}. Then, T' is isomorphic to T_{k-h}.

(c) For every $k \geq 0$, the metric M_k with expansion $\beta \geq 4$ has doubling dimension at most 2.

We will use Proposition 2.4.10(a) crucially in our analysis. Unless otherwise stated, we assume the HST metric M_k is always constructed with separation β large enough such that the statement holds.

We prove the following theorem that states the lower bound result for the HST metrics.

Theorem 2.4.11 *For each integer $k \geq 1$ and any $\epsilon > 0$, there exists an HST metric M_k with large enough separation β such that any $(1+\epsilon)$-spanner on M_k with at most m edges has hop-diameter at least $\Omega(\alpha(m,n))$.*

We observe that HST metrics with large separation embed into the real line with small distortion in the following claim.

Claim 2.4.12 *For each integer $k \geq 1$ and any $\rho > 0$, for sufficiently large $\beta > 0$, the HST metric M_k with separation β embeds into the real line with distortion at most $1+\rho$.*

Proof: We embed the leaves associated with M_k into the real line in their natural ordering, i.e. leaves in the subtree rooted at some internal node are clustered together in the line. The distance between embedded points is the same as that between them in the tree. Such an embedding does not contract distances.

Consider the expansion of the distance between a pair of leaves whose lowest common ancestor is at height r. Hence, their distance in the tree is $2\beta^r$. Observe that their embedded distance is at most $2 \cdot \{2^r + 2^{r-1}\beta + \cdots + 2\beta^{r-1} + \beta^r\}$. Hence, the distortion is at most

$$\begin{aligned} \frac{2^r + 2^{r-1}\beta + \cdots + 2\beta^{r-1} + \beta^r}{\beta^r} &= \frac{2^r}{\beta^r} \cdot \frac{(\beta/2)^r - 1}{\beta/2 - 1} + 1 \\ &\leq \frac{1}{\beta/2 - 1} + 1, \end{aligned}$$

which is at most $1 + \rho$ for $\beta \geq 2(1 + \frac{1}{\rho})$. ∎

Now Theorem 2.4.9, the main result of this section, follows from Theorem 2.4.11 (the result for HSTs) and Claim 2.4.12 (which relates distances in the HST to those on the real line) as follows.

Proof of Theorem 2.4.9: Suppose $n = 2^k$ is a power of two. We construct a line metric M with n points. Let $\epsilon' = 2\epsilon$ and $\rho > 0$ be small enough such that $(1+\epsilon)(1+\rho) \leq 1+\epsilon'$. Suppose the HST metric M_k has large enough separation β such that by Theorem 2.4.11, any $(1+\epsilon')$-spanner for M_k with m edges has hop-diameter $\Omega(\alpha(m,n))$, and by Claim 2.4.12, M_k embeds into some line metric M with distortion at most $1+\rho$.

Suppose P is a $(1+\epsilon)$-spanner for metric M with m edges and hop-diameter at most D. Since $(1+\epsilon)(1+\rho) \leq 1+\epsilon'$, it follows spanner P corresponds to a $(1+\epsilon')$-spanner in M_k with m edges and hop-diameter at most D. Therefore, $D = \Omega(\alpha(m,n))$. ■

In the rest of the section, we will show Theorem 2.4.11, the lower bound result for the HST metrics. To this end, we define a variant of the Ackermann's function.

Definition 2.4.13 *Define the function $G(i,j)$, for $i \geq 0, j \geq 0$ to be:*

$$\begin{array}{ll} G(0,0) = 0, G(0,j) = 2^{\lceil \log_2 j \rceil}; & j \geq 1 \\ G(i,0) = 0, G(i,1) = 1; & i \geq 1 \\ G(i,j) = G(i,j-1)G(i-1, 4G(i,j-1)); & i \geq 1, j \geq 2 \end{array}$$

Proposition 2.4.14 *Suppose $G(i,j)$ is the function defined as above.*

(a) For all $i \geq 0, j \geq 1$, $G(i,j)$ is a power of two.
(b) For $j \geq 1$, $j \leq G(0,j) \leq 2j$.

We now prove the main technical lemma for the lower bound for the HST metrics; as we will see, the proof of Theorem 2.4.11 will follow easily from this lemma.

Lemma 2.4.15 *Suppose $2^k \geq G(i,j)$, where $i \geq 0$ and $j \geq 1$; suppose $\epsilon > 0$ and the HST metric M_k has large enough separation β. Suppose X is a subset of M_k such that $|X| = n \geq 1$. Let $\rho = n/2^k$. Then, any $(1+\epsilon)$-spanner for X with hop-diameter at most $i+1$ must have more than $\frac{1}{4}\rho j n$ edges.*

Proof: We prove the result by induction on the lexicographical order of (i,j).

Base cases. For $i = 0$, $j \geq 1$, any spanner with hop-diameter 1 on n points must have exactly $\frac{1}{2}n(n-1)+n$ edges, recalling that we require that a spanner must contain a self-loop for each point. Hence, observing that $j \leq G(0,j) \leq 2^k$ from Proposition 2.4.14, we conclude that such a spanner cannot have the number of edges less than $\frac{1}{4}\rho j n \leq \frac{1}{4}n^2 < \frac{1}{2}n(n-1)+n$.

For $i \geq 1$, $j = 1$, we observe that any spanner on n points must have at least n edges. Hence, the number of edges in a spanner cannot be less than $\frac{1}{4}\rho n \leq \frac{1}{4}n < n$.

Inductive Step. Suppose X is a subset of M_k such that $2^k \geq G(i,j)$ for some $i \geq 1$ and $j \geq 2$, where $|X| = n$ and $\rho = n/2^k$. For contradiction's sake, assume there is a $(1+\epsilon)$-spanner E with hop-diameter $i+1$ for X such that $|E| \leq \frac{1}{4}\rho j n$.

Let I be the indexing set for the subtrees of T_k, each rooted at some internal node and containing exactly $G(i, j-1)$ leaves. Observing that $G(i, j-1)$ is a power of 2 from Proposition 2.4.14, it follows that

$$\begin{aligned}|I| = 2^k/G(i,j-1) &\geq G(i,j)/G(i,j-1) \\ &= G(i-1, 4G(i,j-1)).\end{aligned}$$

For each $s \in I$, let V_s be the set of leaves contained in the corresponding sub-tree. Let us also define:

- $E_s^1 := \{\{u,v\} \in E : u,v \in V_s\}$, for each $s \in I$, and $E^1 := \cup_{s \in I} E_s^1$.
- $E^2 := \{\{u,v\} \in E : u \in V_s, v \in V_t, s \neq t\}$.

We describe the high level idea to obtain a contradiction. Suppose for each $s \in I$, we replace the subtree containing V_s by a leaf in the same manner as Proposition 2.4.10(b), then we would obtain a tree T' which is isomorphic to $T_{\widehat{k}}$, where $2^{\widehat{k}} = |I| \geq G(i-1, 4G(i,j-1))$.

Let $X_s := X \cap V_s$ and $J := \{s \in I : |X_s| \geq 1\}$. Identifying each X_s's with the corresponding leaf in the modified tree T', consider the submetric of $M_{\widehat{k}}$ induced by the non-empty X_s's, whose point set we write as $X' := \{X_s : s \in J\}$. Hence, X_s is a subset of metric M_k, as well as a point in metric X'.

Define $E' := \{\{X_s, X_t\} : \{u,v\} \in E^2, u \in X_s, v \in X_t\}$. Observe that E' is a $(1+\epsilon)$-spanner for X' with hop diameter at most $i+1$. Since we wish to apply the induction hypothesis, we need to show that the size of E' is small. Moreover, since $|I| \geq G(i-1, 4G(i,j-1))$, the induction hypothesis can only say about spanners of hop-diameter at most i. To resolve this issue, we would remove some points in X' and modify the spanner appropriately such that its hop-diameter is at most i. First observing that $|E'| \leq |E^2|$, it suffices to show that $|E^2|$ is small.

Claim 2.4.16 $|E^2| < \frac{1}{4}\rho n$.

Proof: Let $|X_s| = n_s$ and $\rho_s = n_s/G(i,j-1)$. Observe from Proposition 2.4.10(a) that for each $s \in I$, any $(1+\epsilon)$-path between vertices inside X_s cannot go outside X_s. Hence, for $n_s \geq 1$, it follows E_s^1 is a spanner for X_s having hop-diameter at most $i+1$. Applying the induction hypothesis for $(i, j-1)$, we have for each s, $|E_s^1| > \frac{1}{4}\rho_s(j-1)n_s$. Summing over $s \in I$, we have

$$|E^1| > \sum_{s \in I} \frac{1}{4}\rho_s(j-1)n_s \geq \frac{1}{4} \cdot \frac{j-1}{G(i,j-1)} \sum_{s \in I} n_s^2.$$

Observing that $\sum_{s \in I} n_s = n$ and the fact that $x \mapsto x^2$ is a convex function, the last term is minimized when all n_s's are equal. Hence,

$$|E^1| > \frac{j-1}{4G(i,j-1)} \cdot |I| \cdot (\frac{n}{|I|})^2 = \frac{1}{4}(j-1)\rho n.$$

Since there are at most $\frac{1}{4}\rho j n$ edges in total, it follows that $|E^2| < \frac{1}{4}\rho n$. ■

Next, we describe a procedure that removes some points from X' and modify E' to obtain a spanner with hop-diameter at most i . Note that the points from X' are indexed by J. The procedure labels the removed points *bad*.

1. Place the index set J in a list L in an arbitrary order.
2. Consider each element s in list L according to the ordering:
   (a) If there exists an element t appearing after s in the list L such that any $(1+\epsilon)$-path in E' between X_s and X_t takes at least $i+1$ hops,
      (i) Label s *bad* and remove it from list L.
      (ii) Modify E' so that if X_p is a point in list L closest to X_s, every edge incident on X_s will now be incident on X_p, i.e., X_s and X_p are merged.
   (b) Move on to the next element in list L.

Any two remaining points certainly have a $(1+\epsilon)$-path with at most i hops; otherwise, the one appearing earlier in the list would have been removed. Moreover, observe in step (ii) of the procedure that X_s and X_p are equidistant from any other X_q's in the list. Hence, the length of any $(1+\epsilon)$-path for two points still in the list does not increase. Moreover, since we have merged X_s with X_p, the number of hops for any $(1+\epsilon)$-path cannot increase.

Let B be the set of $s \in J$ that are labelled bad. Let $R := J - B$ be the set of remaining indices. Let $\widehat{E}$ be the modified edge set. It follows that $\widehat{E}$ is a spanner with hop-diameter at most i for $\widehat{X} := \{X_s : s \in R\}$. However, we need to show that not too many bad points are removed.

Claim 2.4.17 $\sum_{s \in R} |X_s| \geq \frac{1}{2}n$.

Proof: For each $s \in B$, there exists $t \in J$ such that any $(1+\epsilon)$-path between X_s and X_t in E' has at least $i+1$ hops. Fix $b \in X_t$ and consider any $a \in X_s$, observe that there is a $(1+\epsilon)$-path P: $a = v_0, v_1, \ldots, v_l = b$ in E such that $l \leq i+1$. For each v, let $\varphi(v)$ be the unique X_q that contains it. Then, it follows there is a $(1+\epsilon)$-path P': $X_s = \varphi(v_0), \varphi(v_1), \ldots, \varphi(v_l) = X_t$, after removing redundant X_q's. Hence, $l = i+1$ and there are no redundant X_q's, otherwise there would be a $(1+\epsilon)$-path from X_s to X_t with less than $i+1$ hops. We associate $a \in X_s$ with the edge $\{a, v_1\} \in E^2$.

It follows for each $s \in B$ and each $a \in X_s$, there exists some edge $\{a, v\} \in E^2$. Each edge can be associated with at most two points in the bad X_s's. Hence, we obtain the following.

$$\sum_{s \in B} |X_s| \leq 2|E^2| < \frac{1}{2}\rho n \leq \frac{1}{2}n,$$

where the middle inequality follows from Claim 2.4.16. Hence, it follows that $\sum_{s \in G} |X_s| \geq \frac{1}{2}n$. ■

We can now obtain a contradiction to the induction hypothesis of Lemma 2.4.15 for $(i-1, 4G(i, j-1))$, which states that if $\hat{X}$ is a sub-metric of $T_{\widehat{k}}$ such that $2^{\widehat{k}} \geq G(i-1, 4G(i, j-1))$

and $\widehat{\rho} = |\widehat{X}|/2^{\widehat{k}}$, then any $(1+\epsilon)$-spanner for $\widehat{X}$ with hop-diameter at most i must have more than $\frac{1}{4}\widehat{\rho}(4G(i,j-1))|\widehat{X}|$ edges.

Now, since for each $s \in R$, $|X_s| \leq G(i,j-1)$, it follows from Claim 2.4.17 that $|\widehat{X}| = |R| \geq \frac{1}{2}n/G(i,j-1)$. Hence, $\widehat{\rho} := |R|/|I| \geq \frac{1}{2}\rho$. Moreover, $n = \rho G(i,j-1)|I| \leq 2|\widehat{X}|\, G(i,j-1)$.

In conclusion, we have a subset $\widehat{X}$ in the metric $T_{\widehat{k}}$ such that $2^{\widehat{k}} = |I| \geq G(i-1, 4G(i,j-1))$ and $\widehat{\rho} = |\widehat{X}|/|I| \geq \rho/2$. Moreover, $\widehat{E}$ is a $(1+\epsilon)$-spanner for $\widehat{X}$ with hop-diameter at most i and has the number of edges less than:

$$\frac{1}{4}\rho n \leq \frac{1}{4} \cdot (2\widehat{\rho}) \cdot 2|\widehat{X}|\, G(i,j-1) = \frac{1}{4}\widehat{\rho}(4G(i,j-1))|\widehat{X}|,$$

obtaining the desired contradiction against the induction hypothesis for $(i-1, 4G(i,j-1))$. This completes the inductive step of the proof of Lemma 2.4.15. ■

If we substitute $\rho = 1$ in Lemma 2.4.15, we obtain the following corollary.

Corollary 2.4.18 *Suppose $n = 2^k \geq G(i,j)$, $j \geq 1$. Let $\epsilon > 0$ and the HST metric M_k have large enough separation β. Then, any $(1+\epsilon)$-spanner for M_k with hop-diameter at most $i+1$ must have more than $\frac{1}{4}jn$ edges.*

In order to get the desired lower bound on the hop-diameter in Theorem 2.4.11, we have to relate the function $G(i,j)$ to the Ackermann function $A(i,j)$; we do this via yet another function $H(i,j)$.

Definition 2.4.19 *Define the function $H(i,j)$, for $i \geq 0, j \geq 0$ to be:*

$$\begin{array}{ll} H(0,j) = 8j^3 & \text{for } j \geq 0 \\ H(i,0) = 0, H(i,1) = 8 & \text{for } i \geq 1 \\ H(i,j) = H(i-1, H(i,j-1)) & \text{for } i \geq 1, j \geq 2 \end{array}$$

Claim 2.4.20 *Let $H(i,j)$ be as defined above.*

- *(a) For $i \geq 0, j \geq 0$, $H(i,j) \leq A(i+4, j+4) - 4$. In particular, $H(i,j) \leq A(i+4, j+4)$.*
- *(b) For $i \geq 0, j \geq 0$, $H(i,j) \geq 4j^2 G(i,j)$. In particular, $H(i,j) \geq G(i,j)$.*

Proof: We prove both results by induction on the lexicographic order of (i,j). Let us prove the claim of part (a) first.

Base cases. For $j \geq 0$, $H(0,j) = 8j^3 \leq A(4, j+4) - 4$. For $i \geq 1$, $H(i,0) = 0 \leq A(i+4, 4) - 4$ and $H(i,1) = 8 \leq A(i+4, 5) - 4$.

Inductive step. Suppose $i \geq 1, j \geq 2$. Then, using the induction hypothesis, we have

$$\begin{aligned} H(i,j) &= H(i-1, H(i,j-1)) \\ &\leq A(i+3, H(i,j-1)+4) - 4 \\ &\leq A(i+3, A(i+4, j+3)) - 4 \\ &= A(i+4, j+4) - 4, \end{aligned}$$

which completes the inductive step of the first result.

We next prove the claim of part (b).

Base cases. For $j \geq 0$, $H(0,j) = 8j^3 \geq 4j^2 G(0,j)$, by Proposition 2.4.14(b). For $i \geq 1$, $H(i,0) \geq 8 \cdot 0^2 G(i,0)$, as both sides are zero; $H(i,1) = 8 \geq 4 = 4G(i,1)$.

Inductive step. Suppose $i \geq 1, j \geq 2$. Then, using the induction hypothesis, we have

$$\begin{aligned} H(i,j) &= H(i-1, H(i,j-1)) \\ &\geq 4H(i,j-1)^2 G(i-1, H(i,j-1)) \\ &\geq 4H(i,j-1)^2 G(i-1, 4(j-1)^2 G(i,j-1)) \end{aligned}$$

Observe that since $i \geq 1$ and $j \geq 2$, $H(i,j-1) \geq 2^{j-1} \geq j$. Hence, $H(i,j) \geq 4j^2 G(i-1, 4G(i,j-1)) = 4j^2 G(i,j)$, completing the induction step of the second result. ■

The following claim describes some properties of the Ackermann function and a functional inverse defined by $a(x,j) := \min\{i \mid i \geq 1, A(i,j) > x\}$; note that this is different from the more commonly used functional inverse α from Definition 2.4.6.

Claim 2.4.21 *Suppose the functional inverse a is defined as above.*

(a) For all $j \geq 0$, if $x \geq y \geq 0$, then $a(x,j) \geq a(y,j)$. In particular, $a(x,j) \geq a(\log_2 x, j)$.
(b) For $k \geq 1$ and $x \geq 0$, $a(x, 4k+4) + 1 \geq a(x, 4k)$.

Proof: The first statement follows trivially from the fact that the Ackermann's function $A(i,j)$ is monotone. For the proof of the second statement, suppose $i = a(x, 4k+4)$. Hence, $i \geq 1$ and $A(i, 4k+4) > x$. Observe that $A(i+1, 4k) = A(i, A(i+1, 4k-1))$ and $A(i+1, 4k-1) \geq 2^{4k-1} \geq 4k+4$, since $k \geq 1$ and $i \geq 1$. Hence, it follows that $A(i+1, 4k) \geq A(i, 4k+4) > x$ and thus $a(x, 4k) \leq a(x, 4k+4) + 1$, as required. ■

We can now prove Theorem 2.4.11 and obtain the lower bound result for the HST metrics.

Proof of Theorem 2.4.11: Suppose E is a $(1+\epsilon)$-spanner E for M_k. Let $j = \lceil \frac{4m}{n} \rceil$. Then, by Corollary 2.4.18, since $m \leq \frac{1}{4} jn$, if $G(i,j) \leq n$, the hop-diameter of E is larger than $i+1$. Hence, the hop-diameter of E is at least the following:

$$\begin{aligned} & \min\{i+1 \mid G(i, \lceil \tfrac{4m}{n} \rceil) > n\} \\ \geq\ & \min\{i+1 \mid H(i, 4\lceil \tfrac{m}{n} \rceil) > n\} && \text{(Claim 2.4.20(b))} \\ \geq\ & \min\{i+1 \mid A(i+4, 4\lceil \tfrac{m}{n} \rceil + 4) > n\} && \text{(Claim 2.4.20(a))} \\ =\ & \min\{i \mid A(i, 4\lceil \tfrac{m}{n} \rceil + 4) > n\} - 3 \\ =\ & a(n, 4\lceil \tfrac{m}{n} \rceil + 4) - 3 \\ \geq\ & a(n, 4\lceil \tfrac{m}{n} \rceil) - 4 && \text{(Claim 2.4.21(b))} \\ \geq\ & a(\log_2 n, 4\lceil \tfrac{m}{n} \rceil) - 4 && \text{(Claim 2.4.21(a))} \end{aligned}$$

The proof is completed from the observation that $a(\log_2 n, 4\lceil \frac{m}{n} \rceil) = \alpha(m,n)$, by the definition of the functions α and a. ■

Chapter 3

Ultra-Low Dimensional Embeddings for Doubling Metrics

3.1 Introduction

We consider the problem of representing a metric (V, d) using a small number of dimensions. Several applications represent data as points in a Euclidean space with thousands of dimensions. However, this high-dimensionality poses significant computational challenges: many algorithms tend to have an exponential dependence on the dimension. Hence we are constantly seeking ways to combat this so-called *curse of dimensionality*, by finding low-dimensional yet faithful representations of the data. In this work, we attempt to maintain all pairwise distances, i.e. we seek to minimize the *distortion* of an embedding.

This computational motivation leads one to an already compelling and fundamental mathematical question: *given a metric space* (which may or may not be Euclidean to begin with), *what is the least number of dimensions in which it can be represented with "reasonable" distortion?*

To answer these questions, dimension reduction in Euclidean spaces have been studied extensively. The celebrated and surprising "flattening" lemma of Johnson and Lindenstrauss [JL84] states that the dimension of any Euclidean metric on n points can be reduced to $O(\frac{\log n}{\epsilon^2})$ with $(1+\epsilon)$ distortion, and moreover, this can be done via a random linear map. This result is existentially tight: a simple packing argument shows that any distortion-D embedding of a uniform metric on n points into Euclidean space requires at least $\Omega(\log_D n)$ dimensions—intuitively, there aren't enough distinct directions in a low dimensional Euclidean space to accommodate a large number of equidistant points. Hence we do need the $\Omega(\log n)$ dimensions, and even allowing $O(\log n)$ distortion cannot reduce the number of dimensions below $\Omega(\log n / \log\log n)$.

It is natural to ask if this "volume" restriction is the only bottleneck to a low-dimensional embedding. In other words, can metrics that do not have such volume hurdles be embedded into low-dimensional spaces with small distortion? The notion of *doubling dimension* [Ass83] makes this very idea concrete: roughly speaking, a metric has doubling dimension $\dim_D = k$ if and only if it has (nearly-)uniform submetrics of size about 2^k, but no larger. A metric (or more strictly, a family of metrics) is simply called *doubling* if the doubling dimension is bounded by a universal constant. (See section 3.1.2 for a more precise definition).

The Questions. The packing lower bound shows that any metric requires $\Omega(\dim_D)$ dimensions for a constant-distortion embedding into Euclidean space: is this lower bound tight? We now know the existence of n-point metrics with $\dim_D = O(1)$ that require $\Omega(\sqrt{\log n})$-distortion into Euclidean space (of any dimension) [GKL03], but can we actually achieve this distortion with $o(\log n)$-dimensions? What if we give up a bit in the distortion? Bourgain's classical result (along with the JL-lemma) shows that all metrics embed into Euclidean space of $O(\log n)$ dimensions and $O(\log n)$ distortion [LLR95], but we do not even know if doubling metrics embed into $O(\log^{1-\epsilon} n)$ dimensions with $O(\log^{1-\epsilon} n)$ distortion.

If we restrict our attention to Euclidean doubling metrics, we know just as little: a tantalizing conjecture of Lang and Plaut [LP01] states that all Euclidean metrics with $\dim_D = O(1)$ embed into $O(1)$ dimensional Euclidean space with $O(1)$ distortion. However, the best result we know is still the JL-Lemma (which is completely oblivious to the doubling dimension, and moreover, is a linear map which is doomed to fail). Again, we do not even know how to take a doubling *Euclidean* point set and flatten it into (say) $O(\log^{1-\epsilon} n)$ dimensions with $O(\log^{1-\epsilon} n)$ distortion!

The Answers. We make progress on the problem of embedding doubling metrics into Euclidean space with small dimension and distortion. (Our results hold for *all* doubling metrics, not just Euclidean ones.)

Theorem 3.1.1 (Ultra-Low-Dimension Embedding) *Any metric space with doubling dimension $\dim_D$ embeds into $O(\dim_D \log\log n)$ dimensions with $O(\log n/\sqrt{\log\log n})$ distortion.*

Hence we can embed the metric into *very few* Euclidean dimensions (i.e., $\tilde{O}(\dim_D)$, where the notation $\tilde{O}(\cdot)$ suppresses a multiplicative factor polynomial in $\log\log n$), and achieve a slightly smaller distortion than even Bourgain's embedding. Note that to achieve distortion $O(\log n)$, any metric with doubling dimension $\dim_D$ requires at least $\Omega(\frac{\dim_D}{\log\log n})$ Euclidean dimensions, and hence we are within an $O(\log\log n)^2$ factor to the *optimal dimension* for this value of distortion.

This is a special case of our general trade-off theorem:

Theorem 3.1.2 (Main Theorem) *Suppose (V, d) is a metric space with doubling dimension $\dim_D$. For any integer T such that $\Omega(\dim_D \log\log n) \leq T \leq \ln n$, there exists $F : V \to \mathbb{R}^T$ into T-dimensional space such that for all $x, y \in V$, $d(x,y) \leq \|F(x) - F(y)\|_2 \leq O\left(\sqrt{\frac{\dim_D}{T}}\ \log n\right) \cdot d(x,y)$.*

Varying the target dimension T, we can get some interesting tradeoffs between the distortion and dimension. For instance, we can balance the two quantities and get $O(\log^{2/3} n)$ dimensions and $O(\log^{2/3} n)$ distortion for doubling metrics, as desired. On the other hand, for large target dimension $T = \ln n$, we get distortion $O(\sqrt{\dim_D \log n})$, which matches the best known result from [KLMN05].

In the interests of clarity of presentation, we only show the *existence* of such embeddings. Standard techniques (e.g,. [Bec91, Alo91, MR98]) can be used to give algorithmic versions of our results.

Techniques. Our embedding can best be thought of as an extension of Rao's embedding [Rao99]: there are $O(\log n)$ copies of coordinates for each distance scale, hence leading to $O(\log n \log \Delta)$ dimensions. As observed in [ABN06], it is possible to sum up the coordinates over different distance scales to form one coordinate, and in expectation the contraction is bounded. Using bounded doubling dimension, we show that there is limited dependency between pairs of points (using the Lovasz Local Lemma), and hence we only need much less than $O(\log n)$ coordinates to ensure that the contraction for all points are bounded.

For the tradeoff between the target dimension and the distortion, we apply a random sign (± 1) to the contribution for each distance scale before summing them up to form a coordinate. This process is analogous to the random projection in JL-type embeddings. Indeed, we use analysis similar to that in [Ach00] to obtain a tradeoff between the target dimension and the expansion, although in our case the original metric needs not be Euclidean.

We give two embeddings: the first one uses a simple decomposition scheme [GKL03, Tal04, CGMZ05] and illustrates the above ideas in bounding both the contraction and the expansion. The resulting embedding has distortion $O(\dim_D / \sqrt{T} \cdot \log n)$ with T dimensions. In order to reduce the dependence on the doubling dimension to $\sqrt{\dim_D}$, we use *uniform* padded decomposition schemes based on [ABN06].

Bibliographic Note. Independently of our work, Abraham, Bartal, and Neiman (personal communication) have obtained results of a very similar nature, showing how to achieve a trade-off between distortion and dimension as a function of the doubling dimension $\dim_D$ and the number of points n. We believe their results are incomparable to ours. For instance, they can achieve $O(\dim_D)$-dimensional embeddings—smaller than ours by an $O(\log \log n)$ factor—though only with slightly super-logarithmic distortion.

Normally, for a pair of points, conventional techniques bound its contraction using only one distance scale. In order to apply the Local Lemma, the probability of the associated bad event has to be small enough (see Lemma 3.2.7) and hence we need $O(\log \log n)$ dimensions. Their idea is to use $O(\log \log n)$ distance scales to bound the contraction. Hence, they do not need the $O(\log \log n)$ factor in the dimension, but the distortion would suffer an extra factor of $O(\log^\epsilon n)$.

However, we use random signs in our embedding to bound the expansion and consequently our trade-off at the higher end of dimension is slightly better than theirs. They also present results on gracefully degrading distortion and average distortion (in the sense defined in [ABC$^+$05, ABN06]).

Moreover, they also show explicitly how to apply techniques [Alo91, MR98] of getting an algorithmic version of the Local Lemma to construct such an embedding in time $k^{2^{O(k)} \log \log n}$, where $k = \dim_D$. Hence, for $\dim_D = o(\log \log n)$, we have a polynomial time algorithm; for $\dim_D = o(\log n)$, we have a sub-exponential time algorithm.

3.1.1 Related Work

Dimension reduction for Euclidean space was first studied by Johnson and Lindenstrauss [JL84], using random projections. The results and techniques have since been sharpened and simplified in [FM88, IM98, DG03, Ach00, AC06]. The embeddings have been derandomized, see [EIO02, Siv02]. Moreover, Matousek [Mat90] has obtained an almost tight tradeoff between the dimension of the target space and the distortion of the embedding. On the other hand, dimension reduction for L_1 space has been shown to be much harder in [BC03, LN03].

The notion of doubling dimension was introduced by Larman [Lar67] and Assouad [Ass83], and first used in algorithm design by Clarkson [Cla99]. The properties of doubling metrics and their algorithmic applications have since been studied extensively, a few examples of which appear in [GKL03, KL03, KL04, Tal04, HPM05, BKL06, CG06b, IN, KRX06, KRX07].

There is extensive work on metric embeddings, see [IM04]. Bourgain [Bou85] gave an embedding whose coordinates are formed by distances from random subsets. Low diameter decomposition is a useful tool and was studied by Awerbuch [Awe85], and Linial and Saks [LS93]. Randomized decompositions for general metrics are given in [Bar96, CKR01, FRT04]. Klein et al. [KPR93] gave decomposition schemes for minor-excluding graphs, which were used by Rao [Rao99] to obtain embeddings for planar graphs into Euclidean space. These ideas were developed further in [KLMN05, ABC$^+$05, ABN06].

On the other hand, there is also research on embeddings into constant dimensional spaces, both for general metrics [BCIS05] and special classes of metrics, for instance ultra-metrics [BCIS06].

3.1.2 Notation and Preliminaries

The reader is referred to standard texts—e.g.,[DL97, Mat02]—for basic definitions of metric spaces. We denote a finite metric space by (V, d), its size by $n = |V|$, and its doubling dimension $\dim_D$ by k. We assume that the minimum distance between two points is 2 (somewhat weird!), and hence its diameter Δ is also (almost) the aspect ratio of the metric. A *ball* $B(x, r)$ is the set $\{y \in V \mid d(x, y) \leq r\}$.

Definition 3.1.3 (Nets) *Given a metric (V, d) and $r > 0$, an r-net N for (V, d) is a subset of V such that*

1. *(Covering Property) For all $x \in V$, there exists $y \in N$ such that $d(x, y) \leq r$.*
2. *(Packing Property) For all $x, y \in N$ such that $x \neq y$, $d(x, y) > r$.*

Definition 3.1.4 (Doubling Dimension $\dim_D$) *The* doubling dimension *of a metric (V, d) is at most k if for all $x \in V$, for all $r > 0$, every ball $B(x, 2r)$ can be covered by the union of at most 2^k balls of the form $B(z, r)$, where $z \in V$.*

Definition 3.1.5 (Padded Decompostion) *Given a finite metric space (V, d), a positive parameter $D > 0$ and $\alpha > 1$, a D-bounded α-*padded decomposition *is a distribution Π over partitions of V such that the following conditions hold.*

(a) For each partition P in the support of Π, the diameter of every cluster in P is at most D.

(b) Suppose $S \subseteq V$ is a set with diameter d. If P is sampled from Π, then the set S is partitioned by P with probability at most $\alpha \cdot \frac{d}{4D}$

Note. We only need a weaker condition implied by item (b): if we set $S := B(x, \frac{D}{\alpha})$, then the ball is partitioned by P with probability at most $\frac{1}{2}$. In other words, we have $Pr[B(x\frac{D}{\alpha}) \subseteq P(x)] \geq \frac{1}{2}$, where $P(x)$ is the cluster in P containing x.

3.2 The Basic Embedding

We give two embeddings: the one from this section is the basic embedding, which achieves the following trade-off between dimension and distortion:

Theorem 3.2.1 (The Basic Embedding) *Given a metric space (V, d) with doubling dimension $\dim_D$, and a target dimension T in the range $\Omega(\dim_D \log\log n) \leq T \leq \ln n$, there exists a mapping $f : V \to \mathbb{R}^T$ such that for all $x, y \in V$, $\Omega\left(\frac{\sqrt{T}}{\dim_D}\right) \cdot d(x,y) \leq ||f(x) - f(y)||_2 \leq O(\log n) \cdot d(x,y)$. Hence, the distortion is $O(\frac{\dim_D \log n}{\sqrt{T}})$.*

Note that this trade-off is slightly worse than than the one claimed in Theorem 3.1.2 in terms of its dependence on the doubling dimension; however, the advantage is that this embedding is easier to state and prove. We will then improve on this embedding in the next section.

3.2.1 Basic Embedding: Defining The Embedding

The embedding $f : (V, d) \to \mathbb{R}^T$ we describe is of the form $f := \oplus_{t\in[T]} \Phi^{(t)}$, where the symbol $\oplus$ is used to denote the concatenation of the various coordinates. Each $\Phi^{(t)} : V \to \mathbb{R}$ is a single coordinate generated independently of the other coordinates according to a probability distribution described as follows. To simplify notation, we drop the superscript t and describe how a random map $\Phi : V \to \mathbb{R}$ is constructed, and f is just the concatenation of T such coordinates.

Let $D_i := H^i$, for some constant $H \geq 2$. (Later we see that H is set large enough to bound the contraction.) Suppose all distances in the metric space are at least 2, and I is the largest integer such that $D_I < \Delta$. The mapping $\Phi : V \to \mathbb{R}$ is of the form $\Phi := \sum_{i\in[I]} \varphi_i$. We describe how $\varphi_i : V \to \mathbb{R}$ is constructed, for each $i \in [I]$.

Fix $i \in [I]$. We view the metric (V, d) as a weighted complete graph, and contract all edges with lengths at most $D_i/2n$. The points that are contracted together in this process would obtain the same value under φ_i. Let the resulting metric be (V, d_i). Here are a few properties of the metric (V, d_i).

Proposition 3.2.2 *Suppose for each $i \in [I]$, the metric (V, d_i) is defined as above. Then, for all $x, y \in V$, the following results hold.*

(a) For all $i \in [I]$, $d_i(x,y) \leq d(x,y) \leq d_i(x,y) + \frac{D_i}{2}$.
(b) For $j \geq i$, $d_j(x,y) \leq d_i(x,y)$.

Observe that Property (a) of Proposition 3.2.2 implies that the metric (V, d_i) gives good approximations of the distances in (V, d) of scales above D_i. In particular, (V, d_i) admits an $O(k)$-padded

D_i-bounded stochastic decomposition.

Proposition 3.2.3 (Padded Decomposition for Doubling Metrics [GKL03, Tal04, CGMZ05]) *Suppose the metric (V, d) has doubling dimension k. Then, there is an α-padded D_i-bounded stochastic decomposition Π_i for the metric (V, d_i), where $\alpha = O(k)$. Moreover, the event $\{B_i(x, D_i/\alpha) \subseteq P_i(x)\}$ is independent of all the events $\{B_i(z, D_i/\alpha) \subseteq P_i(z) : z \notin B_i(x, 3D_i/2)\}$, where $B_i(u, r) := \{v \in V : d_i(u, v) \leq r\}$.*

Suppose P_i is a random partition of (V, d_i) sampled from the padded decomposition Π_i of Proposition 3.2.3. Let $\{\sigma_i(C) : C \text{ is a cluster in } P_i\}$ be uniform $\{0, 1\}$-random variables, and γ_i be a uniform $\{-1, 1\}$-random variable. The random objects P_i, σ_i and γ_i are sampled independently of one another. Define $\varphi_i : V \to \mathbb{R}$ by

$$\varphi_i(x) := \gamma_i \cdot \sigma_i(P_i(x)) \cdot \min\{d_i(x, V \setminus P_i(x)), D_i/\alpha\} \tag{3.1}$$

Hence we take the distance from the point x to the closest point outside its cluster, truncate it at D_i/α (recall that α is as defined in Proposition 3.2.3), and multiply it with the $\{0, 1\}$ r.v. associated with its cluster, and the $\{-1, 1\}$ r.v. associated with the distance scale i. (For brevity, we will use the expression $\kappa_i(x) := \sigma_i(P_i(x)) \cdot \min\{d_i(x, V \setminus P_i(x)), D_i/\alpha\}$; hence $\varphi_i(x) = \gamma_i \cdot \kappa_i(x)$.) We shall see that the σ_i's play an important role in bounding the contraction, while the role of γ_i's is to bound the expansion.

To summarize, the embedding is defined to be:

$$f := \oplus_{t \in [T]} \Phi^{(t)}; \Phi^{(t)} := \sum_{i \in [I]} \varphi_i^{(t)}. \tag{3.2}$$

We rephrase Theorem 3.2.1 in terms of the above randomized construction.

Theorem 3.2.4 *Suppose the input metric (V, d) has doubling dimension k, and the target dimension T is in the range $\Omega(k \log\log n) \leq T \leq \ln n$. Then, with non-zero probability, the above procedure produces a mapping $f : V \to \mathbb{R}^T$ such that for all $x, y \in V$, $\Omega\left(\frac{\sqrt{T}}{\dim_D}\right) \cdot d(x, y) \leq ||f(x) - f(y)||_2 \leq O(\log n) \cdot d(x, y)$. In other words, there exist some realization of the various random objects such that the distortion of the resulting mapping is $O(\frac{\dim_D \log n}{\sqrt{T}})$.*

Note. Before we dive in, let us note that we consider the modified metrics (V, d_i) in order to avoid a dependence on the aspect ratio Δ in the expansion bound for the embedding. Now observe that $|\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y)| \leq \min\{d_j(x, y), D_j/\alpha\}$.

Lemma 3.2.5 *Suppose $x, y \in V$ and for each $j \in [I]$, define $d_j := \min\{d_j(x, y), D_j/\alpha\}$. Then,*

- *(a) For each $i \in [I]$, $\sum_{j \geq i} d_j \leq O(\log_H n) \cdot d_i(x, y)$.*
- *(b) For each $i \in [I]$, $\sum_{j \geq i} d_j^2 \leq O(\log_H n) \cdot d_i(x, y)^2$.*

In particular, for all $t \in [T]$, the contribution $|\sum_{j \geq i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \leq \sum_{j \geq i} d_j \leq O(\log_H n) \cdot d_i(x, y)$.

Moreover, $\sum_{i \in [I]} d_i^2 \leq O(\log_H n) \cdot d(x, y)^2$.

Proof: We prove statements (a) and (b). The other statements follow from the two in a straight forward manner.

For ease of notation, we omit the superscript t in this proof. Observe that for $j \geq i$, $d_j \leq d_j(x,y) \leq d_i(x,y)$, where the second inequality follows from Proposition 3.2.2(b).

There are three cases to consider depending on the value of j. The first is for very large j's when $d(x,y) \leq \frac{D_j}{2n}$: in this case, $d_j(x,y) = 0$. The second case is for moderate values of j when $\frac{D_j}{2n} < d(x,y) \leq D_j$: there are at most $O(\log_H n)$ such j's. In (a), adding these up gives a contribution of $O(\log_H n) \cdot d_i(x,y)$; in (b), we have a contribution of $O(\log_H n) \cdot d_i(x,y)^2$.

Finally, the last case is for small values of j, when $d(x,y) > D_j$. Consider the largest j_0 for which this happens. Then, it follows from Proposition 3.2.2 that $d_i(x,y) \geq d_{j_0}(x,y) > D_{j_0}/2$. Observing that $d_j \leq D_j/\alpha$ and $\{D_j\}$ forms a geometric sequence, it follows that $\sum_{i \leq j \leq j_0} d_j = O(d_i(x,y))$, and $\sum_{i \leq j \leq j_0} d_j^2 = O(d_i(x,y)^2)$.

Combining the three cases gives the result. ■

3.2.2 Basic Embedding: Bounding Contraction

A natural idea to bound the contraction for a particular pair of points x, y is to use the padding property of the random decomposition: if $d(x,y) \approx H^i$, then at the corresponding scale $i \in [I]$ the two vertices will be in different clusters, and will contribute a large distance. This idea has been extensively used in previous work starting with [Rao99]. However, in these previous works, we have a separate coordinate for each distance scale, which leads to a large number of dimensions. Abraham et al. [ABN06] show that the coordinates for distance scales can actually be combined to form one single coordinate, and with constant probability the contraction is still bounded. Now we want to use a small number of coordinates as well: to do this, we exploit small doubling dimension to use the Lovasz Local Lemma and bound the contraction for all pairs of points.

Fixing the γ's. As noted in the description of the embedding, the γ's do not play any role in bounding the contraction. In fact, we will show something *stronger*: for any realization of the γ's, there exists some realization of the P's and σ's for which the contraction of the embedding f is bounded. For the rest of this section, we assume that the γ's are arbitrarily fixed upfront.

For each $i \in [I]$, let the subset N_i be an arbitrary βD_i-net of (V, d_i), for some $0 < \beta < 1$ to be specified later.

Bounding the Contraction for some Special Points. We first bound the contraction for the pairs in $E_i := \{(x,y) \in N_i \times N_i : 3D_i/2 < d_i(x,y) \leq 3HD_i\}$, $i \in [I]$. (Note that from Proposition 3.2.2(a), it follows that for each $(x,y) \in E_i$, $d(x,y) < 4HD_i$.)

For $t \in [T]$, and $(x,y) \in E_i$, define $A^{(t)}(x,y)$ to be the event that *all the following* happens:

- the vertex x is well-padded: i.e., $B_i(x, \frac{D_i}{\alpha}) \subseteq P_i^{(t)}(x)$;
- the vertex y is mapped to 0: $\sigma_i^{(t)}(P_i^{(t)}(y)) = 0$;
- if $|\sum_{j>i}(\varphi_i^{(t)}(x) - \varphi_i^{(t)}(y))| \leq \frac{D_i}{2\alpha}$, then $\sigma_i^{(t)}(P_i^{(t)}(x)) = 1$, otherwise $\sigma_i^{(t)}(P_i^{(t)}(x)) = 0$.

Proposition 3.2.6 (Conditioning on Higher Levels) *Let $(x, y) \in E_i$. Suppose for $j > i$, the random objects $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : t \in [T]\}$ have been arbitrarily fixed. For each $t \in [T]$, sample random partition $P_i^{(t)}$ from Proposition 3.2.3 and random $\{0,1\}$-variables $\{\sigma_i^{(t)}(C) : C$ is a cluster of $P_i^{(t)}\}$ uniformly, all independently of one another. Then, for each $t \in [T]$, with probability at least $\frac{1}{8}$, the event $A^{(t)}(x, y)$ happens independently over the different t's.*

Moreover, if the event $A^{(t)}(x, y)$ happens, then the inequality $|\sum_{j \geq i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \geq \frac{D_i}{2\alpha}$ holds; furthermore, for any realization of the remaining random objects, i.e., $\gamma_i^{(t)}$ and $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : j < i\}$, the inequality $|\sum_{i \in [I]}(\varphi_i^{(t)}(x) - \varphi_i^{(t)}(y))| \geq \frac{D_i}{4\alpha}$ holds, provided $H \geq 8$. (Recall that $D_{i+1} = HD_i$.)

Proof: Given any realization of the random objects of scales larger than i, each of the three defining events for $A^{(t)}(x, y)$ happens independently of one another with probability at least $\frac{1}{2}$, and hence $A^{(t)}(x, y)$ happens with probability at least $\frac{1}{8}$, independently over $t \in [T]$, since the random objects at scale i are sampled independently over $t \in [T]$.

It follows that if $A^{(t)}(x, y)$ happens, then the partial sum from large scales up to scale i is $|\sum_{j \geq i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \geq \frac{D_i}{2\alpha}$. Observe the sum from smaller scales $|\sum_{j<i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))|$ is bounded above by a geometric sum $\sum_{j<i} \frac{D_j}{\alpha}$, which is at most $\frac{D_i}{4\alpha}$, provided that $H \geq 8$. ■

In order to show that the contraction for the pair (x, y) is small, we need to show that the event $A^{(t)}(x, y)$ happens for a constant fraction of t's. We define $C(x, y)$ to be the event that for at least $\frac{T}{16}$ values of t, the event $A^{(t)}(x, y)$ happens. We conclude that the event $C(x, y)$ happens with high probability (as a function of T), by using a Chernoff bound: if X is the sum of i.i.d. Bernoulli random variables, then $Pr[X < (1-\epsilon)E[X]] \leq \exp(-\frac{1}{2}\epsilon^2 E[X])$, for $0 < \epsilon < 1$.

Proposition 3.2.7 (Using Concentration) *Under the sampling procedure described in Proposition 3.2.6, the event $C(x, y)$ fails to happen with probability at most $p := \exp(-\frac{T}{64})$.*

Proof: This follows by applying the Chernoff bound mentioned above with $\epsilon = \frac{1}{2}$. ■

Now that each event $C(x, y)$ happens with high enough probability, we use the Lovasz Local Lemma to show that there is some realization of $\{P_i^{(t)}, \sigma_i^{(t))} : t \in [T]\}$ such that for all $(x, y) \in E_i$, the *events $C(x, y)$ happen simultaneously*. In order to use the Local Lemma, we need to analyze the dependence of these events. Recall that N_i is a βD_i-net of (V, d_i).

Lemma 3.2.8 (Limited Dependence) *For each $(x, y) \in E_i$, the event $C(x, y)$ is independent of all but $B := (\frac{H}{\beta})^{O(k)}$ of the events $C(u, v)$, where $(u, v) \in E_i$.*

Proof: Observe that the event $C(x, y)$ is determined by the random objects $\{P_i^{(t)}, \sigma_i^{(t)} : t \in [T]\}$. More specifically, it is determined completely by the events $\{B_i(w, \frac{D_i}{\alpha}) \subseteq P_i^{(t)}(w) : t \in [T]\}$ and $\{\sigma_i^{(t)}(P^{(t)}(w)) = 0 : t \in [T]\}$, for $w \in \{x, y\}$. Note that if $d_i(x, w) > 3D_i/2$, then the corresponding events for the points x and w are independent. Note that if $d_i(x, w) \leq 3D_i/2$, then $d(x, w) \leq 2D_i$; moreover, any two net-points in (V, d_i) must be more than βD_i apart in (V, d). Hence, observing that the doubling dimension of the given metric is at most k, for each of x and y, only $(\frac{2D_i}{\beta D_i})^{O(k)}$ net points are relevant. Now, each net point can be incident by at most $(\frac{4H}{\beta})^{O(k)}$ edges in E_i. Hence, it follows that $C(x,y)$ is independent of all but $(\frac{H}{\beta})^{O(k)}$ of the events $C(u,v)$, where $(u,v) \in E_i$. ■

Now we can apply the (symmetric form of the) Lovasz Local Lemma.

Lemma 3.2.9 (Lovasz Local Lemma) *Suppose there is a collection of events such that each event fails with probability at most p. Moreover, each event is independent of all but B other events. Then, if $ep(B+1) < 1$, then all the events in the collection happen simultaneously with non-zero probability.*

Proposition 3.2.10 (One More Level) *Suppose for $j > i$, the random objects $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : t \in [T]\}$ have been arbitrarily fixed. If $T = \Omega(k \log \frac{H}{\beta})$, then there is some realization of $\{P_i^{(t)}, \sigma_i^{(t)} : t \in [T]\}$ such that all the events $\{C(x,y) : (x,y) \in E_i\}$ happen. In particular, such a realization does not depend on the γ's at scale i.*

Proof: From Proposition 3.2.7, the failure probability for each event $C(x,y)$ is at most $p := \exp(-\frac{T}{64})$ and from Lemma 3.2.8, the number of dependent events is at most $B = (\frac{H}{\beta})^{O(k)}$. Hence, setting $\Omega(k \log \frac{H}{\beta})$, we have $ep(B+1) < 1$, and we can apply the Local Lemma. ■

Define $\mathcal{E}$ to be the event that for all $i \in [I]$, for all $(x,y) \in E_i$, the event $C(x,y)$ happens. By applying Proposition 3.2.10 repeatedly, we show that the event $\mathcal{E}$ happens with non-zero probability.

Proposition 3.2.11 (Contraction for Nearby Net Points) *Suppose in the construction the γ's are arbitrarily fixed, and the P's and σ's are still random and independent. Moreover, suppose $T = \Omega(k \log \frac{H}{\beta})$. Then, with non-zero probability, our random construction produces an embedding $f : (V,d) \to \mathbb{R}^T$ such that the event $\mathcal{E}$ happens; in particular, there exists some realization of the P's and σ's such that $||f(x) - f(y)||_2 \geq \frac{\sqrt{T}}{4} \cdot \frac{D_i}{4\alpha}$.*

Proof: For each $i \in [I]$, let $\mathcal{E}_i$ denote the event that for all $(x,y) \in E_i$, the event $C(x,y)$ happens. Then, we have $\mathcal{E} = \cap_{i \in [I]} \mathcal{E}_i$.

From Proposition 3.2.10, we have for all $i \in [I]$, $Pr[\mathcal{E}_i | \cap_{j \geq i+1} \mathcal{E}_j] > 0$. Hence, we have $Pr[\mathcal{E}] = \prod_{i \in [I]} Pr[\mathcal{E}_i | \cap_{j \geq i+1} \mathcal{E}_j] > 0$. ■

Bounding the Contraction for All Points. We next bound the contraction for an arbitrary pair (u,v) of points noting that if all net points do not suffer large contraction (by the above argument), and all pairs do not incur a large expansion (by the argument of Lemma 3.2.5), then one can extend the contraction result to all pairs of points. Of course, to do so, the net N_i must be sufficiently fine. Recall that N_i is a βD_i-net for (V, d_i).

Lemma 3.2.12 (Extending to All Pairs) *Suppose the event $\mathcal{E}$ happens. Then, for any $x, y \in V$, there exist $T/16$ values of t's for which*

$$|\Phi^{(t)}(x) - \Phi^{(t)}(y)| = \Omega(d(x,y))/\alpha H.$$

Proof: We can assume $\beta < 1/4$. Let $i \in [I]$ such that $(2+2\beta)D_i \leq d(x,y) \leq (2+2\beta)D_{i+1}$. Suppose $u,v \in N_i$ are the net points such that $d_i(x,u) \leq \beta D_i$ and $d_i(y,v) \leq \beta D_i$. Then, it follows that $(u,v) \in E_i$. Since the event $\mathcal{E}$ happens, the event $C(u,v)$ also occurs, and so there are at least $T/16$ values of t's for which the event $A^{(t)}(u,v)$ occurs. We show that for each such t, $|\Phi^{(t)}(x) - \Phi^{(t)}(y)| = \Omega(d(x,y))/\alpha H$.

Since $A^{(t)}(u,v)$ occurs, it follows that $|\sum_{j\geq i}(\varphi_j^{(t)}(u) - \varphi_j^{(t)}(v))| \geq D_i/2\alpha$. Now using Lemma 3.2.5, it follows that $|\sum_{j\geq i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(u))| \leq O(\log_H n)\cdot d_i(u,x) \leq O(\log_H n)\cdot \beta D_i \leq D_i/8\alpha$, for sufficiently small β, where $\frac{1}{\beta} = \Theta(\alpha \log_H n)$. The same upper bound holds for $|\sum_{j\geq i}(\varphi_j^{(t)}(y) - \varphi_j^{(t)}(v))|$. Hence, since the net points u,v were "far apart", and both x and y were close to their net points, we can use the triangle inequality to infer that $|\sum_{j\geq i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \geq D_i/4\alpha$.

Finally, observing that for $j < i$, $|\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y)| \leq D_j/\alpha = \frac{D_i}{\alpha H^{i-j}}$ and $H \geq 16$, we have $|\sum_{j<i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \leq D_i/8\alpha$. Therefore, $|\sum_{j\in[I]}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \geq D_i/8\alpha$, as required. ■

Hence, by setting $H = 16$ and $\frac{1}{\beta} = \Theta(\alpha \log_H n)$, and observing $\alpha = O(k)$ from Proposition 3.2.3 (where k is the doubling dimension and is at most $\log n$), we have the following result.

Proposition 3.2.13 (Bounding Contraction) *Suppose the γ's are arbitrarily fixed and β is sufficiently small such that $\frac{1}{\beta} = \Theta(\alpha \log_H n)$ and $H \geq 16$. Then, for $T = \Omega(k \log\log n)$, there exists some realization of P's and σ's that produces an embedding $f : V \to \mathbb{R}^T$ such that for all $x,y \in V$, $\|f(x) - f(y)\|_2 \geq \Omega(\frac{\sqrt{T}}{k}) \cdot d(x,y)$.*

3.2.3 Basic Embedding: Bounding Expansion

Recall that $\mathcal{E}$ is the event $\cap_{i\in[I]}\cap_{(x,y)\in E_i} C(x,y)$. We showed in Proposition 3.2.11 that $Pr[\mathcal{E}] > 0$, and if the event $\mathcal{E}$ happens, the resulting embedding $f : V \to \mathbb{R}^T$ has bounded contraction. We now bound the expansion of the embedding $f : V \to \mathbb{R}^T$ for every pair (x,y) of points. In order to bound this expansion, the $\{-1,+1\}$-random variables γ_i will finally be used. Their role is fairly natural: if the contributions from different distance scales are simply summed up, then there would be a factor of $|I|$ (roughly speaking) appearing in the expansion for each coordinate. However, with the random variables γ_i's, the sum starts to behave like a random walk, and the expectation of the sum of the signed contributions would only suffer a factor of $\sqrt{I}$. In order to make this argument formal, we use techniques similar to those used in analyzing the Johnson-Lindenstrauss lemma [Ach00]. The main problem that arises here is that if we condition on the event $\mathcal{E}$, not only the different coordinates of the map but also the γ's are no longer independent, and hence we would not be able to use the "random walk"-like argument. Therefore, we need a more careful analysis to apply the large-deviation arguments.

Fixing the P's and σ's. Suppose the γ's are sampled uniformly and independently. From Proposition 3.2.13, there exists some realization of the P's and the σ's such that the contraction of the embedding f is bounded. Hence, from this point, we can concentrate on bounding the expansion. Since the γ's are randomly drawn, the P's and the σ's are random variables too, and are functions of the γ's. Proposition 3.2.10 gives a clear idea of the dependency between the random variables: the P's and the σ's at scale i are determined only by the random objects at scales strictly larger than i, and in particular are independent of the γ's at scale i.

Let us fix $x, y \in V$ and define the random variable

$$S := ||f(x) - f(y)||_2^2 = \sum_{t \in [T]} (Q^{(t)})^2,$$

where $Q^{(t)} := \Phi^{(t)}(x) - \Phi^{(t)}(y)$. (The coordinates Φ were defined in (3.1). We want to show that for large enough T, the r.v. S does not deviate too much from its mean with high probability. Then, a union bound over all pairs (x, y) of points leads to the conclusion that with non-zero probability, the embedding f has bounded expansion.

Observe that $Q^{(t)} := \sum_{i \in [I]} \gamma_i^{(t)} Y_i^{(t)}$, where $Y_i^{(t)} := \kappa_i^{(t)}(x) - \kappa_i^{(t)}(y)$. Define $d_i := \min\{d_i(x, y), D_i/\alpha\}$. Recall that the random variables $\gamma_i^{(t)}$ are uniformly picked from $\{-1, +1\}$, and $|Y_i^{(t)}| \leq d_i$.

We can illustrate the dependency between the different random objects in the following description.

For i from I down to 0, do:

1. For each $t \in [T]$, the value $Y_i^{(t)}$ is picked adversarially from $[-d_i, d_i]$, hence possibly depending on previously picked values $\{Y_j^{(t)}, \gamma_j^{(t)} : j > i, t \in [T]\}$.
2. For each $t \in [T]$, $\gamma_i^{(t)}$ is picked *uniformly* from $\{-1, +1\}$, and moreover, *independent* of any random objects picked thus far.

Lemma 3.2.14 (Computing the m.g.f.) *Suppose the γ's and Y's are picked according to the above description. Moreover, $\nu^2 := \sum_{i \in [I]} d_i^2$. Then for $0 \leq h\nu^2 < 1/2$,*

$$E[\exp(hS)] \leq (1 - 2h\nu^2)^{-T/2}.$$

Moreover, for $\epsilon > 0$, $Pr[S > (1 + \epsilon)T\nu^2] \leq ((1 + \epsilon)\exp(-\epsilon))^{T/2}$.

The proof of Lemma 3.2.14 appears in Section 3.2.4. Using this lemma, we can bound the expansion of the embedding.

Proposition 3.2.15 (Bounding Expansion) *Suppose the target dimension T is at most $\ln n$. Then, for each pair $x, y \in V$, with probability at least $1 - \frac{1}{n^2}$, $||f(x) - f(y)||_2 \leq O(\log n) \cdot d(x, y)$.*

Proof: Let $\nu^2 := \sum_{i \in [I]} d_i^2$, and recall that $S = ||f(x) - f(y)||_2^2$. Then, from Lemma 3.2.14, we have for $\epsilon > 0$, $Pr[S > (1 + \epsilon)T\nu^2] \leq ((1 + \epsilon)\exp(-\epsilon))^{T/2}$.

Note that for $\epsilon \geq 8$, $(1 + \epsilon)\exp(-\epsilon) \leq \exp(-\epsilon/2)$. Hence, for $T \leq \ln n$, we set $\epsilon := \frac{8 \ln n}{T}$ and from Lemma 3.2.5, we have $\nu^2 = \sum_{i \in [I]} d_i^2 \leq O(\log n) \cdot d(x, y)^2$. Hence, with failure probability at most $\frac{1}{n^2}$, we have $||f(x) - f(y)||_2^2 \leq (1 + \frac{8 \ln n}{T}) \cdot T \cdot O(\log n) \cdot d(x, y)^2 \leq O(\log^2 n) \cdot d(x, y)^2$.
■

Using the union bound over all pairs (x, y) and combining with Proposition 3.2.13, we complete the proof for the low distortion embedding claimed in Theorem 3.2.4, modulo the proof of Lemma 3.2.14 that is given in Section 3.2.4. In Section 3.3, we will give an embedding that improves the dependence on the doubling dimension $\dim_D$.

3.2.4 Resolving Dependency among Random Variables

Suppose we wish to bound the magnitude of the following sum, whose terms are dependent on one another:

$$S := \sum_{t \in [T]} (Q^{(t)})^2, \tag{3.3}$$

where for each $t \in [T]$, $Q^{(t)} := \sum_{i \in [I]} \gamma_i^{(t)} Y_i^{(t)}$. The $\gamma_i^{(t)}$'s are $\{-1, +1\}$ random variables; for each $i \in [I]$, the $Y_i^{(t)}$'s are random variables taking values in the interval $[-d_i, d_i]$. The following procedure specifies how the various random variables are being sampled.

For i from I down to 0, do:

1. For each $t \in [T]$, the value $Y_i^{(t)}$ is picked from $[-d_i, d_i]$, possibly depending on previously picked values $\{Y_j^{(t)}, \gamma_j^{(t)} : j > i, t \in [T]\}$.
2. For each $t \in [T]$, $\gamma_i^{(t)}$ is picked *uniformly* from $\{-1, +1\}$, and moreover, *independent* of any random objects picked thus far.

A standard technique to analyze the magnitude of S defined in (3.3) is to consider the moment generating function (m.g.f.) $E[\exp(hS)]$, for sufficiently small $h > 0$. This is fairly easy when the terms in the summation S are independent: however, observe that each $Y^{(t)}$ is dependent on the random objects indexed by $j > i$. Moreover, the $Q^{(t)}$'s are not independent either. However, we can get around this and prove the following result, via Lemmas 3.2.16 and 3.2.17.

Lemma 3.2.14 (Computing the m.g.f.) Suppose $\nu^2 := \sum_{i \in [I]} d_i^2$. Then for $0 \leq h\nu^2 < 1/2$,

$$E[\exp(hS)] \leq (1 - 2h\nu^2)^{-T/2}.$$

Moreover, for $\epsilon > 0$, $Pr[S > (1 + \epsilon)T\nu^2] \leq ((1 + \epsilon)\exp(-\epsilon))^{T/2}$.

Recall that the problem was that each $Y^{(t)}$ is dependent on the random objects indexed by $j > i$. Moreover, the $Q^{(t)}$'s are not independent either. To get around this, we consider random variables related to $Q^{(t)}$. Define $\widehat{Q}^{(t)} := \sum_{i \in [I]} \gamma_i^{(t)} d_i$ and $\overline{Q}^{(t)} := \sum_{i \in [I]} g_i^{(t)} d_i$, where the $g_i^{(t)}$'s are independent normal $N(0, 1)$ variables. Define $\widehat{S} := \sum_{t \in [T]} (\widehat{Q}^{(t)})^2$ and $\overline{S} := \sum_{t \in [T]} (\overline{Q}^{(t)})^2$ analogously. Observe that both the $\widehat{Q}^{(t)}$'s and the $\overline{Q}^{(t)}$'s are independent over different t's. Define $\nu^2 := \sum_{i \in [I]} d_i^2$. A standard calculation gives us that $E[\exp(h\overline{S})] \leq (1 - 2h\nu^2)^{-T/2}$, for $0 \leq h\nu^2 < 1/2$. We show that $E[\exp(hS)]$ is bounded above by the same quantity.

As observed in [Ach00], by the Monotone Convergence Theorem, we have $E[\exp(hS)] = \sum_{r \geq 0} \frac{h^r}{r!} E[S^r]$. Hence, we compare the even powers of Q, $\widehat{Q}$ and $\overline{Q}$.

Lemma 3.2.16 *The following inequalities hold.*

1. *For any integer* $r \geq 0$, $E[\widehat{Q}^{2r}] \leq E[\overline{Q}^{2r}]$.
2. *For any real number* $h > 0$, $E[\exp(h\widehat{S})] \leq E[\exp(h\overline{S})]$.

Proof: The first statement follows from the observation that $E[\gamma_i^{2r}] = 1 \leq E[g_i^{2r}]$. The second statement follows from the first statement, observing that the $\widehat{Q}^{(t)}$'s and the $\overline{Q}^{(t)}$'s are independent, and using the identity $E[\exp(hZ)] = \sum_{r\geq 0} \frac{h^r}{r!} E[Z^r]$. ∎

The next lemma resolves the issue that the $Q^{(t)}$'s are not independent. The idea is to replace each random variable $Y_i^{(t)}$ by a constant d_i and show that this does not decrease the expectation of the relevant random variables.

Lemma 3.2.17 *The following properties hold.*

1. *For all* $r_t \geq 0$ *(*$t \in [T]$*),* $E[\prod_{t\in[T]} (Q^{(t)})^{2r_t}] \leq E[\prod_{t\in[T]} (\widehat{Q}^{(t)})^{2r_t}]$.
2. *For* $h > 0$, $E[\exp(hS)] \leq E[\exp(h\widehat{S})]$.

Proof: Note the second statement follows from the first using the identity $E[\exp(hZ)] = \sum_{r\geq 0} \frac{h^r}{r!} E[Z^r]$, and hence it suffices to prove the first statement. Let us define the partial sums $Q_i^{(t)} := \sum_{j\geq i} \gamma_i^{(t)} Y_i^{(t)}$ and $\widehat{Q}_i^{(t)} := \sum_{j\geq i} \gamma_i^{(t)} d_i$. We show the following statement by backward induction on i. The case $i = 1$ gives the required result. We show that for $i \in [I]$, for all $r_t \geq 0$ ($t \in [T]$), $E[\prod_{t\in[T]} (Q_i^{(t)})^{2r_t}] \leq E[\prod_{t\in[T]} (\widehat{Q}_i^{(t)})^{2r_t}]$.

The case $i = I$ follows from the fact that for all $r \geq 0$, for all $t \in [T]$, $|Y_I^{(t)}| \leq d_I$. Hence, for all $r_t \geq 0$ ($t \in [T]$), $E[\prod_{t\in[T]} (Q_I^{(t)})^{2r_t}] = E[\prod_{t\in[T]} (Y_I^{(t)})^{2r_t}] \leq E[\prod_{t\in[T]} (d_I)^{2r_t}] = E[\prod_{t\in[T]} (\widehat{Q}_I^{(t)})^{2r_t}]$.

Assume that for all $l_t \geq 0$ ($t \in [T]$), $E[\prod_{t\in[T]} (Q_{i+1}^{(t)})^{2l_t}] \leq E[\prod_{t\in[T]} (\widehat{Q}_{i+1}^{(t)})^{2l_t}]$, for $i \geq 0$. Fix some $r_t \geq 0$ ($t \in [T]$).

$$\begin{aligned}
E[\prod_{t\in[T]} (Q_i^{(t)})^{2r_t}] &= E[\prod_{t\in[T]} (Q_{i+1}^{(t)} + \gamma_i^{(t)} Y_i^{(t)})^{2r_t}] && (3.4)\\
&= E[\sum_{l_1=0}^{r_1} \cdots \sum_{l_t=0}^{r_t} \prod_{t\in[T]} \binom{2r_t}{2l_t} (Q_{i+1}^{(t)})^{2r_t-2l_t} (Y_i^{(t)})^{2l_t}] && (3.5)\\
&\leq E[\sum_{l_1=0}^{r_1} \cdots \sum_{l_t=0}^{r_t} \prod_{t\in[T]} \binom{2r_t}{2l_t} (Q_{i+1}^{(t)})^{2r_t-2l_t} d_i^{2l_t}] && (3.6)\\
&\leq E[\sum_{l_1=0}^{r_1} \cdots \sum_{l_t=0}^{r_t} \prod_{t\in[T]} \binom{2r_t}{2l_t} (\widehat{Q}_{i+1}^{(t)})^{2r_t-2l_t} d_i^{2l_t}] && (3.7)\\
&= E[\prod_{t\in[T]} (\widehat{Q}_i^{(t)})^{2r_t}] && (3.8)
\end{aligned}$$

The equality (3.5) uses the fact that the r.v.'s $\gamma_i^{(t)}$'s are independent of all other random variables and the expectation of an odd power of $\gamma_i^{(t)}$ is 0. The inequality (3.6) follows from the fact that

$|Y_i^{(t)}| \leq d_i$. The inequality (3.7) follows from the linearity of expectation and the induction hypothesis. Finally, equality (3.8) holds for the same reason as that for (3.5). This completes the inductive proof. ■

Finally, we are in a position to prove Lemma 3.2.14:

Proof of Lemma 3.2.14: From Lemma 3.2.17, we have $E[\exp(hS)] \leq E[\exp(h\widehat{S})]$, which is at most $E[\exp(h\overline{S})]$, by Lemma 3.2.16. Finally, from a standard calculation [DG03], $E[\exp(h\overline{S})] \leq (1-2h\nu^2)^{-T/2}$, for $0 \leq h\nu^2 < 1/2$.

To prove the second part of the lemma, let $h\nu^2 = \frac{\epsilon}{2(1+\epsilon)} < \frac{1}{2}$. Then, we have

$$\begin{aligned} Pr[S > (1+\epsilon)T\nu^2] &= Pr[\exp(hS) > \exp((1+\epsilon)Th\nu^2)] \\ &\leq E[\exp(hS)]\exp(-(1+\epsilon)Th\nu^2) \\ &\leq (1-2h\nu^2)^{-T/2} \cdot \exp((1+\epsilon)Th\nu^2) \\ &= ((1+\epsilon)\exp(-\epsilon))^{T/2}. \end{aligned}$$

which proves the large-deviation inequality. ■

3.3 A Better Embedding via Uniform Padded Decompositions

Our basic embedding in the previous section uses a simple padded decomposition [CGMZ05], and serves to illustrate the proof techniques: however, its dependence on $\dim_D$ is sub-optimal. In order to improve the dependence of the distortion on the doubling dimension, we use a more sophisticated decomposition scheme. We modify the uniform padded decomposition in [ABN06], by incorporating the properties of bounded doubling dimension directly within the construction, to achieve both the padding property, as well as independence between distant regions.

3.3.1 Uniform Padded Decompositions

Definition 3.3.1 (Uniform Functions) *Given a partition P of (V,d), a function $\eta : V \to \mathbb{R}$ is* uniform with respect to the partition P *if points in the same cluster take the same value under η, i.e., if $P(x) = P(y)$, then $\eta(x) = \eta(y)$.*

For $r > 0$ and $\gamma > 1$, the "local growth rate" is denoted by $\rho(x,r,\gamma) := \frac{|B(x,r\gamma|)}{|B(x,r/\gamma)|}$, and $\overline{\rho}(x,r,\gamma) := \min_{z\in B(x,r)} \rho(z,r,\gamma)$. All logarithms are based 2 unless otherwise specified.

Claim 3.3.2 (Claim 2 of [ABN06]) *For $x,y \in V$, $\gamma \geq 5$ and $r > 0$ such that $2(1+\frac{1}{\gamma})r < d(x,y) \leq (\gamma - 2 - \frac{1}{\gamma})r$, we have $\max\{\overline{\rho}(x,r,\gamma), \overline{\rho}(y,r,\gamma)\} \geq 2$.*

We show that if (V,d) has bounded doubling dimension, there exists a uniformly padded decomposition: i.e., one where the padding function $\alpha(\cdot)$ is uniform with respect to the partition. The following lemma is similar to [ABN06, Lemma 4], except that it has additional properties about bounded doubling dimension, and also independence between distant regions.

Lemma 3.3.3 (Uniform Padded Decomposition) *Suppose (V,d) is a metric space with doubling dimension k, and $D > 0$. Let $\Gamma \geq 8$. Then, there exists a D-bounded α-padded decomposition Π on (V,d), where $\alpha = O(k)$, with the following properties. For each partition P in the support of Π, there exist uniform functions $\xi_P : V \to \{0,1\}$ and $\eta_P : V \to (0,1)$ such that $\eta_P \geq \frac{1}{\alpha}$. Moreover, if $\xi_P(x) = 1$, then $2^{-7}/\log \rho(x,D,\Gamma) \leq \eta_P(x) \leq 2^{-7}$; if $\xi_P(x) = 0$, then $\eta_P(x) = 2^{-7}$ and $\overline{\rho}(x,D,\Gamma) < 2$.*

Then, for all $x \in V$, the probability of the event $\{B(x, \eta_P(x)D) \subseteq P(x)\}$ is at least $\frac{1}{2}$. Furthermore, the event $\{B(x, \eta_P(x)D) \subseteq P(x)\}$ is independent of all the events $\{B(z, \eta_P(z)D) \subseteq P(z) : z \notin B(x, 3D/2)\}$.

Proof: We first describe how a random decomposition is sampled, and show that it satisfies the claimed properties. We construct a $\frac{D}{4}$-net N for (V,d) in the following way. Initially, no net points are chosen and all points are uncovered. While there are still uncovered points in V, we pick v among the uncovered points that minimizes $\rho(v,D,\Gamma)$. We include v in the set N of net points, and all points in V within distance $\frac{D}{4}$ of v are covered. The process is repeated until all points are covered. Let $N := \{v_1, v_2, \ldots, v_{|N|}\}$ be the net points in the order in which they are picked. Let λ be the maximum number of net points in N in a ball of radius $\frac{3D}{4}$. Since (V,d) has doubling dimension k, $\lambda = 2^{O(k)}$. Without loss of generality, we assume $\lambda \geq 8$.

We next describe how each cluster is formed in a random partition. Initially, all points in V are unclustered. We start from $j = 1$ to $|N|$, and form a cluster C_j (which can be empty) in the following manner. For each j, define

$$\widehat{\chi}_j := \rho(v_j, D, \Gamma), \tag{3.9}$$

$$\chi_j := 2\min\{\max\{\widehat{\chi}_j, \sqrt{8}\}, \lambda\}. \tag{3.10}$$

and the probability density function

$$p(r) := \frac{\chi_j^2}{1-\chi_j^{-2}} \cdot \frac{8\ln \chi_j}{D} \cdot \chi_j^{\frac{-8r}{D}} \qquad \text{for } r \in [\tfrac{D}{4}, \tfrac{D}{2}] \tag{3.11}$$

We sample a random radius r_j from the above probability density function. The cluster C_j consists of the remaining unclustered points in $B(v_j, r_j)$, which can be empty. The non-empty clusters form the random partition.

Next, we define two functions $\xi : V \to \{0,1\}$ and $\eta : V \to (0,1)$. Suppose the cluster C_j is non-empty. For all $x \in C_j$, define

$$\eta_P(x) := \frac{2^{-7}}{\min\{\max\{\log \widehat{\chi}_j, 1\}, \log \lambda\}} \geq \frac{1}{\alpha},$$

for some $\alpha = O(k)$. If $\widehat{\chi}_j \geq 2$, define $\xi_P(x) = 1$; otherwise, $\xi_P(x) := 0$. Hence, by construction, the functions ξ_P and η_P are uniform with respect to the partition P.

Relationship between ξ_P and η_P. Suppose $\xi_P(x) = 1$. Then it follows that $\widehat{\chi}_j \geq 2$. From the way the net N is constructed, observe that when the random radius r_j is picked, all remaining unclustered points z satisfy $\rho(z, D, \Gamma) \geq \rho(v_j, D, \Gamma) = \widehat{\chi}_j$. Hence, it follows that $2^{-7} \geq \eta_P(x) \geq \frac{2^{-7}}{\log \widehat{\chi}_j} \geq \frac{2^{-7}}{\log \rho(x,D,\Gamma)}$.

Suppose $\xi_P(x) = 0$. Then, $\widehat{\chi}_j < 2$, and hence $\eta_P(x) = 2^{-7}$. Moreover, since $d(x, v_j) \leq \frac{D}{2}$, it follows that $\overline{\rho}(x, D, \Gamma) \leq \rho(v_j, D, \Gamma) = \widehat{\chi}_j < 2$.

Independence between distant regions. Define $N_x := \{v \in N : d(x, v) \leq \frac{3D}{4}\}$. Observe that the event $\{B(x, \eta_P(x)D) \subseteq P(x)\}$ is determined completely by the random r_j's for which $v_j \in N_x$. Hence, this event is independent of all the events $\{B(z, \eta_P(z)D) \subseteq P(z) : z \notin B(x, \frac{3D}{2})\}$.

Padding property. Finally, it remains to show that the event $\{B(x, \eta_P(x)D) \subseteq P(x)\}$ happens with probability at least $\frac{1}{2}$. Using the same argument as the proof of Lemma 4 in [ABN06], the probability of the event $\{B(x, \eta_P(x)D) \nsubseteq P(x)\}$ is at most $(1 - \theta)(1 + \theta \sum_{v_j \in N_x} \chi_j^{-1})$, for the particular choice of $\theta = \sqrt{1/2}$. For completeness, we outline the proof of this result in Lemma 3.3.4. Hence, it suffices to show that the sum $\sum_{v_j \in N_x} \chi_j^{-1}$ is at most 1.

Recall from construction (3.10) that $\chi_j := 2 \min\{\max\{\widehat{\chi}_j, \sqrt{8}\}, \lambda\}$. Define $N_1 := \{v_j \in N_x : \chi_j = 2\max\{\widehat{\chi}_j, \sqrt{8}\}\}$, the net points influencing x whose χ_j value is attained by the first argument in the minimum. Define $N_2 := N_x \setminus N_1$ to be the rest of the net points influencing x. Note that for $v_j \in N_2$, $\chi_j = 2\lambda$. Observe that for all $v_j \in N_1$, $\chi_j^{-1} \leq \frac{1}{2} \cdot \frac{|B(v_j, D/\Gamma)|}{|B(v_j, D\Gamma|)} \leq \frac{1}{2} \cdot \frac{|B(v_j, D/\Gamma)|}{|B(x, 3D/4 + D/\Gamma|)}$; the last inequality follows from the fact that $B(v_j, D\Gamma) \supseteq B(x, 3D/4 + D/\Gamma)$. Moreover, observe that $B(v_j, D/\Gamma) \subseteq B(x, 3D/4 + D/\Gamma)$. Since N_1 are points from a $\frac{D}{4}$-net, any two points are more than $\frac{D}{4}$ apart. Finally, the balls $B(v_j, D/\Gamma)$ are disjoint, as $D/\Gamma \leq D/8$. Hence, it follows that $\sum_{v_j \in N_1} \chi_j^{-1} \leq \frac{1}{2}$.

On the other hand, $\sum_{v_j \in N_2} \chi_j^{-1} \leq |N_2|/2\lambda \leq \frac{1}{2}$, because $|N_x| \leq \lambda$. Hence, the sum $\sum_{v_j \in N_x} \chi_j^{-1} \leq 1$, as required. ∎

The following lemma is proved using techniques in Lemma 4 of [ABN06]. For completeness, we give the proof here.

Lemma 3.3.4 *Consider the decomposition Π on (V, d) described in Lemma 3.3.3, and the associated function $\eta_P : V \to (0, 1)$ for each partition P in the support of Π. Fix $x \in V$ and recall $N_x := \{x \in N : d(x, v) \leq \frac{3D}{4}\}$, the net-points used in the decomposition that are close to x. Recall also that for each $v_j \in N$, there is a parameter χ_j for sampling a random radius r_j that is used to create a cluster centering at v_j. Then, the probability of the event $\{B(x, \eta_P(x)D) \nsubseteq P(x)\}$ is at most $(1 - \theta)(1 + \theta \sum_{v_j \in N_x} \chi_j^{-1})$, where $\theta = \sqrt{1/2}$.*

Proof: We first state a property of the probability density function defined in (3.11). For convenience, for two sets A and S, we use $A \bowtie S$ to denote $A \cap S \neq \emptyset$ and $A \cap \overline{S} \neq \emptyset$.

Proposition 3.3.5 (Lemma 5 of [ABN06]) *Suppose $Z \subseteq V$ and $x, v \in Z$. Let $\chi \geq 2$ be a parameter, and $D > 0$ be an upper bound on the diameter of a cluster. Suppose r is sampled from the distribution $p(r) := \frac{\chi^2}{1-\chi^{-2}} \cdot \frac{8 \ln \chi}{D} \cdot \chi^{-8r/D}$, $r \in [D/4, D/2]$. Let $S := B_Z(v, r)$. Suppose $\theta \in (0, 1)$ such that $\theta \geq 2\chi^{-1}$, and let $\eta = \frac{1}{16} \log(1/\theta)/\log \chi$. Then, the following holds:*

$$Pr[B_Z(x, \eta D) \bowtie S] \leq (1 - \theta)[Pr[B_Z(x, \eta D) \cap S \neq \emptyset] + \theta \chi^{-1}].$$

We consider the probability that the ball $B(x, \eta_P(x)D)$ is separated by the partition P. Observe that the ball $B(x, \eta_P(x)D) \subseteq B(x, D/4)$ can only be influenced by net points in $N_x := \{v \in N : d(x, v) \leq 3D/4\}$. For convenience, we relabel the net points $N_x := \{v_1, v_2, \ldots, v_t\}$, while still preserving the relative order in which they are picked. Observe that since $\{\chi_j\}$ is monotonically increasing, $\{\eta_j\}$ is monotonically decreasing. Suppose S_j is the cluster created by using v_j as the center.

Observe there is some j_0 such that $x \in S_{j_0}$. In this case, $\eta_P(x) \leq \eta_{j_0}$. Hence, if the ball $B(x, \eta_P(x)D)$ is not contained in S_{j_0}, it must be the case that there is some $j \leq j_0$ such that $B(x, \eta_P(x)D) \bowtie S_j$. Now, since $\eta_{j_0} \leq \eta_j \leq 1/16$, it follows that $B(x, \eta_j D) \bowtie S_j$. So, it suffices to analyze the event that there exists some j such that $B(x, \eta_j D) \bowtie S_j$.

For $1 \leq m \leq t$, we define the events:

$$\mathcal{Z}_m := \{\forall j, 1 \leq j < m, B(x, \eta_j D) \cap S_j = \emptyset\},$$

$$\mathcal{E}_m = \{\exists j, m \leq j \leq t, B(x, \eta_j D) \bowtie S_j | \mathcal{Z}_m\}.$$

We wish to obtain an upper bound for $Pr[\mathcal{E}_1]$. We prove the following result by induction. The required result comes from the case $m = 1$. For $1 \leq m \leq t$,

$$Pr[\mathcal{E}_m] \leq (1 - \theta)(1 + \theta \sum_{j \geq m} \chi_j^{-1}).$$

We shall use Proposition 3.3.5 repeatedly for the case $\theta = \sqrt{1/2}$. First check that $\theta = \sqrt{1/2} \geq 2\chi_j^{-1}$, for all j. For the base case $m = t$, observe that $\mathcal{Z}_t$ implies that x must be in the cluster S_t. Hence, $Pr[B(x, \eta_t D) \cap S_t \neq \emptyset | \mathcal{Z}_t] = 1$. We apply Proposition 3.3.5 to obtain:

$$Pr[\mathcal{E}_t] \leq (1 - \theta)(1 + \theta \chi_t^{-1}).$$

Suppose the inductive result holds for the case $m+1$ and we consider the case for $m \geq 1$. Define the events:

$$\mathcal{F}_m := \{B(x, \eta_m D) \bowtie S_m | \mathcal{Z}_m\},$$

$$\mathcal{G}_m := \{B(x, \eta_m D) \cap S_m = \emptyset | \mathcal{Z}_m\} = \{\mathcal{Z}_{m+1} | \mathcal{Z}_m\}.$$

We first consider $Pr[\mathcal{F}_m]$. Using Proposition 3.3.5, we have

$$Pr[\mathcal{F}_m] \leq (1 - \theta)(Pr[\overline{\mathcal{G}_m}] + \theta \chi_m^{-1}).$$

Hence, using the induction hypothesis, we complete the inductive step:

$$\begin{array}{rcl} Pr[\mathcal{E}_m] & \leq & Pr[\mathcal{F}_m] + Pr[\mathcal{G}_m]Pr[\mathcal{E}_{m+1}] \\ & \leq & (1-\theta)(Pr[\overline{\mathcal{G}_m}] + \theta\chi_m^{-1}) + Pr[\mathcal{G}_m] \cdot (1-\theta)(1+\theta\sum_{j\geq m+1}\chi_j^{-1}) \\ & \leq & (1-\theta)(1+\theta\sum_{j\geq m}\chi_j^{-1}). \end{array}$$

■

3.3.2 The Better Embedding: Defining the Embedding

The new embedding is quite similar to the basic embedding of Section 3.2.1. We use the uniform padded decomposition of Lemma 3.3.3 to define the new embedding $f : (V, d) \to \mathbb{R}^T$. As before, the metric (V, d) has doubling dimension $\dim_D = k$, and suppose $\alpha = O(k)$ is the padding parameter in Lemma 3.3.3. Let $D_i := H^i$, and assume that the distances in (V, d) are between 2 and H^I.

Again, the embedding is of the form $f := \oplus_{t\in[T]}\Phi^{(t)}$, where each $\Phi^{(t)} : V \to \mathbb{R}$ is generated independently according to some distribution; for ease of notation, we drop the superscript t in the following. Also, each Φ is of the form $\Phi := \sum_{i\in[I]}\varphi_i$. We next describe how each $\varphi_i : V \to \mathbb{R}$ is constructed.

For each $i \in [I]$, let P_i be a random partition of (V, d) sampled from the decomposition scheme as described in Lemma 3.3.3. Suppose $\xi_{P_i} : V \to \{0, 1\}$ and $\eta_{P_i} : V \to (0, 1)$ are the associated uniform functions with respect to the partition P_i. Let $\{\sigma_i(C) : C \text{ is a cluster of } P_i\}$ be uniform $\{0, 1\}$-random variables and γ_i be a uniform $\{-1, +1\}$-random variable. The random objects P_i's, σ_i's and γ_i's are independent of one another. Then φ_i is defined by the realization of the various random objects as:

$$\varphi_i(x) := \gamma_i \cdot \sigma_i(P_i(x)) \cdot \min\left\{\xi_{P_i}(x) \cdot \eta_{P_i}(x)^{-1/2} \cdot d(x, V \setminus P_i(x)), \frac{D_i}{\sqrt{\alpha}}\right\}. \tag{3.12}$$

Note the similarities and difference with (3.1). Again, we let
$\kappa_i(x) := \sigma_i(P_i(x)) \cdot \min\{\xi_{P_i}(x)\eta_{P_i}(x)^{-1/2}d(x, V \setminus P_i(x)), \frac{D_i}{\sqrt{\alpha}}\}$ denote the right half of the expression above.

The proof bounding the distortion will proceed similarly: we show that with non-zero probability, the embedding $f : V \to \mathbb{R}^T$ has low distortion.

3.3.3 The Better Embedding: Bounding Contraction for Nearby Net Points

As before, we use the bounded growth-rate of the metric to bound the contraction of the embedding; however, the proofs are now somewhat more involved. Again, we assume that the γ's are arbitrarily fixed, and the P's and σ's are random and independent. For each $i \in [I]$, let the subset N_i be an arbitrary βD_i-net of (V, d), for some $0 < \beta < 1$ to be specified later. Note that N_i is different from the net used for obtaining the D_i-bounded decomposition P_i. As in the basic embedding, we first bound the contraction for the pairs in $E_i := \{(x, y) \in N_i \times N_i : 3D_i < d(x, y) \leq 4HD_i\}$, $i \in [I]$, and then extend it to all pairs in Section 3.3.5.

Let us fix a pair $(x, y) \in E_i$. Suppose $2(1 + 1/\Gamma) \leq 3$ and $4H \leq (\Gamma - 2 - 1/\Gamma)$: Claim 3.3.2 implies that $\max\{\overline{\rho}(x, D_i, \Gamma), \overline{\rho}(y, D_i, \Gamma)\} \geq 2$. Without loss of generality, we assume the maximum is attained by x. Lemma 3.3.3 now implies that $\xi_{P_i}(x) = 1$.

For $t \in [T]$, define $A^{(t)}(x, y)$ to be the event that all the following happens:

- $B(x, \eta_{P_i}(x)D_i) \subseteq P_i^{(t)}(x)$;
- $\sigma_i^{(t)}(P_i^{(t)}(y)) = 0$;
- if $|\sum_{j>i}(\varphi_i^{(t)}(x) - \varphi_i^{(t)}(y))| \leq \frac{D_i}{2\sqrt{\alpha}}$, then $\sigma_i^{(t)}(P_i^{(t)}(x)) = 1$, otherwise $\sigma_i^{(t)}(P_i^{(t)}(x)) = 0$.

Proposition 3.3.6 *Let $(x, y) \in E_i$. Suppose for $j > i$, the random objects $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : t \in [T]\}$ have been arbitrarily fixed. For each $t \in [T]$, sample random partition $P_i^{(t)}$ from Lemma 3.3.3 and random $\{0,1\}$-variables $\{\sigma_i^{(t)}(C) : C$ is a cluster in $P_i^{(t)}\}$ uniformly, all independently of one another. Then, for each $t \in [T]$, with probability at least $\frac{1}{8}$, the event $A^{(t)}(x, y)$ happens independently over different t's.*

Moreover, if the event $A^{(t)}(x, y)$ happens, then the inequality $|\sum_{j\geq i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \geq \frac{D_i}{2\sqrt{\alpha}}$ holds. Also, in this case, for any realization of the remaining random objects, i.e., $\gamma_i^{(t)}$ and $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : j < i\}$, the inequality $|\sum_{i\in[I]}(\varphi_i^{(t)}(x) - \varphi_i^{(t)}(y))| \geq \frac{D_i}{4\sqrt{\alpha}}$ holds, provided $H \geq 8$. (Recall $D_{i+1} = HD_i$.)

Proof: Because of the independence of $P_i^{(t)}$ and $\sigma_i^{(t)}$, and observing that x and y are separated by $P_i^{(t)}$, the event $A^{(t)}(x, y)$ happens with probability at least $1/8$. Now, suppose the event $A^{(t)}(x, y)$ happens. For ease of notation, we omit the superscript t. Then, it follows that from $B(x, \eta_{P_i}(x)D_i) \subseteq P_i(x)$ that $d(x, V \setminus P_i(x)) \geq \eta_{P_i}(x)D_i$. Recalling that $\xi_{P_i}(x) = 1$, we have $\xi_{P_i}(x)\eta_{P_i}(x)^{-1/2}d(x, V \setminus P_i(x)) \geq \eta_{P_i}(x)^{1/2}D_i \geq D_i/\sqrt{\alpha}$. Hence, irrespective of whether $\sigma_i(P_i(x))$ is 0 or 1, we have $|\sum_{j\geq i}(\varphi_j^{(t)}(x) - \varphi_j^{(t)}(y))| \geq \frac{D_i}{2\sqrt{\alpha}}$. The rest of the results follow from straight forward calculation, observing that $\{D_j\}$ forms a geometric sequence. ■

As before, we define $C(x, y)$ to be the event that for at least $\frac{T}{16}$ values of t, the event $A^{(t)}(x, y)$ happens.

Using the same Chernoff bound as in Proposition 3.2.7, we can show a similar result.

Proposition 3.3.7 *Suppose $(x, y) \in E_i$, and for $j > i$, the random objects $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : t \in [T]\}$ have been arbitrarily fixed. Then, the event $C(x, y)$ fails to happen with probability at most $p := \exp(-\frac{T}{64})$.*

We next use the Lovasz Local Lemma to show that there is some realization of $\{P_i^{(t)}, \sigma_i^{(t))} : t \in [T]\}$ such that for all $(x, y) \in E_i$, the events $C(x, y)$ happen simultaneously. In order to use the Local Lemma, we need to analyze the dependency of these events. Recall that N_i is a βD_i-net of (V, d_i).

Lemma 3.3.8 *For each $(x, y) \in E_i$, the event $C(x, y)$ is independent of all but $B := (\frac{H}{\beta})^{O(k)}$ of $C(u, v)$, where $(u, v) \in E_i$.*

Proof: Observe that the event $C(x,y)$ is determined by the random objects $\{P_i^{(t)}, \sigma_i^{(t)} : t \in [T]\}$. More specifically, it is determined completely by the events $\{B_i(w, \frac{D_i}{\alpha}) \subseteq P_i^{(t)}(w) : t \in [T]\}$ and $\{\sigma_i^{(t)}(P^{(t)}(w)) = 0 : t \in [T]\}$, for $w \in \{x, y\}$. Note that if $d_i(x,w) > 3D_i/2$, then the corresponding events for the points x and w are independent. Note that if $d_i(x,w) \leq 3D_i/2$, then $d(x,w) \leq 2D_i$; moreover, any two net-points in (V, d_i) must be more than βD_i apart in (V,d). Hence, observing that the doubling dimension of the given metric is at most k, for each of x and y, only $(\frac{2D_i}{\beta D_i})^{O(k)}$ net points are relevant. Now, each net point can be incident by at most $(\frac{4H}{\beta})^{O(k)}$ edges in E_i. Hence, it follows that $C(x,y)$ is independent of all but $(\frac{H}{\beta})^{O(k)}$ of the events $C(u,v)$, where $(u,v) \in E_i$. ∎

By the Local Lemma, if $ep(B+1) < 1$, then all the events $C(x,y)$, where $(x,y) \in E_i$ happen with positive probability.

Proposition 3.3.9 *Suppose for $j > i$, the random objects $\{\gamma_j^{(t)}, P_j^{(t)}, \sigma_j^{(t)} : t \in [T]\}$ have been arbitrarily fixed. If $T = \Omega(k \log \frac{H}{\beta})$, then there is some realization of $\{P_i^{(t)}, \sigma_i^{(t)} : t \in [T]\}$ such that all the events $\{C(x,y) : (x,y) \in E_i\}$ happen.*

Again, we define $\mathcal{E}$ to be the event that for all $i \in [I]$, for all $(x,y) \in E_i$, the event $C(x,y)$ happens. As in the basic embedding, Proposition 3.3.9 can be used repeatedly to show the following result.

Proposition 3.3.10 (Contraction for Nearby Net Points) *Suppose $T = \Omega(k \log \frac{H}{\beta})$. Moreover, the γ's are arbitrarily fixed, and the P's and σ's remain random and independent. Then, the event $\mathcal{E}$ happens with non-zero probability. In particular, there exists some realization of the P's and σ's such that the embedding $f : (V,d) \rightarrow \mathbb{R}^T$ satisfies for all $i \in [I]$, for all $(x,y) \in E_i$, $||f(x) - f(y)||_2 \geq \frac{\sqrt{T}}{4} \cdot \frac{D_i}{4\sqrt{\alpha}}$.*

3.3.4 The Better Embedding: Bounding the Expansion

We use the same argument as the basic embedding to bound the expansion. We sample the γ's uniformly and independently, and use Proposition 3.3.10 to show there exists some realization of the P's and σ's such that the resulting mapping $f : V \rightarrow \mathbb{R}^T$ has the guaranteed contraction. Hence, we can focus on analyzing the expansion.

Again, fix $x, y \in V$ and let $S := ||f(x) - f(y)||_2^2 = \sum_{t \in [T]} (Q^{(t)})^2$, where $Q^{(t)} := \Phi^{(t)}(x) - \Phi^{(t)}(y)$. In turn, $Q^{(t)} := \sum_{i \in [I]} \gamma_i^{(t)} Y_i^{(t)}$, where $Y_i^{(t)} := \kappa_i^{(t)}(x) - \kappa_i^{(t)}(y)$. Recall that $\gamma_i^{(t)}$ is uniformly picked from $\{-1, +1\}$.

We next bound the magnitude of Y_i in the following Lemma, whose proof depends on the uniformity of ξ_{P_i} and η_{P_i}. The proof follows the same argument as in [ABN06, Lemma 8], which we include here for completeness.

Lemma 3.3.11 *Consider a particular $Y_i = \kappa_i(x) - \kappa_i(y)$. Then, the following holds.*

1. *We have $|Y_i| \leq \max\{\xi_{P_i}(x)\eta_{P_i}(x)^{-1/2}, \xi_{P_i}(y)\eta_{P_i}(y)^{-1/2}\} \cdot d(x,y)$.*
2. *For all $z \in V$, $\xi_{P_i}(z)\eta_{P_i}(z)^{-1} \leq 2^7 \log \rho(z, D_i, \Gamma)$.*

Proof: We first prove the first statement. Note that it suffices to show that $\kappa_i(x) - \kappa_i(y) \leq \xi_{P_i}(x)\eta_{P_i}(x)^{-\frac{1}{2}} \cdot d(x,y)$, because by symmetry we would have $\kappa_i(x) - \kappa_i(y) \leq \xi_{P_i}(y)\eta_{P_i}(y)^{-\frac{1}{2}} \cdot d(x,y)$, which gives the required result.

Recall that $\kappa_i(x) := \sigma_i(P_i(x)) \cdot \min\{\xi_{P_i}(x)\eta_{P_i}(x)^{-\frac{1}{2}} d(x, V \setminus P_i(x)), \frac{D_i}{\sqrt{\alpha}}\}$.

We first consider the case $P_i(x) \neq P_i(y)$. Notice that in this case $d(x, V \setminus P_i(x)) \leq d(x,y)$. Hence, we have $\kappa_i(x) - \kappa_i(y) \leq \kappa_i(x) \leq \xi_{P_i}(x)\eta_{P_i}(x)^{-\frac{1}{2}} \cdot d(x, V \setminus P_i(x)) \leq \xi_{P_i}(x)\eta_{P_i}(x)^{-\frac{1}{2}} \cdot d(x,y)$.

For the case $P_i(x) = P_i(y)$, we use the uniformity of the functions ξ_{P_i} and η_{P_i}. If $\kappa_i(y) = \sigma_i(P_i(y)) \cdot \frac{D_i}{\sqrt{\alpha}}$, then since $\kappa_i(x) \leq \sigma_i(P_i(x)) \cdot \frac{D_i}{\sqrt{\alpha}}$, it follows that $\kappa_i(x) - \kappa_i(y) \leq 0$; otherwise, $\kappa_i(x) - \kappa_i(y) \leq \xi_{P_i}(x)\eta_{P_i}(x)^{-\frac{1}{2}} \cdot |d(x, V \setminus P_i(x)) - d(y, V \setminus P_i(y))| \leq \xi_{P_i}(x)\eta_{P_i}(x)^{-\frac{1}{2}} \cdot d(x,y)$.

The second statement follows from the construction of ξ_{P_i} and η_{P_i} as in Lemma 3.3.3. If $\xi_{P_i}(z) = 1$, then $\eta_{P_i}(z)^{-1} \leq 2^7 \log \rho(z, D_i, \Gamma)$. ∎

We have $|Y_i| \leq d_i := \max\{\sqrt{O(\log \rho(x, D_i, \Gamma))}, \sqrt{O(\log \rho(y, D_i, \Gamma))}\} \cdot d(x,y)$.

Denote $\nu^2 := \sum_{i \in [I]} d_i^2$. We bound the magnitude of ν in the following proposition. The first statement follows from a telescoping sum, and the second follows from the first, using the definition of d_i.

Proposition 3.3.12 *The following inequalities hold.*

1. *For all $z \in V$, $\sum_{i \in [I]} \log \rho(z, D_i, \Gamma) = O(\log_H \Gamma) \cdot \log n$.*
2. *$\nu^2 = O(\log_H \Gamma \log n) \cdot d(x,y)^2$.*

The proof now proceeds in the same fashion as in Section 3.2.3; setting $H := 16$ and $\Gamma := 128$, we have $\nu^2 = O(\log n) \cdot d(x,y)^2$. Hence, applying Lemma 3.2.14, and setting $\epsilon := \frac{8 \ln n}{T}$ as before, we have the following result.

Lemma 3.3.13 (Bounding Expansion) *Suppose $T \leq \ln n$. Then, for each pair $x, y \in V$, with probability at least $1 - \frac{1}{n^2}$,*

$$||f(x) - f(y)||_2 \leq O(\log n) \cdot d(x,y).$$

3.3.5 The Better Embedding: Bounding Contraction for All Pairs

Now that we have proved that with non-zero probability, the expansion for every pair of points is at most $O(\log n)$, and the contraction for nearby net points is bounded, we next show that if the βD_i-net N_i for (V,d) is fine enough (i.e., β is small enough), then the contraction bound can be extended to *all pairs*.

Lemma 3.3.14 (Bounding Contraction for All Pairs) *Suppose the event $\mathcal{E}$ holds and the expansion of the embedding f is bounded in the manner described in Lemma 3.3.13. Suppose $\beta > 0$ is small enough such that $\beta^{-1} = \Theta(\sqrt{\alpha} \log n)$, where $\alpha = O(k)$. Then, for all $x, y \in V$, $||f(x) - f(y)||_2 \geq \Omega(\sqrt{T/\alpha}) \cdot d(x,y)$.*

Proof: Without loss of generality, we can assume $\beta < \frac{1}{4}$. Suppose $x, y \in V$. Let $i \in [I]$ such that $(3+2\beta)D_i < d(x,y) \leq (3+2\beta)HD_i$. Suppose $u, v \in N_i$ are net points closest to x and y respectively. Then, it follows that $3D_i < d(u,v) \leq 4HD_i$, and so $(u,v) \in E_i$. Hence, by Theorem 3.3.10, $||f(u) - f(v)||_2 \geq \frac{\sqrt{T}}{4} \cdot \frac{D_i}{4\sqrt{\alpha}}$. On the other hand, both $d(u,x)$ and $d(v,y)$ are at most βD_i. Since the expansion of the embedding f is bounded by $O(\log n)$, it follows that both $||f(u) - f(x)||_2$ and $||f(v) - f(y)||_2$ are at most $O(\log n) \cdot \beta D_i$.

Finally, we set β to be small enough such that $\frac{1}{\beta} = \Theta(\sqrt{\alpha} \log n)$. By the triangle inequality, $||f(x) - f(y)||_2 \geq ||f(u) - f(v)||_2 - ||f(x) - f(u)||_2 - ||f(y) - f(v)||_2 \geq \Omega(\sqrt{\frac{T}{\alpha}}) \cdot D_i \geq \Omega(\sqrt{\frac{T}{\alpha}}) \cdot d(x,y)$. ■

Putting Lemmas 3.3.13 and 3.3.14 together proves Theorem 3.1.2.

Chapter 4

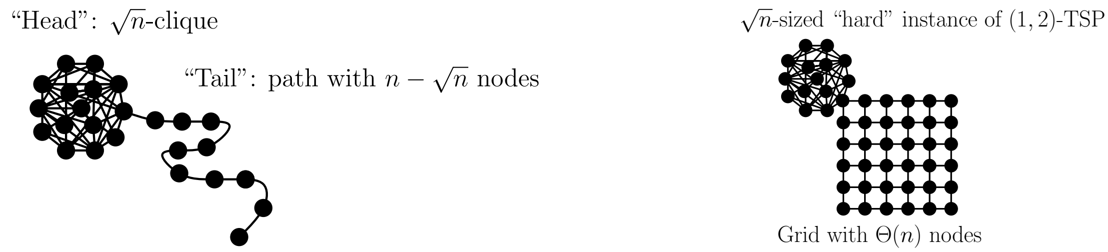
Approximating TSP on Metrics with Bounded Global Growth

4.1 Introduction

Distance functions are ubiquitous, arising as distances from home to work, round-trip delays between hosts on the Internet, dissimilarity measures between documents, and many other applications. As a simplifying assumption, theoreticians often assume that the distance function in question forms a metric. 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 and satisfy the triangle inequality. Unless specified otherwise, we assume that the set V is finite.

However, some problems remain hard even when the underlying distance function is a metric, an example of which is the Traveling Salesman Problem (TSP). Papadimitriou and Yannakakis [PY93] showed that TSP is MAX-SNP hard in general for metrics whose distances are either 1 or 2. Indeed, even for more structured metrics such as Euclidean metrics, Trevisan [Tre00] showed that the problem remains MAX-SNP hard if the Euclidean dimension is unbounded. On the other hand, Arora [Aro98] gave the first PTAS for TSP on low dimensional Euclidean metrics. A natural and basic question that arises in the study of metric spaces is: *How do we quantify the complexity of metric spaces?* More specifically, *which classes of metric spaces admit efficient algorithms for TSP?* The class of tree metrics trivially admits efficient exact solution for TSP. It is not surprising that metrics induced by special classes of graphs admit efficient TSP algorithms. For instance, for graphs with bounded tree widths, Arnborg and Proskurowski [AP89] gave a dynamic program that solves TSP on the induced metrics exactly in linear time. For metrics induced by weighted planar graphs, the best known algorithm is by Klein [Kle05], who gave a $(1 + \epsilon)$-approximation algorithm that runs in linear time $O(c^{1/\epsilon^2} n)$, where $c > 0$ is some constant. Grigni [Gri00] gave QPTAS's for metrics induced by minor-forbidding graphs and bounded-genus graphs.

The above examples were situations where the simplicity was in the representation: one can ask if there are some parameters that capture the complexity of metric spaces. For Euclidean metrics, the underlying dimension is such a good candidate. However, not all metrics are Euclidean, and



(a) Lollipop (b) Augmented Grid

Figure 4.1: Very simple examples of metrics with low correlation dimension.

a general metric embeds into L_2 with distortion as large as $\Omega(\log n)$ [Mat97], even with no restriction on the number of dimensions. A question one can ask is: are there other parameters that can capture the intrinsic algorithmic complexity of an abstract metric space (i.e., independent of its representation)? What is the *intrinsic dimension* of $M = (V, d)$?

Building on a definition of [Ass83], researchers considered the *doubling dimension* $\dim_D(M)$ of a metric M [GKL03]: this concept generalized the notion of dimension in geometric spaces, i.e., $\dim_D(\mathbb{R}^d, \ell_p) = \Theta(d)$. Doubling dimension proved to be a very useful parameter: in the past three years, many algorithms have since been developed whose performance (run-time, space) can be given by functions $F(|V|, \dim_D(M))$, which give better quantification than those obtained for general metrics. For instance, Talwar [Tal04] gave a $(1+\epsilon)$-approximation algorithm for TSP such that for metrics with doubling dimension $\dim_D(M)$ at most k, the algorithm runs in time $2^{(\frac{k}{\varepsilon}\log n)^{O(k)}}$. While this result is potentially worse for large dimensions, it is *much* better for well-behaved metrics, and arguably having this extra parameter to work with allows us to develop more nuanced algorithms.

Despite its popularity, doubling dimension has some drawbacks: perhaps the biggest one is being that a space with low $\dim_D$ cannot have "large dense clusters".[1] This strict definition makes it difficult to use it to model real networks, which tend to be well-behaved "on the average", but often have a few regions of "high density". We define a new notion of dimension, the **correlation dimension** which captures the idea of being *"low-dimensional on average"*. We give structural results as well as algorithms for spanners and TSP for metrics with low correlation dimension. Our definitions are inspired by work on the *correlation fractal dimension* in physics [GP83] and in databases [BF95].

Note that correlation dimension is not the only (or even the first) idea to incorporate dense regions in graphs (see [KL06] for another exciting, and somewhat different direction of relaxing doubling dimension, thereby obtaining both PTAS and QPTAS for TSP). But it gives a different (global) way of measuring the complexity, and can be useful in contexts where stricter, local ways of measuring dimension are not applicable.

Our Results and Techniques. Given a finite metric $M = (V, d)$, let $\mathbf{B}(x, r)$ denote the ball

[1]More precisely, the doubling dimension is defined so that any set that is almost equilateral in a metric of dimension $\dim_D$ can only have $2^{\dim_D}$ points in it; the precise definition of doubling appears in Section 4.2.

around u of radius r. The *correlation dimension* is defined as the smallest constant k such that

$$\sum_{x\in V}|\mathbf{B}(x,2r)| \leq 2^k \cdot \sum_{x\in V}|\mathbf{B}(x,r)|, \tag{4.1}$$

and moreover, this inequality must hold under taking any net of the metric M. (A more formal definition is given in Section 4.2.) Note that this definition is an "average" version of the bounded-growth rate used by [PRR99, KR02], and hence should be more general than that notion. We show that in fact, correlation dimension is *even more general than doubling dimension*:

Theorem 4.1.1 (Correlation Generalizes Doubling) *Given a metric M, the correlation dimension is bounded above by a constant times the doubling dimension.*

Moreover, correlation dimension is strictly more general than doubling dimension: adding a clique of size $O(\sqrt{n})$ to a doubling metric does not change its correlation dimension by much, but completely destroys its doubling dimension. (Some examples are given in Figure 4.1. One can be convinced that each of these example metrics has "low complexity on average," which is precisely what correlation dimension tries to capture.)

The following theorems show the algorithmic potential of this definition.

Theorem 4.1.2 (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 $\tilde{O}_{k,\epsilon}(\sqrt{n})$ and distortion $1+\epsilon$.*

This immediately allows us to get $2^{\tilde{O}(\sqrt{n}))}$-time algorithms for all problems that can be solved efficiently on small-treewidth graphs, including the traveling salesman problem. Moreover, Theorem 4.1.2 is tight, since metrics with bounded $\dim_C$ can contain $O(\sqrt{n})$-sized cliques.

However, 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 4.1.3 (Approximation Schemes for TSP) *Given any metric M with $\dim_C(M) = k$, the TSP can be solved to within an expected $(1+\varepsilon)$-factor in time $2^{O(n^\delta \varepsilon^{-k})}$ for* any constant $\delta > 0$.

Hence, given constants ε, k, the algorithm runs in *sub-exponential* time. (Recall that sub-exponential time is $\cap_{\delta>0}\mathsf{DTIME}(2^{n^\delta})$.) As we will see later, the best exponent in the expression above that we can show is $(\varepsilon^{-1}2^{\sqrt{\log n \log\log n}})^{4k}$.

While metrics with bounded correlation dimension cannot in general have $(1+\varepsilon)$-stretch spanners with a linear number of edges, we can indeed get some improvement over general metrics.

Theorem 4.1.4 (Sparse Spanners) *Given any $0 < \varepsilon < 1$, any metric with correlation dimension k has a spanner with $O(n^{3/2}\varepsilon^{-O(k)})$ edges and stretch $(1+\varepsilon)$. Moreover, there exist metrics with $\dim_C = 2$ and for each of which any 1.5-stretch spanner has $\Omega(n^{3/2})$ edges.*

4.1.1 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 *low-growth rate* of a metric space requires that for all $x \in V$ and all r, $|\mathbf{B}(x, 2r)|$ is comparable to $|\mathbf{B}(x, r)|$. This was used in [PRR99, KR02, HKRZ02] to develop algorithms for object location in general metrics, and in [KK77, AM05], for routing problems.

A large number of algorithms have been developed for doubling metrics; e.g., for NN-searching [Cla99, KL04, KL05, BKL06, HPM05, CG06b], for the TSP and other optimization problems [Tal04], for low-stretch compact routing [Tal04, CGMZ05, Sli05, AGGM06, XKR06, KRX06], for sparse spanners [CGMZ05, HPM05], and for other applications [KSW04, KMW05]. Many algorithms for Euclidean space have been extended to work for doubling metrics.

For Euclidean metrics, the first approximation schemes for TSP and other problems were given by Arora [Aro98] and Mitchell [Mit99]; see, e.g., [CL98, ARR99, CLZ02, KR99] for subsequent algorithms, and [CL00] for a derandomization. The runtime of Arora's algorithm [Aro98] was $O(n(\log n)^{O(\sqrt{k}\cdot\frac{1}{\varepsilon})^{k-1}})$, which was improved to $2^{(\frac{k}{\varepsilon})^{O(k)}}n + O(kn\log n)$ [RS99]. For $(1+\varepsilon)$-approximation for TSP on doubling metrics, the best known running time is $2^{(\frac{k}{\varepsilon}\log n)^{O(k)}}$ [Tal04]. Here, the parameter k is the doubling dimension or the Euclidean dimension in the corresponding cases.

Finally, the concept of correlation fractal dimension was studied by Belussi and Faloutsos [BF95, PKF00] for estimating the selectivity of spatial queries; Faloutsos and Kamel [FK94] also used fractal dimension to analyze R-trees.

Earlier Notions of Correlation Dimension 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 $\lim_{r\to 0}\lim_{\epsilon\to 0}\frac{\log C((1+\epsilon)r) - \log C(r)}{\log(1+\epsilon)}$. 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. In the next section, we will define a version of this definition for finite sets.

4.2 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 $d(x, y) \le \varepsilon$. A subset $N \subseteq V$ is an *ε-packing* if for all $x, y \in N$ such that $x \neq y$, $d(x, y) > \varepsilon$. 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 4.1.1, we give the following definition:

Definition 4.2.1 (correlation dimension) *A metric $M = (V, d)$ has* correlation dimension $\dim_C(M)$ *at most k if for all $r > 0$, the inequality*

$$\textstyle\sum_{x \in N} |\mathbf{B}_N(x, 2r)| \le 2^k \cdot \sum_{x \in N} |\mathbf{B}_N(x, r)| \tag{4.2}$$

holds for all nets $N \subseteq V$.

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. Recall that the *doubling dimension* $\dim_D(M)$ is the least k such that every ball $\mathbf{B}(x, r)$ of radius r can be covered by at most 2^k balls of radius $r/2$ [GKL03]. The *strong doubling dimension*[2] is the least k such that

$$|\mathbf{B}(x, 2r)| \le 2^k |\mathbf{B}(x, r)| \tag{4.3}$$

for all $x \in V$ and radius r. We know that the strong doubling dimension is no more than $4 \dim_D$ [GKL03]. It follows directly from the definition (4.3) that the correlation dimension is no more than the strong doubling dimension; more surprisingly, the following result is true as well. We give its proof in Section 4.3.

Theorem 4.2.2 *For any metric space M,* $\dim_C(M) \le O(\dim_D(M))$.

Hence the *class of bounded correlation dimension metrics contains the class of doubling metrics*. The converse is not true: metrics with bounded $\dim_C$ can be much richer. Consider, for instance, the unweighted 2-d grid with $\dim_D = \dim_C = O(1)$. Now attaching an unweighted clique (or, say, a metric with all distances between 1 and 2) on $O(\sqrt{n})$ vertices to one of the vertices of the grid: one can verify that the induced metric still has $\dim_C = O(1)$, but the $\dim_D$ jumps to $\frac{1}{2} \log n$.

The reader wondering about why the bounded average growth property (4.2) is required to hold for *every net* of M in Definition 4.2.1 is referred to Section 4.3.2: loosely, the definition becomes too inclusive without this restriction.

A very useful property of correlation dimension is that it still has "small" nets. (Of course, since we allow large cliques, they cannot be as small as for doubling dimension):

Lemma 4.2.3 (Small Nets) *Consider a metric $M = (V, d)$ with $\dim_C(M) \le k$. Suppose S is an R-packing with diameter D. If we add more points to S and obtain an R-net N for (V, d), then the size of the packing satisfies $|S| \le (2D/R)^{k/2} \cdot \sqrt{|N|}$.*

[2]This quantity has been described as the KR-dimension in [GKL03]; we use this name due to [BKL06] to keep matters simple.

Proof: Observe that $|S|^2 \leq \sum_{x \in N} |\mathbf{B}_N(x, D)|$. By applying the definition of correlation dimension repeatedly, we have for each integer $t \geq 0$,

$$\textstyle\sum_{x \in N} |\mathbf{B}_N(x, D)| \leq 2^{kt} \sum_{x \in N} |\mathbf{B}_N(x, D/2^t)|. \tag{4.4}$$

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

Hence, given any metric with $\dim_C = O(1)$, any near-uniform set in the metric has size at most $O(\sqrt{n})$, and hence λ, the doubling constant [GKL03] of this metric is also $O(\sqrt{n})$.

At this point, it is worthwhile to mention that because property (4.2) is required to hold for *every net* of M in Definition 4.2.1, it is hard to approximate the correlation dimension of a given metric.

Theorem 4.2.4 *Given a metric $M = (V, d)$ with n points, it is NP-hard to distinguish between the cases $\dim_C(M) = O(1)$ and $\dim_C(M) = \Omega(\log n)$.*

The proof of Theorem 4.2.4 involves a reduction from the MAXIMUM INDEPENDENT SET [Has96] problem, and is given in Section 4.4. Observe that this result rules out any non-trivial approximation of the correlation dimension; however, this does not necessarily rule out using correlation dimension for the design of algorithms. In particular, the algorithms we design do not require us to know the correlation dimension of the input metric up-front; while the TSP approximation algorithm of Section 4.7 seems to require this information at first glance, this issue can be resolved using standard "guess-and-double" ideas.

4.3 Relating Doubling and Correlation Dimensions

In this section, we study the inter-relationships between doubling dimension and correlation dimension. We show that the correlation dimension of any metric is at most $O(\dim_D(M))$, but that the converse is not true.

4.3.1 Correlation Dimension Generalizes Doubling

Let us prove the following theorem.

Theorem 4.3.1 (Doubling metrics have bounded $\dim_C$) *Let $M = (V, d)$ be a metric space. Then, $\dim_C(M) \leq 8 \dim_D(M) + 1$.*

Proof: While the proof of this theorem is somewhat long, it is conceptually not very difficult. Suppose the doubling dimension $\dim_D(M) = k$ and $\lambda = 2^k$; to prove the theorem it is enough to show that

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

We can then apply this result to every net $N \subseteq V$ (since $\dim_D(N) \leq 2 \dim_D(V)$) to complete the proof of the theorem.

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

$$\mathbf{B}(x, 2r) \subseteq \cup_{y \in Y_x} B_y, \tag{4.5}$$

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

However, we may be unlucky and have several $y \in Y_x$ such that $|B_y| > |\mathbf{B}(x,r)|$. Define the *small* centers $S_x = \{y \in Y_x \mid |\mathbf{B}(y, \frac{r}{2})| \leq |\mathbf{B}(x,r)|\}$, and the set of the *large* centers $L_x := Y_x \setminus S_x$. Note that $|S_x|, |L_x| \leq |Y_x| \leq \lambda^4$. Plugging into (4.5), we get

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

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

$$\textstyle\sum_{x \in V} |\mathbf{B}(x,2r)| \leq \lambda^4 \eta \sum_{x \in V} |\mathbf{B}(x,r)| + \sum_{x \in V} \sum_{y \in L_x} |B_y| \tag{4.6}$$

The first term is what we want: we just need to bound the second term on the right hand side of (4.6). Call this term $\mathbf{E}$.

Changing the order of summation, and defining $N_y := \{x \in V : y \in L_x\}$, we have

$$\mathbf{E} := \textstyle\sum_{x \in V} \sum_{y \in L_x} |B_y| = \sum_{y \in Y} \sum_{x: y \in L_x} |B_y| = \sum_{y \in Y} |N_y| \cdot |B_y|. \tag{4.7}$$

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

A change in perspective. 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 $|N'_y| \leq \lambda^3$. 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 $|N_y| \leq \lambda^3 |B_y|$. Plugging this into (4.7), we get

$$\mathbf{E} \leq \textstyle\sum_{y \in Y} \lambda^3 |B_y|^2. \tag{4.8}$$

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 $|B_y|^2 \leq \sum_{z \in B_y} |\mathbf{B}(z,r)|$. This implies that

$$\mathbf{E} \leq \lambda^3 \textstyle\sum_{y \in Y} \sum_{z \in B_y} |\mathbf{B}(z,r)| = \lambda^3 \sum_{z \in V} \sum_{y: z \in B_y} |\mathbf{B}(z,r)|. \tag{4.9}$$

The second equality is a change in the order of summation. We still have to show this quantity is at most $\lambda^4 \sum_{x \in V} |\mathbf{B}(x,r)|$; for this it suffices to show that $|\{y \in Y \mid z \in B_y\}| \leq \lambda$.

The Home Stretch. Consider $M_z := \{y \in Y \mid z \in B_y\}$: we want to show $|M_z| \leq \lambda$. Note that M_z is contained in a ball of radius $\frac{r}{2}$ centered at z and any two distinct points in M_z is more than $\frac{r}{2}$ apart. From the doubling property of V, M_z contains at most λ points! Combining this with (4.6) and (4.9), we have

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

completing the proof. ∎

4.3.2 The Converse is False

Given that the correlation dimension of a metric is at most $4\dim_D(M) + 1$, one can ask if the two quantities are essentially the same; however, the converse of Theorem 4.3.1 is not true. In particular, we can show that a metric with bounded correlation dimension does not necessarily have bounded doubling dimension. Consider the "$\sqrt{n}$-lollipop" metric induced by the graph obtained by attaching a path metric with $n - \sqrt{n}$ nodes to a clique of size $\sqrt{n}$: the doubling dimension of this metric is clearly at least $\log_2 \sqrt{n} = \frac{1}{2}\log n$. However, note that the quantity $\sum_x |\mathbf{B}(x, r)|$ starts off at n (for $r = 0$), and is about $\Theta(nr)$ for arbitrary $r \leq n$. Moreover, this also holds true for any ε-net N, with $\sum_{x \in N} |\mathbf{B}_N(x, r)|$ being $|N|$ for $r \leq \varepsilon$, and being $\Theta(|N|r/\varepsilon)$ for general $r \geq \varepsilon$. Hence the correlation dimension of this metric is $O(1)$.

Why require closure under taking nets?

Let us consider defining a metric to have correlation dimension k if

$$\textstyle\sum_{x\in V}|\mathbf{B}(x, 2r)| \leq 2^k \cdot \sum_{x\in V}|\mathbf{B}(x, r)| \tag{4.10}$$

holds *only for the original metric* and not for all nets N. In this case, we can show that the definition is too inclusive: in particular,

Proposition 4.3.2 *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 dimension of M' is 2 (under this new notion of dimension).*

Hence, if we do not require the closure under taking sub-nets, we can realize *any metric* as a submetric of a (slightly larger) low-dimensional metric, making the definition completely uninteresting (at least for TSP).

Proof: Without loss of generality, let the minimum inter-point 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 attach it to some point in V. If we view the original metric as a complete graph on V, the distances metric d' are the shortest-path distances in the new graph formed by adding this "tail". It is an easy calculation to check that the resulting metric has small correlation dimension (under this definition of correlation dimension). ■

This shows that the weaker definition (without the closure under taking subnets) has limited application, and motivates why we need to restrict the definition further. Taking subnets is perhaps the minimal restriction we can add to make it possible to say interesting things about the resulting metrics.

4.4 Hardness of Approximating Correlation Dimension

In this section, we show that it is NP-hard to approximate the correlation dimension of a metric better than $O(\log n)$; since the correlation dimension always lies in the interval $1, \ldots, \log n$, this proves that only trivial approximation guarantees are possible unless $P = NP$.

Theorem 4.4.1 *Given a metric $M = (V, d)$ with n points, it is NP-hard to distinguish between the cases $\dim_C(M) = O(1)$ and $\dim_C(M) = \Omega(\log n)$.*

The proof is by reduction from the hardness of approximation of INDEPENDENT SET [Has96].

Proposition 4.4.2 ([Has96]) *There exists $0 < k_1 < k_2 < 1$ such that given a graph with n vertices, it is NP-hard to distinguish whether the size of a maximum independent set is smaller than n^{k_1} or larger than n^{k_2}.*

Proof: Let $G = (V, E)$ be an instance of the independent set problem, 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 $\alpha(G) = \Omega(\log n)$.

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 useful for the hardness proof.

Let us define a parameter $l = 2(1 - k_1)$, where k_1 is the smaller constant in the hardness result for independent set, and let $K = n^l$; note that $1 < K \le n^2$; this will be the size parameter. Define $R = 2n^2$; this will 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 one other.

1. **Super-clique.** 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 and unit inter-point distance. The distance between points from adjacent clusters is 2, and hence between points in the i^{th} and j^{th} 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.

The Analysis. We now begin to examine the correlation dimension of this metric M. Note that bounding the correlation dimension amounts to 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 $1 \le \epsilon < 2$.

Lemma 4.4.3 *If N is an ϵ-net for the metric M where $\epsilon < 1$ or $\epsilon \ge 2$, then $\sum_{x \in N} |B_N(x, 2r)| \le O(1) \sum_{x \in N} |B_N(x, r)|$ for any $r > 0$.*

Proof: Let us consider ϵ-nets for $\epsilon < 1$. Since the smallest distance in M is 1, by the covering property of a net, 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 Super-clique, or within each cluster in the Chain-of-clusters can see one another. This gives a contribution of $2K \times \binom{n}{2} = \Theta(Kn^2)$ 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 gradually 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 super-clique does not also cause any sudden jumps in $F_N(r)$. Finally, as r increases beyond R, nothing interesting happens.

For $\epsilon \geq 2$, at most one point in each copy of M_G and each cluster remain in the net N. It is easy to check 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 of the chain-of-clusters; moreover, for each copy of M_G in the super-clique, 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 only give a gradual contribution of $\Theta(K^2) = o(n^2K)$; hence, if there is a large contribution to $F_N(r)$ due to the Super-clique 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 Super-clique (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 two following lemmas make this intuition formal.

Lemma 4.4.4 *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 corresponds to an independent set in G, the contribution to $F_N(2r)$ due to the Super-clique 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 4.4.5 *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 $F_N(r) \leq O(n^2K)$. Observe that the super-clique contributes at least $(n^{k_2}K)^2 = n^{2k_2+2l}$. Hence, $F_N(2r)/F_N(r) \geq \Omega(n^{2k_2+2l-2-l}) = \Omega(n^{2(k_2-k_2)})$. ■

Combining the lemmas completes the proof of the hardness reduction. ■

4.5 Sparse Spanners

We begin our study of metrics with small correlation dimension with a simple construction of sparse spanners; this will also serve to introduce the reader to some of the basic concepts we will use later. 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 $\Delta = \frac{\max_{x,y} d(x,y)}{\min_{x,y} d(x,y)}$ is the *aspect ratio* of the metric. This should be contrasted with a trivial lower bound for general metrics: any spanner with stretch less than 3 for $K_{n,n}$ requires $\Omega(n^2)$ edges.

4.5.1 Sparse Spanners: Upper Bound

Theorem 4.5.1 (Sparse Spanner Theorem) *Given a metric $M = (V, d)$ with $\dim_C(M) \leq k$, and $\varepsilon > 0$, there exists a $(1+\varepsilon)$-spanner with $\varepsilon^{-O(k)} \min\{n^{1.5}, n \log \Delta\}$ edges.*

The algorithm for constructing sparse spanners for metrics with bounded correlation dimension is the same as that for doubling metrics in Section 2.2; the proofs, of course, are different. For completeness, we briefly describe the algorithm here again.

Construction for sparse spanners. Given a metric (V, d) and a parameter $\varepsilon > 0$, let us 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 inter-vertex distance is at least 1, $Y_i = V$ for $-p \leq i < 0$.) For each net Y_i in the sequence, we add edges between vertices which are in the net Y_i and "close together". In particular, for $i \geq -p$, define the edges at level i to be $E_i = \{(u, v) \in Y_i \times Y_i \mid \gamma \cdot 2^{i-1} < d(u, v) \leq \gamma \cdot 2^i\}$. The union of all these edge sets $\widehat{E} = \cup_i E_i$ is the spanner returned by the construction.

The following lemma (appearing as Lemma 2.2.5 in Section 2.2) states that the spanner $\widehat{E}$ preserves distances well:

Lemma 4.5.2 (Low Stretch) *The set of edges $\widehat{E}$ forms a $(1+\varepsilon)$-spanner for (V, d).*

Hence, it suffices to show that $\widehat{E}$ has a small number of edges. We first show that for each i, the set E_i contains a small number of edges, compared to the size of the net Y_i.

Lemma 4.5.3 *If the metric (V, d) has correlation dimension at most k, the size $|E_i| \leq 2^{kp}|Y_i|$.*

Proof: Observe that $|E_i| \leq \sum_{v \in Y_i} |\mathbf{B}_{Y_i}(v, \gamma \cdot 2^i)|$. By using Definition 4.2 for correlation dimension repeatedly, and the fact that $p = \lceil \log_2 \gamma \rceil + 1$, it follows that the sum is bounded by $2^{kp}|Y_i|$. ■

We can now prove half of Theorem 4.5.1; since each $|Y_i| \leq n$ and $2^p = O(\varepsilon^{-1})$, summing the above 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. However, this bound may be $\Theta(n^2)$ if the aspect ratio is large, and we have to work harder to get a bound depending only on n and ε. 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 4.5.4 *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 $|E_i| \leq \frac{1}{2}|U|(|U| + 1)$.*

Proof: By the construction of the edge set E_i, note that if $(u, v) \in E_i$, then $d(u, v) \leq \gamma \cdot 2^i$. 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 $d(x, y) \leq \gamma \cdot 2^{i+1}$. 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 $|E_i| \leq \binom{|U|}{2} + |U|$, which completes the proof. ■

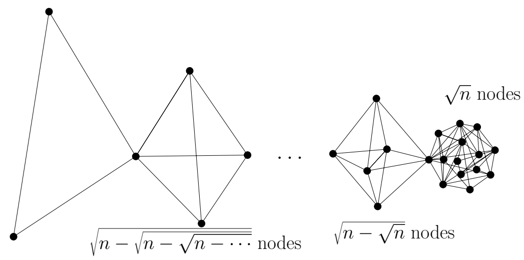


Figure 4.2: Lower bound example for Sparse Spanners.

Lemma 4.5.5 *For any $r \in \{0, 1, \dots, p-1\}$, the edges in all the E_i's with $i \equiv r \pmod p$ is $\sum_j |E_{jp+r}| \leq O(2^{kp/2} n^{1.5})$.*

Proof: Define an upper bound function $F(\cdot)$ such that for any j_0, if $|Y_{j_0 p+r}| = a$, then $\sum_{j \geq j_0} |E_{jp+r}| \leq F(a)$; we want to find the sharpest upper found function $F(\cdot)$ possible. Lemma 4.5.4 implies that if $|U| = b$, then $F(a) \leq \max_b\{\frac{1}{2}b(b+1) + F(a-b)\}$. 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 4.5.3 we have $|E_{j_0 p+r}| \leq 2^{kp} a$. Putting these together forces $F(a) \leq 2^{kp} a + F(a - 2^{kp/2}\sqrt{a})$, and implies that $F(a) = O(2^{kp/2} a^{1.5})$. Since any $|Y_i| \leq n$, the result follows. ■

Applying Lemma 4.5.5 for each $r \in [p]$ and summing up the resulting bounds gives us $|\widehat{E}| \leq O(2^{kp/2} p n^{1.5}) \leq (2 + \frac{1}{\varepsilon})^{O(k)} n^{1.5}$, proving the second part of Theorem 4.5.1.

Note that for metrics with bounded doubling dimension, one can get a $(1+\epsilon)$-spanners with $O(n\varepsilon^{-O(k)})$ edges [CGMZ05, HPM05]. However, we show that such a result is not possible with bounded correlation dimension, and that the upper bound in Theorem 4.5.1 is indeed tight.

4.5.2 Sparse Spanners: Lower Bound

Theorem 4.5.6 (Lower Bound on Sparsity) *There exists a family of metrics with bounded correlation dimension such that for each metric in the family, any 1.5-stretch spanner has at least $\Omega(n^{1.5})$ edges.*

The metric in the lower bound is roughly represented by the picture in Figure 4.2; note that it is essential that this lower bound metric has super-polynomial aspect ratio Δ, for we can obtain such a spanner with $O(n \log \Delta)$ edges from Theorem 4.5.1.

We give a construction for a family of metrics that has bounded correlation dimension, but any 1.5-spanner for any metric in the family must have at least $\Omega(n^{1.5})$ edges.

Let $A \geq 4$ be a parameter, which specifies the difference in distance scales in different levels of the recursive construction. The construction algorithm 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 clarity, we omit all ceilings or floors from the description. For ease of description, 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 inter-point distance α; set u to be any point.
2. Otherwise, construct $M' := M(n - \sqrt{n}, \alpha A)$, together with the special point u'. Replace u' with a uniform metric U with $\sqrt{n} + 1$ points having inter-point distance α. Each point in U has distance to any other point the same as that from u'. Set the special point u to be any point in U.

Lemma 4.5.7 *For all $n \geq 1$, the metric $M(n, 1)$ has correlation dimension at most $O(1)$.*

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

(1) If $2r < A^i$, 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 $2r \geq A^i > r$, 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 $2r \geq A^{\widehat{i}} > r$, where $\widehat{i} > i$. Let $p := |\mathbf{B}_N(u_i, r)|$ 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 $A^{\widehat{i}+1} > 2r > r \geq A^{\widehat{i}}$, 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 (4.2). ■

Theorem 4.5.8 *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 edges in U. Suppose we contract U to a single point in H. Then, the resulting graph is a 1.5-spanner for $M(n - \sqrt{n}, A)$, and hence contain 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 $h(n) \geq \Omega(n^{1.5})$. ■

4.6 Algorithms for Metrics with Bounded Correlation Dimension

Having defined the notion of correlation dimension, and having seen a simple warm-up (obtaining sparse spanners), we now turn to devising algorithms for metric spaces, whose performance is parameterized by the correlation dimension of the underlying metric space. This task is complicated by two issues:

- **Global versus Local Properties.** The notion of correlation dimension is *global*, in the sense that while there may be pockets of "high-complexity" in a metric with low $\dim_C$, the complexity is "low on the average". One should compare this to previous notions of dimension like doubling, where the metric was well-structured in every region and at every scale, and thus local arguments would usually suffice to give good algorithms. In sharp contrast, *we are forced to develop algorithms that take into account this global averaging*.

  As an example, consider the TSP: suppose the input graph consists of a max-SNP hard $(1, 2)$-TSP instance on $\sqrt{n}$ nodes, which is attached to one vertex of a unit grid. If we want to obtain a $(1 + \varepsilon)$ approximation to TSP, our algorithm would have to cluster the graph into the "easy" part (the grid), and the "complicated" part (the $(1, 2)$-TSP instance), and perhaps run a (Q)PTAS on the former part and a constant approximation algorithm on the latter part. Of course, the input metric with $\dim_C = O(1)$ may not have such an obvious clustering.

- **Doubling results may not be applicable.** As noted in the discussion after Lemma 4.2.3, metrics with $\dim_C = O(1)$ cannot have near-uniform sets of size $\omega(\sqrt{n})$, and hence their doubling dimension is at most $\frac{1}{2}\log_2 n + O(1)$. Hence, while we can conceivably use results for doubling metrics, most of the current results are no longer interesting for that range of doubling dimension: e.g., the results for TSP have a running time of $\exp\{(\varepsilon^{-1}\log n)^{O(\dim_D)}\}$, and hence plugging in $\dim_D = \frac{1}{2}\log_2 n$ does worse than $n!$, the running time for an exact algorithm. Again, our algorithms will try to avoid this simple-minded reduction to doubling, even though they will rely on many ideas developed in the doubling metrics literature.

In the rest of the paper, the two main algorithmic results we present are:

- **Weak TSP Approximation & Embedding into Small Treewidth Graphs.** We first show how to solve the TSP on metrics with low correlation dimension within $(1 + \varepsilon)$ in time $2^{\sqrt{n}\cdot(\varepsilon^{-1}\log n)^{O(\dim_C)}}$. As a by-product, we also get Theorem 4.1.2: a random embedding of the original metric into a graph with treewidth $\sqrt{n} \cdot (\varepsilon^{-1}\log n)^{O(\dim_C)}$. Details of this result appear in Section 4.6.1.

  To prove these results, we adopt, adapt and extend the ideas of Arora [Aro02] and Talwar [Tal04]. The main conceptual hurdle to our result is that all the previous proofs use "O(1)-padded decompositions," and metrics with small $\dim_C$ may not admit such good padded decompositions, since padding is a local property, and our metric may have some dense regions. We show how to get around this requirement: we use known padded decompositions with poorer padding guarantees, and show that carefully altering the boundaries suffices for our purpose.

- $(1 + \epsilon)$**-Approximations in Sub-exponential Time.** The ideas we use for the previous algorithm are still fairly local, and hence do not fully use the power of having small correlation dimension. In Section 4.7, we show how to improve our partitioning scheme, and use an improved global charging scheme to get our main result Theorem 4.1.3: an approximation scheme for TSP that runs in sub-exponential time.

4.6.1 An Algorithm for TSP in Time $2^{\tilde{O}(\sqrt{n})}$

Given an $\epsilon \leq 1$, we consider randomized $(1+\epsilon)$-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.

As is well-known, we can assume the aspect ratio is n/ϵ (see, e.g., [Aro02, Tal04]), by the following simple argument. Suppose Δ is the aspect ratio of the metric M, and the minimum distance in the metric is 1. Let V_a be an $\epsilon_a \Delta/n$-net of M. Suppose OPT_a is the length of an optimal tour for points in V_a only. 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 $n \cdot 2e_a\Delta/n = \epsilon_a \cdot 2\Delta \leq \epsilon_a\mathsf{OPT}$. Hence, we will assume that $V_a = V$, and that our metric has an aspect ratio of at most n/ϵ.

Moreover, we assume that $\epsilon > 1/n$, or else we can solve it exactly in $2^{O(\epsilon^{-1}\log\epsilon^{-1})}$-time. We use the following main ideas, which were also used in obtaining known (Q)PTAS's for TSP [Aro02, Tal04]:

(a) We find a good probabilistic hierarchical decomposition into clusters with geometrically decreasing diameters, **(b)** we choose 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, i.e., the tour is *portal-respecting*. The main structure lemma shows that the expected cost of the best portal-respecting tour is at most $(1+\epsilon)$ times its original cost. Finally, **(c)** we find the best portal respecting tour using dynamic programming: 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 4.6.4.) Since the total number of clusters is $\mathrm{poly}(n)$, total runtime is $\mathrm{poly}(n)2^{O(B\log B)}$. Note that for doubling metrics, since each cluster had only $2^{O(\dim_D)}$ child clusters, each with $O(\epsilon^{-1}\log n)^{O(\dim_D)}$ portals, the runtime is quasi-polynomial [Tal04].

The two main problems that we face are the following:

(i) Metrics with low correlation dimension do not admit $O(1)$-padded decompositions which are traditionally used in step (a) above, and
(ii) While we can ensure that the number of portals in any single cluster are at most $\approx O(\sqrt{n})$ using Lemma 4.2.3, each cluster may have as many as $\sqrt{n}$ child clusters, and hence the size B of the union of portals for all the child clusters may be close to $\Theta(n)$.

To take care of these problems, we need to find a new partitioning and portaling scheme, such that the union of the portals in each cluster *and in all its child clusters* has size only $\tilde{O}(\sqrt{n})$; clearly this will require us to do the partitioning and portal-creation steps in a dependent fashion, with each step guiding the other. (Moreover, we will argue that the lack of $O(1)$-padded decompositions does not hurt us much; this will turn out to be the easy part.)

We formalize the above ideas in the Sections 4.6.2, 4.6.3 and 4.6.4.

4.6.2 Hierarchical Decomposition and Portal-Respecting Tour

In this section, we show how probabilistic hierarchical decomposition and portal assignment can be used to approximate TSP. In particular, we show that it is sufficient to restrict our attention to *portal-respecting tours* in order to get an $(1+\epsilon)$-approximation.

Given a metric (V, d), we assume unit minimum distance, and hence the aspect ratio and the largest distance are denoted by Δ. Recall that we can assume $\Delta \leq \frac{n}{\epsilon}$.

Hierarchical Decomposition. Let $L := \lceil \log_H(n/\epsilon) \rceil$ be the number of levels in the system, with $D_L := \Delta$ and $D_{i-1} := D_i/H$, where $H \geq 4$ is a parameter that can possibly depend on n. For each i, $\mathcal{P}_i$ will be a partition of V such that each cluster has diameter at most D_i. Note that if $\mathcal{P}_L$ consists of just one cluster containing all points in V, then each point in V forms a separate cluster in the partition $\mathcal{P}_0$. The family of partitions $\{\mathcal{P}_i\}$ is *hierarchical* if each height-i cluster is contained in some height-$(i+1)$ cluster.

Portal Assignment. For each $0 \leq i < L$, each height-i cluster C has a set $U(C)$ of points called *portals* such that $U(C)$ *is a* βD_i*-covering of* $U(C)$, where $0 < \beta < 1$ is a parameter to be determined later. The portals will satisfy the condition that if a point is a portal for a height-i cluster, then it must be a portal for all lower height clusters. A *child portal* for a cluster is a portal in one of its child clusters. We are looking for a tour that satisfies the *portal condition*:

> 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.** In order to show that the expected length of the restricted tour following the portal condition is not too much larger than that of the optimal tour, we require that the random D_i-bounded partition $\mathcal{P}_i$ sampled from the hierarchical decomposition satisfies the α-padded property. Recall this means that if a set $S \subseteq V$ has diameter d, then it is partitioned by $\mathcal{P}_i$ with probability at most $\alpha \cdot \frac{d}{D_i}$. In particular, the following condition must be satisfied.

> Suppose $u, v \in V$. Suppose B_u and B_v are balls of radius r around u and v respectively. Then, the probability that the set $S := B_u \cup B_v$ is partitioned by $\mathcal{P}_i$ is at most $\alpha \cdot \frac{2r + d(u,v)}{D_i}$.

Given a partition P and a point x, we use $P(x)$ to denote the cluster in P that contains x. Observe that a standard probabilistic decomposition like those by Bartal [Bar96] and Fakcharoenphol et al. [FRT04] gives $\alpha = O(\log n)$.

Lemma 4.6.1 *Suppose $\{\mathcal{P}_i\}$ is an α-padded hierarchical decomposition of (V, d), with portals for each cluster as described above. Then, for any $u, v \in V$, the expected increase in the shortest path obeying the portal condition is at most $6L\alpha\beta \cdot d(u, v)$.*

Proof: Consider the event that u and v are separated in $\mathcal{P}_i$, but not separated in $\mathcal{P}_{i+1}$. This probability is at most $\alpha \cdot d(u, v)/D_i$. Under this event, the shortest path from u to v satisfying the portal condition is at most $(1 + 6\beta D_i)d(u, v)$, i.e., the distance from u to v increases by at most $6\beta D_i$. The bound is $6\beta D_i$, instead of $4\beta D_i$, because it might not be possible for u to reach its closest height-i portal directly. It might have to go through all the lower height portals first. Hence, summing over all heights i, we have shown that the expected increase in the shortest path 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 4.6.1, we can show that by forcing the tour to satisfy the portal condition, the length of the resulting optimal tour does not increase too much.

Proposition 4.6.2 *Suppose OPT_0 is the length of the optimal tour for points in V, satisfying the portal condition with respect to the hierarchical decomposition $\{\mathcal{P}_i\}$ and the corresponding portals for each cluster. Then, $E[\mathsf{OPT}_0] \leq (1 + 6L\alpha\beta)\mathsf{OPT}$.*

4.6.3 A Partitioning and Portaling Algorithm

In the previous section, we showed how a suitable hierarchical decomposition and portaling scheme can restrict the search space of potential tours. In this section we give a concrete construction of a probabilistic hierarchical decomposition and portaling scheme such that both the padding parameter α and the number B of child portals for each cluster are small.

Observe that if the child portals of each cluster form a packing, then using the bounded correlation dimension assumption and Lemma 4.2.3, we can show that B is small for each cluster. If we use a standard hierarchical decomposition (e.g. one by Bartal [Bar96] or FRT [FRT04]) and choose an appropriate net for each cluster to be its portals, then the child portals of a cluster need not be a packing, because portals near the boundary of 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 adjustment to the boundaries of clusters. Here is the main result that describes the properties of the hierarchical decomposition and portaling scheme.

Theorem 4.6.3 (Main Partition-&-Portal Theorem) *Given a metric (V, d) with $\dim_C = k$, and a parameter $\beta \leq 1$, there is a polynomial-time procedure that returns a probabilistic hierarchical partition of the metric with*

- *(A1) The diameter of a height-i cluster is guaranteed to be at most $D_i + \beta D_{i-1}$, where $D_i = 4^i$.*
- *(A2) The probability of (u, v) being separated at height i is at most $O(\log^2 n) \times \frac{d(u,v)}{D_i}$.*

Moreover, each cluster C is equipped with a set of portals $U(C)$ *such that the following properties hold:*

- *(B1) For each non-root cluster C at height-i, the set of portals $U(C)$ forms a $\beta\, D_i$-covering of C.*
- *(B2) Moreover, the set of portals in C and all its children form a $(\beta/4)\, D_{i-1}$-packing.*

The Randomized Partitioning and Portaling Algorithm

Consider the metric (V, d) with unit minimum distance, and hence the aspect ratio being the diameter Δ of the metric. (Moreover, $\Delta \leq n/\epsilon$, as noted before.) Let $H := 4$, and $L := \lceil \log_H(n/\epsilon) \rceil$. Set $D_L := \Delta$, and $D_{i-1} := D_i/4$, as discussed before. We will give a hierarchical decomposition of (V, d) such that for each height-i cluster cluster C, the set $U(C)$ of portals is a is a βD_i-covering of C and its child portals is a $\frac{1}{4}\beta D_{i-1}$-packing, as described in the statement of Theorem 4.6.3.

1. Let $\mathcal{P}_L = \{V\}$ and $U(V) = \emptyset$.
2. For $i = L - 1$ down to 0,

For each height-$(i+1)$ cluster $C \in \mathcal{P}_{i+1}$,

(a) Apply Bartal's probabilistic decomposition [Bar96] on cluster C, using n as an upper bound on the number of points in C, such that the diameter of each resulting sub-cluster is at most D_i. This induces an initial partition $\widetilde{\mathcal{P}}_i$ on C.

(b) Boundary Adjustment using Bartal's decomposition [Bar96]

i. Note that $U(C)$ is a $\frac{1}{4}\beta D_{i+1}$-packing and $D_{i+1} = 4D_i$. Augment $U(C)$ to a βD_i-net $\widehat{U}(C)$ of C. Let Z be the set of points z in C that has no point in $\widehat{U}(C) \cap \widetilde{\mathcal{P}}_i(z)$ within distance βD_i.

ii. Let $W := Z$, $X := C$, and $\overline{U}(C) := \emptyset$.

iii. While W is non-empty,

A. Pick any point u from W. Let $r := \beta D_i / 4 \ln n$. Pick $z \in [0, \frac{1}{4}\beta D_i]$ randomly from the distribution $p(z) := \frac{n}{n-1} \cdot \frac{1}{r} e^{-z/r}$. Let $B := B(u, \frac{1}{4}\beta D_i + z)$.

B. If B contains some point c in $\widehat{U}(C)$, then all points in $B \cap X$ are moved to the height-i cluster currently containing c, otherwise, add u to $\overline{U}(C)$, and move all points in $B \cap X$ to the height-i cluster currently containing u.

C. Remove points in B from both X and W.

iv. Let the new partition on C be $\mathcal{P}_i$. For each new height-i cluster C', let $U(C') := C' \cap (\widehat{U}(C) \cup \overline{U}(C))$.

The Analysis

Lemma 4.6.4 (Correctness) *For $i < L$, for any height-$(i+1)$ cluster C produced by the Decomposition Algorithm, then (1) for any child cluster C' of C, the set $U(C')$ is a βD_i-covering 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 for a height-$(i+1)$ cluster C, the set $U(C)$ is a $\frac{1}{4}\beta D_{i+1}$-packing, then for any child cluster C' of C, $U(C')$ is a βD_i-covering 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. Then, the result follows by induction on i, with the base case starting at $i = L$, as the empty set $U(V)$ is trivially $\frac{1}{4}\beta D_L$-separated.

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 the subset $U(C)$ is a $\frac{1}{4}\beta D_{i+1}$-packing and $D_{i+1} = HD_i \geq 4D_i$, it can be augmented to be a βD_i-net $\widehat{U}(C)$ for C. Observe that points in $\widehat{U}(C)$ are not reassigned to different height-i clusters in the boundary adjustment step.

Let x be a point in C. We show that 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. Recall that Z is the set of points z in C that has no point in $\widehat{U}(C) \cap \widetilde{\mathcal{P}}_i(z)$ within distance βD_i.

Suppose x is not in Z. 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 again that points in $\widehat{U}(C)$ stay in the same clusters. Hence, if point x is not reassigned to another height-i cluster, then it is still covered by the point v after boundary adjustment. Otherwise, point x is in some ball B with radius at most βD_i, which contains a point in $\widehat{U}(C) \cup \overline{U}(C)$. After that, all points in B will be removed from G and stay in the same height-i cluster throughout the boundary adjustment process.

If x is in Z, then at some point it must be removed from list L. Then, by a similar argument, at some point x must be in some ball B with diameter at most βD_i, which contains a point in $\widehat{U}(C) \cup \overline{U}(C)$. The same argument follows.

We next show that $\widehat{U}(C) \cup \overline{U}(C)$ 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 distance more than $\frac{1}{4}\beta D_i$ from $\widehat{U}(C)$ and existing points in $\overline{U}(C)$. Hence, the packing property follows. ∎

Lemma 4.6.5 (Separation Probability) *For each level i, $\Pr[(u,v)$ separated by $\mathcal{P}_i] \leq O(\log^2 n)\frac{d(u,v)}{D_i}$.*

To prove Lemma 4.6.5, we use the following results which can be proved using techniques in [Bar96].

Lemma 4.6.6 *Suppose B_u and B_v are balls centered at $u, v \in V$ respectively with radius r. Then, for each i, the probability that the union $B_u \cup B_v$ is cut apart by $\widetilde{\mathcal{P}}_i$ in the first phase is at most $O(\log n) \cdot \frac{d(u,v)+2r}{D_i}$.*

Lemma 4.6.7 *Suppose $u, v \in V$. Then, the probability that u and v are separated in the boundary adjustment step is at most $O(\log n) \cdot \frac{d(u,v)}{\beta D_i}$, and this is independent of what happens in the first phase.*

Proof of Lemma 4.6.5: Consider $u, v \in V$. Let B_u and B_v be the balls centered at u and v respectively with radius $\frac{1}{2}\beta D_i$. First consider the case when $d(u,v) \geq \beta D_i$. Note that if the union of B_u and B_v is not separated by $\widetilde{\mathcal{P}}_i$, then u and v cannot be separated by $\mathcal{P}_i$. Hence, the probability that $\mathcal{P}_i$ separates u and v is upper bounded by that of the former event, which is at most $O(\log n) \cdot \frac{d(u,v)+\beta D_i}{D_i}$, by Lemma 4.6.6. By the assumption that $d(u,v) \geq \beta D_i$, the probability is at most $O(\log n) \cdot \frac{d(u,v)}{D_i}$.

Consider the case when $d(u,v) < \beta D_i$. Note that if u and v are separated eventually, then the union of B_u and B_v must be cut apart by $\widetilde{\mathcal{P}}_i$. Moreover, in the boundary adjustment step, the points u and v must also be separated.

Hence, by Lemmas 4.6.6 and 4.6.7, this probability is upper bounded by

$$O(\log n) \cdot \frac{d(u,v)+\beta D_i}{D_i} \cdot O(\log n)\frac{d(u,v)}{\beta D_i} \leq O(\log^2 n)\frac{d(u,v)}{D_i}.$$

Thus, we have analyzed both cases and this completes the proof. ∎

Observe that we have not used the notion of correlation dimension so far. In the following lemma, we use the definition of correlation dimension to bound the number of child portals in a cluster.

Lemma 4.6.8 (Small Number of Child Portals) *Suppose the metric space (V, d) has correlation dimension at most k. For all clusters C, 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 Lemma 4.6.4, 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 from the construction, all points in C is contained in a ball with radius at most $(D_{i+1} + \beta D_i)/2$, though not necessarily centered at a point in N. Since N is a $4^i\beta$-net, C is contained in a ball of radius at most $D_{i+1} + \beta D_i$ centered at some net point $u \in N$. Hence, $S \subseteq \mathbf{B}_N(u, D_{i+1} + \beta D_i)$, which by Lemma 4.2.3 has size at most $(16/\beta + 4)^{k/2}\sqrt{|N|} \leq (16/\beta + 4)^{k/2}\sqrt{n}$. ■

4.6.4 Dynamic Programming 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 by Arnbourg and Proskurowski [AP89] and Arora [Aro02], and we give the details here for completeness.

For each cluster C with its portals $U(C)$, there are entries indexed by (J, I), where J is a set of unordered 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 non-portals of cluster C using entry and exit portals described by pairs in J. Moreover, for each point in I, the two points adjacent to it in the tour are in the cluster C. Hence, the points in I are not behaving as portals in the sense that the tour does not enter or exit the cluster C through the points in I. 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 only count the part of tour that is between u and v for some pair $\{u, v\}$ in J. The entry indexed by (J, I) keeps the smallest possible sum of the lengths of the 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.

There are special entries each of which is indexed by only a single portal $x \in U(C)$. This corresponds to the (sub-optimal) case where 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 outlined in [AP89], for a cluster such that the number of child portals is at most B, the time to complete all entries for that cluster is $2^{O(B \log B)}$. Note that 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.

4.6.5 The First TSP Algorithm

Using the partitioning and portaling scheme described in Section 4.6.3 and the dynamic program described in Section 4.6.4, we have an algorithm for approximating TSP.

Theorem 4.6.9 (The First TSP Algorithm) *There is a randomized algorithm for metric TSP, which for metrics with* $\dim_C = k$*, returns a tour of expected length at most* $(1+\epsilon)\mathsf{OPT}$ *in time* $2^{((\log n)/\epsilon)^{O(k)}\sqrt{n}}$.

Proof: Since the aspect ratio of the metric is at most n/ϵ, and the hierarchical partition decreases the diameter of components by a constant factor at each level, the height of the decomposition is $L = O(\log \frac{n}{\epsilon})$. By Theorem 4.6.3, each edge (u, v) of the optimal tour is cut at height-i with probability $\alpha \frac{d(u,v)}{D_i}$ with $\alpha = O(\log^2 n)$.

We set $\beta := \frac{\epsilon}{6L\alpha}$. By Proposition 4.6.2, the expected length of this tour (and hence the length of the optimal portal-respecting tour) is at most $(1 + 6L\alpha\beta) = (1+\epsilon)\mathsf{OPT}$.

We need to also 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 $B^{O(B)}$ runtime.

By Lemma 4.6.8, it follows that $B \leq (16/\beta+4)^{k/2}\sqrt{n}$. Hence, the running time of the algorithm is $nL \cdot 2^{O(B \log B)} = \exp\{(\epsilon^{-1}\log n)^{O(k)}\sqrt{n}\}$, as required. ∎

4.6.6 Embedding into Small Treewidth Graphs

Observe that our probabilistic hierarchical decomposition procedure actually gives an embedding into a distribution of low treewidth graphs. 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 4.6.1, the expected distortion of the distance between any pair of points is small. Using the same parameters as in the proof of Theorem 4.6.9, we have the following theorem.

Theorem 4.6.10 (Embedding into Small Treewidth Graphs) *Given any constant* $0 < \epsilon < 1$ *and* k*, metrics with correlation dimension at most* k *can be embedded into a distribution of graphs with treewidth* $((\log n)/\epsilon)^{O(k)}\sqrt{n})$ *and distortion* $1+\epsilon$.

4.7 A Sub-Exponential Time $(1+\epsilon)$-Approximation for TSP

In the previous section, we saw how to get a $(1+\epsilon)$-approximation algorithm for TSP on metrics with bounded correlation dimension, essentially using the idea of random embeddings into small treewidth graphs. The approach gives approximations for any problem on metric spaces which can be solved for small-treewidth graphs: however, it is limited by the fact that the $\sqrt{n}$-lollipop graph metric has bounded correlation dimension, and randomly $(1+\epsilon)$-approximating this graph requires the use of graphs with large treewidth.

In this section, we get an improved approximation for TSP using another useful observation. Consider the bad examples in Figure 4.1: the contribution to OPT due to the dense structure is much smaller than that from the low-dimensional ambient structure. For example, for the sub-grid with a $(1, 2)$-TSP instance tacked onto it (Figure 4.1(b)), we can obtain a $(1 + \epsilon)$-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 only contributes $\Theta(\sqrt{n})$ to OPT). Of course, this is a simple case where the clustering is obvious; our algorithm must do some kind of clustering for all instances. Moreover, this indicates that we need to do a global accounting of cost: the sloppy approximation of the "hard" subproblem needs to be charged to the entire OPT, and not just the optimal tour on the subproblem.

Here are some of the issues we need to address (most of which are tied to each other), along with descriptions of how we handle them:

- **Avoiding Large Tables.** The immediate hurdle to a better runtime is that some cluster may have $\Theta(\sqrt{n})$ child portals and we have to spend $\sqrt{n}^{\sqrt{n}}$ 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 more than $B > B_0$ portals among its children, we compute, in linear time, a tour on C that *only enters and leaves C once*, but now we incur an extra length of $B \times \mathrm{diam}(C)$ in the final tour we compute. In the sequel, we call this extra length the "MST-loss". This step implies that we need only spend $\min\{O(B), 2^{O(B_0 \log B_0)}\}$ time on any table computation. The patching procedure used here is reminiscent of the patching from [Aro96], and is described in Section 4.7.2.
- **Paying for this 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 at 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 $\epsilon\mathsf{OPT}$.
- **A Potential Charging Scheme.** To be able to charge MST-losses in a global manner, we look at the hierarchical decomposition. The extra length incurred for patching height-i clusters is proportional to the number of child portals of the clusters to which patching is applied. *If the union of all the height-$(i-1)$ portals in the decomposition satisfied some packing condition*, we could use Lemma 4.2.3 to bound the number of them, and hence the total MST-loss at height-i of the decomposition tree. However, the techniques developed so far (in Section 4.6.1) can only ensure that the child portals of a *single cluster* form a packing: we clearly need new techniques.
- **A New Partitioning & Portaling Procedure.** The method in the last section took a cluster C at height-$(i + 1)$, cut it up, and then adjusted the boundaries of the subclusters created at height-i to ensure that the union of the portals in these subclusters formed a packing. However, the portals in all the grand-children of C (i.e., all the clusters at height-$(i - 1)$ below C) may not form a packing: hence we have to re-adjust the boundaries created at height-i yet again. In fact, when clusters at a certain level are created, the boundaries for clusters in all higher levels have to be readjusted. This can potentially increase the probability that a pair of points are separated at each level. This is resolved by ensuring

that cluster diameters fall by logarithmic factors instead of by constants. The details are given in Section 4.7.1.

- **Avoiding Computation of Correlation Dimension.** As given in Theorem 4.2.4, it is hard to approximate the correlation dimension of a given metric. However, the algorithm can guess the correlation dimension k of the input metric. It starts from small values of k and for each net encountered, it takes polynomial time to verify the bounded average growth rate property (4.2). Whenever property (4.2) is violated for some net, we know the current estimation of the correlation dimension is too small. The value of k is increased and the algorithm is restarted. 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 would not affect the asymptotic running time.

We formalize the ideas sketched above in the following. The general framework described in Section 4.6.2 of using hierarchical decomposition and portals to approximate TSP still applies here. We give a more sophisticated partitioning and portaling scheme in Section 4.7.1, and analyze the MST-loss incurred from patching in Section 4.7.2.

4.7.1 The Modified Partitioning and Portaling Algorithm

The main difference is that when a height-i partition is performed, all higher height partitions are modified, in order to ensure that all height-i portals form a packing. Let $H \geq 4$ be a parameter (possibly depending on n) that will be determined later. Let $L := \lceil \log_H (n/\epsilon) \rceil$. Set $D_L := \Delta$, the diameter of (V, d); $D_{i-1} := D_i / H$.

We are going to give a hierarchical decomposition of (V, d) such that for each height i, U_i is the set of height-i portals such that for each height-i cluster C, the set $U_i \cap C$ of portals is a is a βD_i-covering of C and U_i is a $\frac{1}{4}\beta D_{i-1}$-packing. Observe that once a U_i is formed, it will not be modified; moreover, once a point is chosen to be a portal for a cluster, it will not be moved to another cluster.

1. Let $\mathcal{P}_L = \{V\}$ and $U_L = \emptyset$.
2. For $i = L - 1$ down to 0,
   (a) For each height-$(i+1)$ cluster $C \in \mathcal{P}_{i+1}$, apply Bartal's probabilistic decomposition [Bar96] on cluster C, using n as an upper bound on the number of points in C, such that the diameter of each resulting sub-cluster is at most D_i. This induces a temporary partition $\widetilde{\mathcal{P}}_i$ on C.
   (b) Boundary Adjustment using Bartal's decomposition [Bar96]:
      i. Note that U_{i+1} is a $\frac{1}{4}\beta D_{i+1}$-packing and $D_{i+1} = HD_i \geq 4D_i$. Augment U_{i+1} to a βD_i-net $\widehat{U}_i$ of V. Let Z be the set of points z in V that has no point in $\widehat{U}_i \cap \widetilde{\mathcal{P}}_i(z)$ within distance βD_i.
      ii. Let $W := Z$, $X := V$, and $\overline{U}_i := \emptyset$.
      iii. While W is non-empty,
         A. Pick any point u from W. Let $r := \beta D_i / 4 \ln n$. Pick $z \in [0, \frac{1}{4}\beta D_i]$ randomly from the distribution $p(z) := \frac{n}{n-1} \cdot \frac{1}{r} e^{-z/r}$. Let $B := B(u, \frac{1}{4}\beta D_i + z)$.

B. If B contains some point c in $\widehat{U}_i$, then all points in $B \cap X$ are moved to the height-i cluster currently containing c, otherwise, add u to $\overline{U}_i$, and move all points in $B \cap X$ to the height-i cluster currently containing u.

C. Remove points in B from both X and W.

iv. Observe that the partitions $\mathcal{P}_j$ for $j > i$ can be modified. Let the new height-i partition on V be $\mathcal{P}_i$. Set $U_i := \widehat{U}_i \cup \overline{U}_i$.

Analyzing the probability of a pair being separated

We first analyze the probability that a pair of points u, v are separated right after some partition $\mathcal{P}_i$ is formed for the first time. Since the decomposition procedure is quite sophisticated, the analysis is done more carefully than before. First, we rephrase a result concerning the Bartal's decomposition [Bar96].

Fact 4.7.1 *There exists $t > 0$ such that any n point metric space can be probabilistically decomposed 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 partitioned is at most*

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

Throughout this subsection, the parameter t refers to the one that comes from Fact 4.7.1. However, we prove the following lemma, which is more general and is used later. Recall that $D_{i+1} := HD_i$. For technical reason, we assume that $H \geq 4t \log n$, which we shall see is not a problem.

Lemma 4.7.2 *Suppose $u, v \in V$, and B_u and B_v are balls of radius r centered at u and v respectively. Then, the probability that $B_u \cup B_v$ is separated by $\mathcal{P}_i$, right after $\mathcal{P}_i$ is formed for the first time, is at most $4t^2 \log^2 n \cdot \frac{d(u,v)+2r}{D_i}$.*

Proof: We show 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 $r' := r + \frac{1}{2}\beta D_i$. Observe that if $\mathcal{P}_i$ separates $B_u \cup B_v$, then one or both of the following events happen.

1. Event A: The partition $\mathcal{P}_{i+1}$ separates $B(u, r') \cup B(v, r')$ right after it is formed.
2. Event B: The partition $\tilde{\mathcal{P}}_i$ separates $B(u, r') \cup B(v, r')$.

The probability of event A is, by the induction hypothesis, at most $4t^2 \log^2 n \cdot \frac{\delta+\beta D_i}{D_{i+1}}$; the probability of the event $B \setminus A$ is at most $t \log n \cdot \frac{\delta+\beta D_i}{D_i}$, by Fact 4.7.1. Hence, observing that $D_{i+1} = HD_i \geq 4tD_i \log n$, the probability of the event $A \cup B$ is at most

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

Case 1: $\delta \geq \beta D_i$. Then, the above probability is at most $4t \log n \cdot \frac{\delta}{D_i}$.

Case 2: $\delta < \beta D_i$. Observe that in order for $\mathcal{P}_i$ to separate $B_u \cup B_v$, in addition to the event $A \cup B$, the event that $B_u \cup B_v$ is separated during the boundary adjustment step must also occur. Note that the probability that this latter event happens given the event $A \cup B$ is at most $t \log n \cdot \frac{\delta}{\beta D_i}$. Hence, it follows that the required probability is at most

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

■

Using Lemma 4.7.2, we show the following lemma.

Lemma 4.7.3 *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 $\mathcal{P}_i$ must separates u and v right after some $\mathcal{P}_j$, where $j \leq i$, is formed. Let this event be E_j. We consider the probability of such event E_j. Observe that in order for this to happen, then for each $j \leq l < i$, the partition $\mathcal{P}_i$ has to separate $B(u, \beta D_{l-1}) \cup B(v, \beta D_{l-1})$, due to boundary adjustment at height l, right after $\mathcal{P}_l$ is formed. 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 E_j is at most:

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

where the first term comes from Lemma 4.7.2, and each subsequent terms comes from Fact 4.7.1 applied to each boundary adjustment step. Now, summing $Pr[E_j]$ over $j \leq i$ shows that the probability that (u, v) is cut by the final $\mathcal{P}_i$ is at most $(4t \log n)^L \cdot \frac{d(u,v)}{D_i}$. ■

All portals in each level form a packing

Using the same argument as in Lemma 4.6.4, we can prove the following lemma.

Lemma 4.7.4 *For each height i, the set U_i of height-i portals is a $\frac{1}{4}\beta D_{i-1}$-packing and for each height-i cluster C, the set $U_i \cap C$ of portals is a βD_i-covering of C.*

The only thing to watch out is that when a point x is being assigned to another cluster during boundary adjustment at height i, how do we know x still has a near portal for higher heights? The observation is that portals are not re-assigned to another cluster once they are chosen. Since the point x is near some height-i portal y, which has all higher height portals nearby, we conclude that x still has higher height portals nearby.

4.7.2 Handling Large Portal Sets via Patching

Patching a single cluster

If a cluster C has many child portals (say about $\sqrt{n}$ portals), it is too expensive to compute the entries corresponding to C. In particular, computing the standard TSP table for this cluster would require $O(\sqrt{n}^{\sqrt{n}}) = 2^{\tilde{O}(\sqrt{n})}$ time, which in itself would dash all hopes of obtaining a subexponential time algorithm. To avoid this, we do a two step patching described in the following. The first idea is simple: if we are willing to pay an extra $O(BD)$ amount, where B is the number of portals, and D the diameter of the cluster, then we can find a tour that enters and leaves at a single portal. Indeed, we can find a tour that enters cluster C through some portal x, performs a traveling salesperson tour on points in cluster C, and leaves cluster C through x.

Proposition 4.7.5 (Patching to get a Single Portal) *Suppose cluster C has diameter D, and that there are at most B portals in the cluster C. Then, given any tour on the vertices V, the tour can be modified such that it enters and leaves the cluster C through a* single portal *with additional length at most BD.*

However, *computing such a tour* requires work as well, and we need to ensure that this computation can be done fast: if cluster C has too many child portals, it would be too expensive to compute the optimal tour inside C. Hence, we need a second patching step.

Proposition 4.7.6 *Consider the dynamic program in Section 4.6.4, and look at a cluster C with diameter D and 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 $l + BD$, but now only takes time $O(B)$.*

Proof: From each child cluster C_λ of C, pick the entry such that the length l_λ of its partial segments is smallest. 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 Propositions 4.7.5 and 4.7.6, 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 $2BD$.

Applying Patching Technique in Dynamic Program

We analyze the increase in tour lengths when we apply the patching procedure described in Section 4.7.2. Recall that OPT_0 is the length of the optimal tour returned by the dynamic program (without patching) described in Section 4.6.4.

Suppose patching is applied for clusters with more than B_0 child portals, but only up to height-i clusters, and no patching is applied for clusters in height higher than i. Let the length of the optimal tour returned in such a way be OPT_i. Observe that OPT_L is the length of the tour returned by the dynamic program if patching is applied whenever appropriate.

The following lemma shows that the extra length incurred by patching all clusters in one level is small. Recall k is the correlation dimension of the metric.

Lemma 4.7.7 *For $0 \le i < L$, $\mathsf{OPT}_{i+1} \le \mathsf{OPT}_i + \frac{1}{B_0}(\frac{8H}{\beta})^{k+1}\mathsf{OPT}$.*

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 4.7.2, it follows the extra length to patch up all appropriate height-$(i+1)$ clusters is at most $2\sum_\lambda B_\lambda D_{i+1}$. Now, from the definition of correlation dimension, we have for all integers t,

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

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

$$\textstyle\sum_\lambda B_\lambda^2 \leq \sum_{x\in N_i} |\mathbf{B}_{N_i}(x, D_{i+1})| \leq (\frac{8H}{\beta})^k |N_i|. \tag{4.11}$$

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

$$|\Lambda| \leq \tfrac{1}{B_0^2}(\tfrac{8H}{\beta})^k |N_i|. \tag{4.12}$$

Using the Cauchy-Schwartz inequality, we have

$$\textstyle\sum_\lambda B_\lambda \leq \sqrt{|\Lambda| \cdot \sum_\lambda B_\lambda^2}. \tag{4.13}$$

By substituting (4.11) and (4.12) into (4.13), we have

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

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

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

■

Lemma 4.7.7 implies that the total extra length incurred by patching is small.

$$\mathsf{OPT}_L \leq \mathsf{OPT}_0 + \tfrac{L}{B_0}(\tfrac{8H}{\beta})^{k+1}\mathsf{OPT}. \tag{4.14}$$

4.7.3 The Second TSP Algorithm

Theorem 4.7.8 (Sub-exponential time algorithm for TSP) *For any metric with correlation dimension k, we can give a randomized $(1+\epsilon)$-approximation for TSP in time $\exp\{(\epsilon^{-1}2^{\sqrt{\log n \log\log n}})^{4k}\} = 2^{O_{\epsilon,k}(n^\delta)}$, 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 $H \geq 4$. Hence the depth of the tree is $L := \Theta(\log_H(n/\epsilon))$. As indicated above (and proved in Lemma 4.7.3), 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-covering 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 $\epsilon/2$, we set $\beta = O(\epsilon/L\,\alpha)$.

Finally, from an analysis in Section 4.7.2, the length increase from patching (the "MST-loss") is $\frac{L}{B_0}(\frac{8H}{\beta})^{k+1}\mathsf{OPT}$. To make this at most $\epsilon/2$ as well, we set pick B_0 such that $\frac{L}{B_0}(\frac{8H}{\beta})^{k+1} = \epsilon/2$.

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

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

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 $\log H = \sqrt{\log \frac{n}{\epsilon} \log\log n}$.

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

$$nL \cdot 2^{O(B \log B)} = \exp\{(\epsilon^{-1} \cdot 2^{\sqrt{\log n \log\log n}})^{4k}\} = 2^{O_{\epsilon,k}(n^\delta)},$$

for any $\delta > 0$.

■

4.8 Summary and Conclusions

We have 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 dense clusters (of size up to $O(\sqrt{n})$) but have small average growth-rate. We show 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 defining other notions of dimension for metric spaces: it is fairly unlikely that one notion can capture the complexity of metrics (both the local complexity, as in doubling, as well as the global behavior). Since one definition may not fit all situations, it seems reasonable to consider several definitions, whose properties can then be exploited under the appropriate circumstances.

Chapter 5

Conclusion

We have seen in this thesis that there are notions of dimension that are useful in measuring the complexity of a general metric with respect to certain problems. There are other classes of metrics for which good algorithmic guarantees can be obtained, for instance metrics induced by planar graphs have been extensively studied [OS81, GT87, KPR93, AGK$^+$98]. However, it is not always clear how these classes of metrics can be used to measure the complexity of an arbitrary metric. For example, there are polynomial time algorithms [AGK$^+$98, Kle05] for approximating TSP for planar graphs, yet such algorithms do not apply to general metrics, nor do they provide guarantees in terms of the "planarness" of a given metric, a concept which is hard to define or be exploited in the first place. We conclude this thesis by discussing extensions and future directions for the work in each chapter.

5.1 Spanners for Doubling Metrics

In this thesis, the compactness of a spanner is measured by the number of edges in the spanner. However, for applications in which a spanner is used for maintaining physical connections between sites, the weight of the spanner serves as a better objective than the number of edges. For any bounded dimensional Euclidean metric, Narasimhan and Smid [NS07] showed that for any $t > 1$, there is a t-spanner with weight $O(MST)$, where MST is the weight of a minimum spanning tree.

The construction of low weight spanner for Euclidean metrics relies highly on the geometric properties of Euclidean space. However, some standard ideas are still applicable, such as the Kruskal-like construction.

> **Kruskal's algorithm for constructing low weight spanner.** Arrange the edges (u, v) in increasing order of $d(u, v)$ in a list. Start with an empty spanner and consider each edge $e = (u, v)$ in the list. If the distance between u and v in the current spanner is already at most $t \cdot d(u, v)$, then discard the edge e, otherwise include edge e in the spanner.

Such a construction would guarantee that the spanner returned has stretch at most t. The more technical part is how to show that the spanner has low weight. It would be interesting to see if the geometric assumptions used for Euclidean metrics can be replaced by properties ensured by bounded doubling dimension, which are more combinatorial in nature.

5.2 Low Dimensional Embeddings for Doubling Metrics

In Chapter 3, we show that for embedding doubling metrics into Euclidean space, there is a tradeoff between the target dimension and the distortion of the embedding: given an n-point metric (V, d) with doubling dimension $\dim_D$, and any target dimension T in the range $\Omega(\dim_D \log\log n) \leq T \leq O(\log n)$, we show that the metric embeds into Euclidean space $\mathbb{R}^T$ with $O(\log n \sqrt{\dim_D / T})$ distortion.

A question one can ask is: does this tradeoff extend to smaller values of T? We know that for large $T = O(\log n)$. The result is tight with respect to n, because there exists doubling metrics that embed into Euclidean space with distortion at least $\Omega(\sqrt{\log n})$. Yet our result does not apply when the target dimension is small, e.g. $T = O(\dim_D)$. However, from a manuscript of Abraham, Bartal and Neiman, it is possible to obtain target dimension $O(\dim_D)$ at the expense of increasing the distortion to $O(\log^{1+\epsilon} n)$, for some small $\epsilon > 0$. An interesting question would be if our tradeoff actually holds for small target dimension as well. In particular, is it possible to obtain target dimension $O(\dim_D)$ with distortion $O(\log n)$?

Another question is how good this tradeoff is. For example, is it possible to embed a doubling metric into $O(\log\log n)$ dimensions with $O(\sqrt{\log n})$ distortion? Observe that some lower bound for this tradeoff would imply some kind of lower bound for dimension reduction of Euclidean metrics with constant doubling dimension. In particular, the following ideal result would be impossible.

> **Ideal Result for Dimension Reduction in Euclidean Spaces.** Any Euclidean metric with doubling dimension $\dim_D$ can be reduced to $O(\dim_D)$ dimensions with $O(\dim_D)$ distortion.

Suppose we have an arbitrary metric (V, d) with constant doubling dimension. Then, it can be embedded into Euclidean space with distortion $O(\sqrt{\log n})$ (and $O(\log n)$ dimensions). Observe this would increase the doubling dimension up to at most $O(\log\log n)$. Hence, the ideal result would imply that any doubling metric can be embedded into Euclidean space with $O(\log\log n)$ dimensions and $O(\sqrt{\log n}\log\log n)$ distortion. Hence, the ideal result would be impossible if there exists some $\epsilon > 0$, such that there are doubling metrics for which embedding them into Euclidean space with $O(\log\log n)$ dimensions would incur distortion at least $\Omega(\log^{0.5+\epsilon} n)$.

5.3 Global Notion of Dimension

We have introduced net correlation dimension, a global notion of dimension for which we give a sub-exponential time algorithm for approximating TSP on such globally bounded metrics. The question is whether such notion of global dimension has application to other problems. It is conceivable that a different notion of global dimension needs to be considered for a different problem. For instance, with respect to TSP, a global notion of dimension should rule out a metric with a short tail with linear number of points as a simple metric (see Section 4.3.2). However, for a problem like nearest-neighbor query, this metric should be considered as simple, because if query points are uniformly sampled, then for a constant fraction of the time, the simple part of the metric is being queried.

One can imagine that our techniques could be applied to the setting described in Arora's survey on approximation schemes for hard geometric optimization problems [Aro03] to get subexponential time algorithms, in a way analogous to how we tackle TSP. Although our definition of correlation dimension was not tailored to specifically solve TSP, in retrospect TSP has a nice structure which allows the techniques of hierarchical decomposition, portal assignment and dynamic program (DP) to be employed for metrics with bounded correlation dimension. We describe some properties of TSP that are essential for our techniques to be applied, and mention how some of the other geometric problems do not satisfy them.

1. **Each entry in the DP can be computed in $2^{\tilde{O}(B)}$ time, where B is the number of child portals.** For TSP, each entry can be computed in $2^{O(B \log B)}$ time. However, for other problems such as k-connectivity, the time for computing an entry in the dynamic program described in [CL00] is doubly exponential in B. Hence, even if B is $\mathrm{polylog}(n)$, the algorithm is still too inefficient.
2. **In the DP, any valid configurations for child entries can be combined to form a valid configuration for the parent entry.** This is essential because we cannot even afford to consider more than one configuration per child entry, because there can be as many as $O(\sqrt{n})$ child entries. For TSP, we show how this can be done in Proposition 4.7.6. However, in the dynamic program for minimum latency described in [AK01], each configuration carries too much information and so arbitrary valid configurations of child entries cannot be combined to form a consistent valid parent configuration.
3. **In the case when the number B of child portals in a cluster is too large, a patching argument should be applicable to reduce the number of active portals, at a cost of $O(BD)$, where D is the diameter of the cluster in question.** Perhaps this is the most restricting condition in using our definition of correlation dimension. This condition allows us to use our definition in making a global charging argument (see Section 4.7.2). However, for problems like k-median, such patching argument cannot be applied. Although k-median can be somehow tackled for low-dimensional Euclidean metrics [ARR99] without reducing the number of portals, in our case the number of portals can be potentially too large to be handled without any reduction technique.

In light of the limitations outlined above, we think that for approximating the hard geometric optimization problems on metrics that somehow behave well globally, one would need to find alternative ways to characterize the global behavior of a metric or employ different techniques outside Arora's framework [Aro03, Aro02].

Bibliography

[ABC+05] Ittai Abraham, Yair Bartal, T-H. Hubert Chan, Kedar Dhamdhere, Anupam Gupta, Jon Kleinberg, Ofer Neiman, and Aleksandrs Slivkins. Metric embeddings with relaxed guarantees. In *FOCS '05: Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science*, pages 83–100, Washington, DC, USA, 2005. IEEE Computer Society.

[ABN06] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. In *Proceedings of the 38th ACM Symposium on the Theory of Computing (STOC)*, pages 271–286, 2006.

[AC06] Nir Ailon and Bernard Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. In *Proceedings of the 38th ACM Symposium on the Theory of Computing (STOC)*, pages 557–563, 2006.

[Ach00] Dimitris Achlioptas. Database friendly random projections. In *Proceedings of the Twenteenth ACM Symposium on Principles of Database Systems*, 2000.

[ADD+93] Ingo Althöfer, Gautam Das, David Dobkin, Deborah Joseph, and José Soares. On sparse spanners of weighted graphs. *Discrete Comput. Geom.*, 9(1):81–100, 1993.

[ADM+95] Sunil Arya, Gautam Das, David M. Mount, Jeffrey S. Salowe, and Michiel H. M. Smid. Euclidean spanners: short, thin, and lanky. *STOC*, pages 489–498, 1995.

[AGGM06] Ittai Abraham, Cyril Gavoille, Andrew Goldberg, and Dahlia Malkhi. Routing in networks with low doubling dimension. In *The 26th International Conference on Distributed Computing Systems (ICDCS)*, 2006.

[AGK+98] Sanjeev Arora, Michelangelo Grigni, David Karger, Philip Klein, and Andrzej Woloszyn. A polynomial-time approximation scheme for weighted planar graph tsp. In *Proceedings of the9thAnnual ACM-SIAM Symposium on Discrete Algorithms*, 1998.

[AK01] Sanjeev Arora and Ravi Kannan. Learning mixtures of arbitrary Gaussians. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 247–257. ACM Press, 2001.

[Alo91] Noga Alon. A parallel algorithmic version of the local lemma. In *Proceedings of the32ndAnnual IEEE Symposium on Foundations of Computer Science*, 1991.

[AM05] Ittai Abraham and Dahlia Malkhi. Name independent routing for growth bounded networks. In *17th ACM Symposium on Parallelism in Algorithms and Architectures*

(SPAA), 2005.

[AP89] S. Arnborg and A. Proskurowski. Linear time algorithms for np-hard problems restricted to partial k-trees. In *Discrete Applied Mathematics, Volume 23*, 1989.

[Aro96] Sanjeev Arora. Polynomial time approximation schemes for Euclidean TSP and other geometric problems. In *37th Annual Symposium on Foundations of Computer Science (Burlington, VT, 1996)*, pages 2–11. IEEE Comput. Soc. Press, Los Alamitos, CA, 1996.

[Aro98] S. Arora. Polynomial time approximation schemes for Euclidean traveling salesman and other geometric problems. *J. ACM*, 45(5):753–782, 1998.

[Aro02] Sanjeev Arora. Approximation algorithms for geometric TSP. In *The traveling salesman problem and its variations*, volume 12 of *Comb. Optim.*, pages 207–221. Kluwer Acad. Publ., Dordrecht, 2002.

[Aro03] Sanjeev Arora. Approximation schemes for NP-hard geometric optimization problems: a survey. *Math. Program.*, 97(1-2, Ser. B):43–69, 2003. ISMP, 2003 (Copenhagen).

[ARR99] Sanjeev Arora, Prabhakar Raghavan, and Satish Rao. Approximation schemes for Euclidean k-medians and related problems. In *STOC '98 (Dallas, TX)*, pages 106–113. ACM, New York, 1999.

[AS87] N. Alon and B. Schieber. Optimal preprocessing for answering on-line product queries. *Tech. report 71/87*, 1987.

[Ass83] Patrice Assouad. Plongements lipschitziens dans $\mathbf{R}^n$. *Bull. Soc. Math. France*, 111(4):429–448, 1983.

[Awe85] Baruch Awerbuch. Complexity of network synchronization. *Journal of the ACM*, 32(4):801–823, 1985.

[Bar96] Yair Bartal. Probabilistic approximations of metric spaces and its algorithmic applications. In *Proceedings of the 37th Symposium on the Foundations of Computer Science (FOCS)*, pages 184–193, 1996.

[BC03] Bo Brinkman and Moses S. Charikar. On the impossibility of dimension reduction in l1. In *Proceedings of the 44th Symposium on the Foundations of Computer Science (FOCS)*, pages 514–523, 2003.

[BCIS05] Mihai Bădoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropoulos. Low-distortion embeddings of general metrics into the line. In *Proceedings of the thirty-seventh annual ACM symposium on Theory of computing*, pages 225–233, New York, NY, USA, 2005. ACM Press.

[BCIS06] Mihai Bădoiu, Julia Chuzhoy, Piotr Indyk, and Anastasios Sidiropou. Embedding ultrametrics into low-dimensional spaces. In *Proceedings of the 22nd Annual Symposium on Computational Geometry*, 2006.

[Bec91] József Beck. An algorithmic approach to the Lovász local lemma. I. *Random Structures Algorithms*, 2(4):343–365, 1991.

[BF95] Alberto Belussi and Christos Faloutsos. Estimating the selectivity of spatial queries using the 'correlation' fractal dimension. In *Proc. 21st Int. Conf. Very Large Data Bases, VLDB*, 11–15 1995.

[BKL06] A. Beygelzimer, S. Kakade, and J. Langford. Cover trees for nearest neighbor. In *In Proceedings of 23rd International Conference on Machine Learning*, 2006.

[Bou85] Jean Bourgain. On Lipschitz embeddings of finite metric spaces in Hilbert space. *Israel Journal of Mathematics*, 52(1-2):46–52, 1985.

[CDNS95] Barun Chandra, Gautam Das, Giri Narasimhan, and José Soares. New sparseness results on graph spanners. *Internat. J. Comput. Geom. Appl.*, 5(1-2):125–144, 1995. Eighth Annual ACM Symposium on Computational Geometry (Berlin, 1992).

[CG06a] T-H. Hubert Chan and Anupam Gupta. Small hop-diameter sparse spanners for doubling metrics. In *Proceedings of the17thAnnual ACM-SIAM Symposium on Discrete Algorithms*, 2006.

[CG06b] Richard Cole and Lee-Ad Gottlieb. Searching dynamic point sets in spaces with bounded doubling dimension. In *The thirty-eighth annual ACM symposium on Theory of computing (STOC)*, 2006.

[CGMZ05] Hubert T-H. Chan, Anupam Gupta, Bruce M. Maggs, and Shuheng Zhou. On hierarchical routing in doubling metrics. *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 762–771, 2005.

[Cha87] Bernard Chazelle. Computing on a free tree via complexity-preserving mappings. In *Algorithmica 2*, pages 337–361, 1987.

[CK95] Paul B. Callahan and S. Rao Kosaraju. A decomposition of multidimensional point sets with applications to k-nearest-neighbors and n-body potential fields. *J. Assoc. Comput. Mach.*, 42(1):67–90, 1995.

[CKR01] Gruia Călinescu, Howard Karloff, and Yuval Rabani. Approximation algorithms for the 0-extension problem. In *Proceedings of the twelfth annual ACM-SIAM symposium on Discrete algorithms*, pages 8–16. ACM Press, 2001.

[CL98] Artur Czumaj and Andrzej Lingas. A polynomial time approximation scheme for Euclidean minimum cost k-connectivity. In *Automata, languages and programming (Aalborg, 1998)*, volume 1443 of *Lecture Notes in Comput. Sci.*, pages 682–694. Springer, Berlin, 1998.

[CL00] Artur Czumaj and Andrzej Lingas. Fast approximation schemes for Euclidean multi-connectivity problems (extended abstract). In *Automata, languages and programming (Geneva, 2000)*, volume 1853 of *Lecture Notes in Comput. Sci.*, pages 856–868. Springer, Berlin, 2000.

[Cla99] K. L. Clarkson. Nearest neighbor queries in metric spaces. *Discrete Comput. Geom.*, 22(1):63–93, 1999.

[Cla06] K. L. Clarkson. Nearest-neighbor searching and metric space dimensions. (survey). In *G. Shakhnarovich, T. Darrell, and P. Indyk, editors, Nearest-Neighbor Methods for Learning and Vision: Theory and Practice. MIT Press*, 11–15 2006.

[CLZ02] Artur Czumaj, Andrzej Lingas, and Hairong Zhao. Polynomial-time approximation schemes for the Euclidean survivable network design problem. In *Automata, languages and programming*, volume 2380 of *Lecture Notes in Comput. Sci.*, pages 973–984. Springer, Berlin, 2002.

[DCKM04] F. Dabek, R. Cox, F. Kaashoek, and R. Morris. Vivaldi: A decentralized network coordinate system. In *Proceedings of the ACM SIGCOMM*, 2004.

[DG03] Sanjoy Dasgupta and Anupam Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures Algorithms*, 22(1):60–65, 2003.

[DL97] Michel Marie Deza and Monique Laurent. *Geometry of cuts and metrics*, volume 15 of *Algorithms and Combinatorics*. Springer-Verlag, Berlin, 1997.

[EIO02] Lars Engebretsen, Piotr Indyk, and R. O'Donnell. Derandomized dimensionality reduction with applications. In *Proceedings of the 13th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 715–712, 2002.

[FK94] Christos Faloutsos and Ibrahim Kamel. Beyond uniformity and independence: Analysis of r-trees using the concept of fractal dimension. In *Proceedings of the Thirteenth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)*, pages 4–13, 1994.

[FM88] Péter Frankl and Hiroshi Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *J. Combin. Theory Ser. B*, 44(3):355–362, 1988.

[FRT04] Jittat Fakcharoenphol, Satish Rao, and Kunal Talwar. A tight bound on approximating arbitrary metrics by tree metrics. *J. Comput. System Sci.*, 69(3):485–497, 2004.

[FT03] Jittat Fakcharoenphol and Kunal Talwar. An improved decomposition theorem for graphs excluding a fixed minor. *RANDOM-APPROX*, pages 36–46, 2003.

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low–distortion embeddings. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, pages 534–543, 2003.

[GP83] P. Grassberger and I. Procaccia. Measuring the strangeness of strange attractors. In *Physica 9D*, 1983.

[Gri00] Michelangelo Grigni. Approximate tsp in graphs with forbidden minors. In *Proceedings of the 27st International Colloquium on Automata, Languages and Programming (ICALP)*, pages 869–877, 2000.

[GT87] J. R. Gilbert and R. E. Tarjan. The analysis of a nested dissection algorithm. In *Numerische Mathematik, Volume 50 , Issue 4*, 1987.

[Has96] J. Hastad. Clique is hard to approximate within $n^{1-\epsilon}$. In *Proceedings of the 37th Symposium on the Foundations of Computer Science (FOCS)*, pages 627–636, 1996.

[HKRZ02] K. Hildrum, J. D. Kubiatowicz, S. Rao, and B. Y. Zhao. Distributed object location in a dynamic network. In *Proceedings of the Fourteenth ACM Symposium on Parallel Algorithms and Architectures*, pages 41–52, August 2002.

[HPM05] Sariel Har-Peled and Manor Mendel. Fast constructions of nets in low dimensional metrics, and their applications. In *Proceedings of the twenty-first annual symposium on Computational geometry*, pages 150–158, 2005.

[IA04] Dahlia Malkhi Ittai Abraham. Compact routing on euclidian metrics. In *PODC*, pages 141–149, 2004.

[IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *Proceedings of the 30th ACM Symposium on the Theory of Computing (STOC)*, pages 604–613, 1998.

[IM04] Piotr Indyk and Jiri Matoušek. Low-distortion embeddings of finite metric spaces. In Jacob E. Goodman and Joseph O'Rourke, editors, *Handbook of Discrete and Computational Geometry*, Discrete Mathematics and its Applications (Boca Raton), chapter 8, pages xviii+1539. Chapman & Hall/CRC, Boca Raton, FL, second edition, 2004.

[IN] Piotr Indyk and Assaf Naor. Nearest neighbor preserving embeddings. In *ACM Transactions on Algorithms (To appear)*.

[Ind01] Piotr Indyk. Algorithmic aspects of geometric embeddings. In *Proceedings of the 42nd Symposium on the Foundations of Computer Science (FOCS)*, pages 10–33, 2001.

[JL84] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[KK77] Leonard Kleinrock and Farouk Kamoun. Hierarchical routing for large networks. Performance evaluation and optimization. *Comput. Networks*, 1(3):155–174, 1976/77.

[KL03] Robert Krauthgamer and James R. Lee. The intrinsic dimensionality of graphs. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 438–447. ACM Press, 2003.

[KL04] Robert Krauthgamer and James R. Lee. Navigating nets: simple algorithms for proximity search. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 798–807. Society for Industrial and Applied Mathematics, 2004.

[KL05] Robert Krauthgamer and James R. Lee. The black-box complexity of nearest-neighbor search. *Theoret. Comput. Sci.*, 348(2-3):262–276, 2005.

[KL06] Robert Krauthgamer and James R. Lee. Algorithms on negatively curved spaces. In *Proceedings of the 47th Symposium on the Foundations of Computer Science (FOCS)*, page to appear, 2006.

[Kle05] Philip N. Klein. A linear-time approximation scheme for planar weighted tsp. In *Proceedings of the 46th Symposium on the Foundations of Computer Science (FOCS)*, pages 647–657, 2005.

[KLMN05] R. Krauthgamer, J. R. Lee, M. Mendel, and A. Naor. Measured descent: a new embedding method for finite metrics. *Geom. Funct. Anal.*, 15(4):839–858, 2005.

[KMW05] F. Kuhn, T. Moscibroda, and R. Wattenhofer. On the locality of bounded growth. In *In Proc. of the 24th Annual Symposium on Principles of Distributed Computing (PODC)*, 2005.

[KPR93] Philip Klein, Serge A. Plotkin, and Satish B. Rao. Excluded minors, network decomposition, and multicommodity flow. In *Proceedings of the 25th ACM Symposium on the Theory of Computing (STOC)*, pages 682–690, 1993.

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

[KR02] David R. Karger and Matthias Ruhl. Finding nearest neighbors in growth-restricted metrics. In *Proceedings of the 34th Annual ACM Symposium on the Theory of Computing*, pages 63–66, 2002.

[KRX06] Goran Konjevod, Andréa W. Richa, and Donglin Xia. Optimal-stretch name-independent compact routing in doubling metrics. In *The twenty-fifth annual ACM symposium on Principles of distributed computing*, 2006.

[KRX07] Goran Konjevod, Andra W. Richa, and Donglin Xia. Optimal scale-free compact routing schemes in doubling networks. In *Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2007.

[KSW04] Jon M. Kleinberg, Aleksandrs Slivkins, and Tom Wexler. Triangulation and embedding using small sets of beacons. In *45th FOCS*, 2004.

[Lar67] D. G. Larman. A new theory of dimension. In *Proc. London Math. Soc., 17*, 1967.

[LGS07] Jonathan Ledlie, Paul Gardner, and Margo Seltzer. Network coordinates in the wild. In *Proceedings of the4th USENIX Symposium on Networked Systems Design & Implementation*, 2007.

[LLR95] Nathan Linial, Eran London, and Yuri Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995. (Preliminary version in *35th FOCS*, 1994).

[LN03] James R. Lee and Assaf Naor. Embedding the diamond graph in L_p and dimension reduction in L_1. In *Geometric and Functional Analysis (GAFA)*. Springer Verlag, 2003.

[LP01] U. Lang and C. Plaut. Bilipschitz embeddings of metric spaces into space forms. In *Geom. Dedicata, 87(1-3)*, pages 285–307, 2001.

[LS93] Nathan Linial and Michael Saks. Low diameter graph decompositions. *Combinatorica*, 13(4):441–454, 1993. (Preliminary version in *2nd SODA*, 1991).

[Mat90] Jiří Matoušek. Bi-Lipschitz embeddings into low dimensional Euclidean spaces. *Commentationes Mathematicae Universitatis Carolinae*, 31(3):589–600, 1990.

[Mat97] Jiří Matoušek. On embedding expanders into l_p spaces. *Israel J. Math.*, 102:189–197, 1997.

[Mat02] Jiří Matoušek. *Lectures on discrete geometry*, volume 212 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 2002.

[Mit99] Joseph 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(4):1298–1309 (electronic), 1999.

[MR98] Michael Molloy and Bruce Reed. Further algorithmic aspects of the local lemma. In *Proceedings of the30thAnnual ACM Symposium on Theory of Computing*, 1998.

[NS07] Giri Narasimhan and Michiel Smid. *Geometric Spanner Networks*. Cambridge University Press, 2007.

[NZ02] T. Ng and H. Zhang. Predicting Internet network distance with coordinates-based approaches. In *Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM)*, 2002.

[OS81] Haruko Okamura and Paul D. Seymour. Multicommodity flows in planar graphs. *Journal of Combinatorial Theory Series B*, 31(1):75–81, 1981.

[PB94] Charles E. Perkins and Pravin Bhagwat. Highly dynamic destination-sequenced distance-vector routing (dsdv) for mobile computers. In *SIGCOMM*, pages 234–244, 1994.

[PBR99] Charles E. Perkins and Elizabeth M. Belding-Royer. Ad-hoc on-demand distance vector routing. In *WMCSA*, pages 90–100, 1999.

[PC97] Vincent D. Park and M. Scott Corson. A highly adaptive distributed routing algorithm for mobile wireless networks. In *INFOCOM*, pages 1405–1413, 1997.

[PKF00] Bernd-Uwe Pagel, Flip Korn, and Christos Faloutsos. Deflating the dimensionality curse using multiple fractal dimensions. In *ICDE*, pages 589–598, 2000.

[PRR99] C. G. Plaxton, R. Rajaraman, and A. W. Richa. Accessing nearby copies of replicated objects in a distributed environment. *Theory Comput. Syst.*, 32(3):241–280, 1999. ACM Symposium on Parallel Algorithms and Architectures (Padua, 1996).

[PS89] David Peleg and Alejandro A. Schäffer. Graph spanners. *J. Graph Theory*, 13(1):99–116, 1989.

[PY93] Christos H. Papadimitriou and Mihalis Yannakakis. The traveling salesman problem with distances one and two. In *Mathematics of Operations Research, Volume 18, Issue 1*, 1993.

[Rao99] Satish B. Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *15th Annual ACM Symposium on Computational Geometry*, pages 300–306, 1999.

[RS99] Satish B. Rao and Warren D. Smith. Approximating geometrical graphs via "spanners" and "banyans". In *STOC '98 (Dallas, TX)*, pages 540–550. ACM, New York, 1999.

[Sal91] J.S. Salowe. Constructing multidimensional spanner graphs. In *Internat. J. Comput. Geom. Appl. 1*, pages 99–107, 1991.

[Siv02] D. Sivakumar. Algorithmic derandomization via complexity theory. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 619–626. ACM Press, 2002.

[Sli05] Aleksandrs Slivkins. Distance estimation and object location via rings of neighbors. In *The 24th Annual Symposium on Principles of Distributed Computing (PODC)*, 2005.

[Tal04] Kunal Talwar. Bypassing the embedding: Algorithms for low-dimensional metrics. In *Proceedings of the 36th ACM Symposium on the Theory of Computing (STOC)*, pages 281–290, 2004.

[Tar75] R. E. Tarjan. Efficiency of a good but not linear set union algorithm. *Journal of ACM 22*, pages 215–225, 1975.

[Tre00] Luca Trevisan. When hamming meets euclid: The approximability of geometric tsp and steiner tree. In *SIAM J. Comput. 30(2)*, 2000.

[Vai91] P.M. Vaidya. A sparse graph almost as good as the complete graph on points in k dimensions. In *Discrete Comput. Geom. 6*, pages 369–381, 1991.

[XKR06] D. Xia, G. Konjevod, and A. Richa. On sampling in higher-dimensional peer-to-peer systems. In *Proceedings of Latin American Theoretical Informatics (LATIN)*, 2006.

[Yao82] Andrew C. Yao. Space-time tradeoff for answering range queries. In *Proceedings of the 14th Annual ACM Symposium on Theory of Computing*, pages 128–136, 1982.