Metric Embedding via Shortest Path Decompositions*

Ittai Abraham[1], Arnold Filtser[2], Anupam Gupta[3], and Ofer Neiman[2]

[1]VMWare. Email: iabraham@vmware.com
[2]Ben-Gurion University of the Negev. Email: {arnoldf,neimano}@cs.bgu.ac.il
[3]Carnegie Mellon University. Email: anupamg@cs.cmu.edu

March 13, 2018

Abstract

We study the problem of embedding weighted graphs of pathwidth k into ℓ_p spaces. Our main result is an $O(k^{\min\{1/p,1/2\}})$-distortion embedding. For $p = 1$, this is a super-exponential improvement over the best previous bound of Lee and Sidiropoulos. Our distortion bound is asymptotically tight for any fixed $p > 1$.

Our result is obtained via a novel embedding technique that is based on low depth decompositions of a graph via shortest paths. The core new idea is that given a geodesic shortest path P, we can probabilistically embed all points into 2 dimensions with respect to P. For $p > 2$ our embedding also implies improved distortion on bounded treewidth graphs $(O((k \log n)^{1/p}))$. For asymptotically large p, our results also implies improved distortion on graphs excluding a minor.

*A full version of the paper can be found at https://www.cs.bgu.ac.il/~arnoldf/PathPaperFull.pdf

1 Introduction

Low-distortion metric embeddings are a crucial component in the modern algorithmist toolkit. Indeed, they have applications in approximation algorithms [LLR95], online algorithms [BBMN11], distributed algorithms [KKM+12], and for solving linear systems and computing graph sparsifiers [ST04]. Given a (finite) metric space (V, d), a map $\phi : V \to \mathbb{R}^D$, and a norm $\|\cdot\|$, the *contraction* and *expansion* of the map ϕ are the smallest $\tau, \rho \geq 1$, respectively, such that for every pair $x, y \in V$,

$$\frac{1}{\tau} \leq \frac{\|\phi(x) - \phi(y)\|}{d(x,y)} \leq \rho \quad .$$

The *distortion* of the map is then $\tau \cdot \rho$ In this paper we will investigate embeddings into ℓ_p norms; the most prominent of which are the Euclidean norm ℓ_2 and the cut norm ℓ_1; the former for obvious reasons, and the latter because of its close connection to graph partitioning problems, and in particular the Sparsest Cut problem. Specifically, the ratio between the Sparsest Cut and the multicommodity flow equals the distortion of the optimal embedding into ℓ_1 (see [LLR95, GNRS04] for more details).

We focus on embedding of metrics arising from certain graph families. Indeed, since general n-point metrics require $\Omega(\log n/p)$-distortion to embed into ℓ_p-norms, much attention was given to embeddings of restricted graph families that arise in practice. (Embedding an (edge-weighted) graph is short-hand for embedding the shortest path metric of the graph generated by these edge-weights.) Since the class of graphs embeddable with some distortion into some target normed space is closed under taking minors, it is natural to focus on minor-closed graph families. A long-standing conjecture in this area is that all non-trivial minor-closed families of graphs embed into ℓ_1 with distortion depending only on the graph family and not the size n of the graph.

While this question remains unresolved in general, there has been some progress on special classes of graphs. The class of outerplanar graphs (which exclude $K_{2,3}$ and K_4 as a minor) embeds isometrically into ℓ_1; this follows from results of Okamura and Seymour [OS81]. Following [GNRS04], Chakrabarti et al. [CJLV08] show that every graph with treewidth-2 (which excludes K_4 as a minor) embeds into ℓ_1 with distortion 2 (which is tight, as shown by [LR10]). Lee and Sidiropoulos [LS13] showed that every graph with pathwidth k can be embedded into ℓ_1 with distortion $(4k)^{k^3+1}$. See Section 1.3 for additional results.

We note that ℓ_2 is a potentially more natural and useful target space than ℓ_1 (in particular, finite subsets of ℓ_2 embed isometrically into ℓ_1). Alas, there are only few (natural) families of metrics that admit constant distortion embedding into Euclidean space, such as "snowflakes" of a doubling metrics [Ass83], doubling trees [GKL03] and graphs of bounded bandwidth [BCMN13]. All these families have bounded doubling dimension. (For definitions, see Section 2.)

1.1 Our Results

In this paper we develop a new technique for embedding certain graph families into ℓ_p spaces with constant distortion. Our main result is an improved embedding for bounded pathwidth graphs (see Section 2 for the definition of pathwidth and other graph families).

Theorem 1 (Pathwidth Theorem)**.** *Any graph with pathwidth k embeds into ℓ_p with distortion $O(k^{1/p})$.*

Note that this is a super-exponential improvement over the best previous distortion bound of $O(k)^{k^3}$, by Lee and Sidiropoulos. Their approach was based on probabilistic embedding into trees, which implies embedding only into ℓ_1. Such an approach cannot yield distortion better than $O(k)$, due to known lower bounds for the diamond graph [GNRS04], that has pathwidth $k+1$. Our embedding holds for any ℓ_p space, and we can overcome the barrier of $O(k)$ (since finite subsets of ℓ_2 embed isometrically into ℓ_p, the distortion of Theorem 1 is never larger than $O(\sqrt{k})$). In particular, we obtain embeddings of pathwidth-k graphs into both ℓ_2 and ℓ_1 with distortion $O(\sqrt{k})$. Moreover, an embedding with this distortion can be found efficiently via semidefinite-programming; see, e.g., [LLR95], even without access to the actual path decomposition (which is NP-hard even to approximate [BGHK92]). We remark that graphs of bounded pathwidth can have arbitrarily large doubling dimension (exhibited by star graphs that have pathwidth 1), and thus our result is a noteworthy example of a non-trivial Euclidean embedding with constant distortion for a family of metrics with unbounded doubling dimension.

Since graphs of treewidth k have pathwidth $O(k \log n)$ (see, e.g., [GTW13]), Theorem 1 provides an embedding of such graphs into ℓ_p with distortion $O((k \log n)^{1/p})$. This strictly improves the best previously known bound, which follows from a theorem in [KLMN05] (who obtained distortion $O(k^{1-1/p} \log^{1/p} n)$), for any $p > 2$, and matches it for $1 \leq p \leq 2$. While [KK16] obtained recently a distortion bound with improved dependence on k, their result $O((\log(k \log n))^{1-1/p}(\log^{1/p} n))$ has sub-optimal dependence on n.

A General Embedding Framework. The embedding of Theorem 1 follows as a special case of a more general theorem: we devise embeddings for any graph family which admits "shortest path decompositions" (SPDs) of "low depth". Every (weighted) path graph has an SPD of depth 1. A graph G has an SPD of depth k if after removing some *shortest path* P, every connected component in $G \setminus P$ has an SPD of depth $k-1$. (An alternative definition appears in Definition 1.) Our main technical result is the following.

Theorem 2 (Embeddings for SPD Families)**.** *Let $G = (V, E)$ be a weighted graph with an SPD of depth k. Then there exists an embedding $f : V \to \ell_p$ with distortion $O(k^{1/p})$.*

Since bounded-pathwidth graphs admit SPDs of low depth, we get Theorem 1 as a simple corollary of Theorem 2. Moreover, we derive several other results, which are summarized in Figure 1.1; these results either improve on the state-of-the-art, or provide matching bounds using a new approach.

Graph Family	**Our results.**	**Previous results**	
Pathwidth k	$O(k^{1/p})$	$(4k)^{k^3+1}$ into ℓ_1	[LS13]
Treewidth k	$O((k \log n)^{1/p})$	$O(k^{1-1/p} \cdot \log^{1/p} n)$	[KLMN05]
		$O((\log(k \log n))^{1-1/p}(\log^{1/p} n))$	[KK16]
Planar	$O(\log^{1/p} n)$	$O(\log^{1/p} n)$	[Rao99]
H-minor-free	$O((g(H) \log n)^{1/p})$	$O(\lvert H\rvert^{1-1/p} \log^{1/p} n)$	[AGG+14]+[KLMN05]

Figure 1: *Our and previous results for embedding certain graph families into ℓ_p. (For H-minor-free graphs, $g(H)$ is some function of $|H|$.)*

Our result of Theorem 1 (and thus also Theorem 2) is asymptotically tight for any fixed $p > 1$. The family exhibiting this fact is the diamond graphs.

Theorem 3 ([NR03, LN04, MN13])**.** *For any fixed $p > 1$ and every $k \geq 1$, there exists a graph $G = (V, E)$ with pathwidth-k, such that every embedding $f : V \to \ell_p$ has distortion $\Omega(k^{\min\{1/p, 1/2\}})$.*

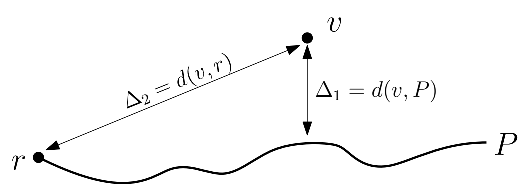
The bound in Theorem 3 was proven first for $p = 2$ in [NR03], generalized to $1 < p \leq 2$ in [LN04] and for $p \geq 2$ by [MN13] (see also [JS09, JLM11]). The proofs of [NR03, LN04] were done using the diamond graph, while [MN13] used the Laakso graph. For completeness, we provide a proof of the case $p \geq 2$ using the diamond graph in Appendix C. We note that for ℓ_1 only the trivial $\Omega(\log k)$ lower bound is known.

1.2 Technical Ideas

Many known embeddings [Bou85, Rao99, KLMN05, ABN11] are based on a collection of 1-dimensional embeddings, where we embed each point to its distance from a given subset of points. We follow this approach, but differ in two aspects. Firstly, the subset of points we use is not based on random sampling or probabilistic clustering. Rather, inspired by the works of [And86] and [AGG+14], the subset used is a geodesic shortest path. The second is that our embedding is not 1-dimensional but 2-dimensional: this seemingly small change crucially allows us to use the structure of the shortest paths to our advantage.

The SPD induces a collection of shortest paths (each shortest path lies in some connected component). A natural initial attempt is to embed a vertex v relative to a geodesic path P using two dimensions:[1]

- The *first* coordinate Δ_1 is the distance to the path $d(v, P)$.
- The *second* coordinate Δ_2 is the distance $d(v, r)$ to the endpoint of the path, called its "root".



Unfortunately, this embedding may have unbounded expansion: If two vertices u, v are separated by some shortest path, in future iterations v may have a large distance to the root of a path P in its component, while u has zero in that coordinate (because it's not in that component), incurring a large stretch. The natural fix is to enforce a Lipschitz condition on every coordinate: for v in cluster X, we *truncate* the value v can receive at $O(d_G(v, V \setminus X))$. I.e., a vertex close to the boundary of X cannot get a large value. Using the fact that the SPD has depth k, each vertex will have only $O(k)$ nonzero coordinates, which implies expansion $O(k^{1/p})$.

To bound the contraction, for each pair u, v we consider the *first* path P in the SPD that lies "close" to $\{u, v\}$ or separates them to different connected components. Then we show that at least one of the two coordinates should give sufficient contribution.

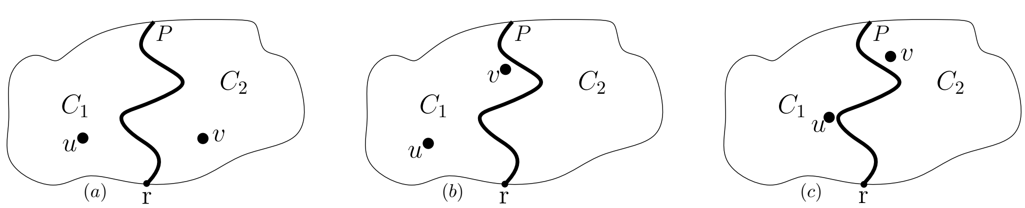


Figure 2: *A shortest path P, rooted at r, partitions cluster X into clusters C_1 and C_2. In cases (a) and (b), the first coordinate (the distance to P) provides sufficient contribution. In case (c) the second coordinate (the distance to root r) provides the contribution.*

[1] In fact, we use different dimensions for each connected components.

But what about the effect of truncation on contraction? A careful recursive argument shows that the contribution to u, v from the first coordinate (the distance from the path P) is essentially not affected by this truncation. Hence the argument in cases (a) and (b) of Figure 2 still works. However, the argument using the distance to the root of P, case (c), can be ruined. Solving this issue requires some new non-trivial ideas. Our solution is to introduce a probabilistic sawtooth function that replaces the simple truncation. The main technical part of the paper is devoted to showing that a collection of these functions for all possible distance scales, with appropriate random shifts, suffices to control the expected contraction in case (c), for *all relevant pairs simultaneously.*

1.3 Other Related Work

There has been work on embedding several other graph families into normed spaces: Chekuri et al. [CGN+06] extend the Okamura and Seymour bound for outerplanar graphs to k-outerplanar graphs, and showed that these embed into ℓ_1 with distortion $2^{O(k)}$. Rao [Rao99] (see also [KLMN05]) embed planar graphs into ℓ_p with distortion $O(\log^{1/p} n)$. For graphs with genus g, [LS10] showed an embedding into Euclidean space with distortion $O(\log g + \sqrt{\log n})$. Finally, for H-minor-free graphs, combining the results of [AGG+14, KLMN05] give ℓ_p-embeddings with $O(|H|^{1-1/p} \log^{1/p} n)$ distortion.

Following [And86, Mil86], the idea of using geodesic shortest paths to decompose the graph has been used for many algorithmic tasks: MPLS routing [GKR04], directed connectivity, distance labels and compact routing [Tho04], object location [AG06], and nearest neighbor search [ACKW15]. However, to the best of our knowledge, this is the first time it has been used directly for low-distortion embeddings into normed spaces.

2 Preliminaries and Notation

For $k \in \mathbb{Z}$, let $[k] := \{1, \ldots, k\}$. For $p \geq 1$, the ℓ_p-norm of a vector $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$ is $\|x\|_p := (\sum_{i=1}^d |x_i|^p)^{1/p}$, where $\|x\|_\infty := \max_i |x_i|$.

Doubling dimension. The doubling dimension of a metric is a measure of its local "growth rate". Formally, a metric space (X, d) has *doubling dimension* λ_X if for every $x \in X$ and radius r, the ball $B(x, r)$ can be covered by 2^{λ_X} balls of radius $\frac{r}{2}$. A family is *doubling* if the doubling dimension of all metrics in it is bounded by some universal constant.

Graphs. We consider connected undirected graphs $G = (V, E)$ with edge weights $w : E \to \mathbb{R}_{>0}$. Let d_G denote the shortest path metric in G; we drop subscripts when there is no ambiguity. For a vertex $x \in V$ and a set $A \subseteq V$, let $d_G(x, A) := \min_{a \in A} d(x, a)$, where $d_G(x, \emptyset) := \infty$. For a subset of vertices $A \subseteq V$, let $G[A]$ denote the induced graph on A, and let $d_A := d_{G[A]}$ be the shortest path metric in the induced graph. Let $G \setminus A := G[V \setminus A]$ be the graph after deleting the vertex set A from G.

Special graph families. Given a graph $G = (V, E)$, a *tree decomposition* of G is a tree T with nodes $B_1, \ldots, B_s$ (called *bags*) where each B_i is a subset of V such that the following properties hold:

- For every edge $\{u, v\} \in E$, there is a bag B_i containing both u and v.
- For every vertex $v \in V$, the set of bags containing v form a connected subtree of T.

The *width* of a tree decomposition is $\max_i\{|B_i| - 1\}$. The *treewidth* of G is the minimal width of a tree decomposition of G.

A *path decomposition* of G is a special kind of tree decomposition where the underlying tree is a path. The *pathwidth* of G is the minimal width of a path decomposition of G.

A graph H is a *minor* of a graph G if we can obtain H from G by edge deletions/contractions, and vertex deletions. A graph family $\mathcal{G}$ is *H-minor-free* if no graph $G \in \mathcal{G}$ has H as a minor.

2.1 The Sawtooth function

Another component in our embeddings will be the following sawtooth function. For $t \in \mathbb{N}$, we define $g_t : \mathbb{R}_+ \to \mathbb{R}$ the *sawtooth* function w.r.t. 2^t as follows. For $x \geq 0$, if $q_x := \lfloor x/2^{t+1} \rfloor$ then

$$g_t(x) = 2^t - \left| x - \left(q_x \cdot 2^{t+1} + 2^t \right) \right| .$$

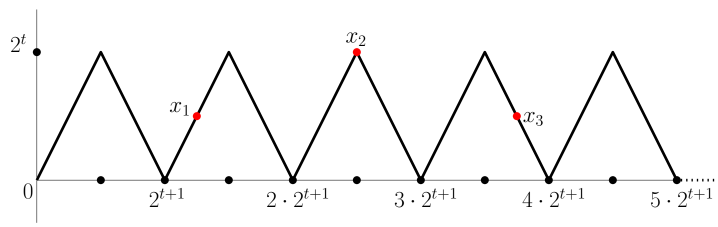


Figure 3: *The graph of the "sawtooth" function g_t. The points $x_1 = 5 \cdot 2^{t-1}$ and $x_3 = 15 \cdot 2^{t-1}$ are mapped to 2^{t-1}, while $x_2 = 10 \cdot 2^{t-1}$ is mapped to 2^t.*

The following observation is straightforward.

Observation 1. *The Sawtooth function g_t is 1-Lipschitz, bounded by 2^t, periodic with period 2^{t+1}.*

The proof of the following lemma appears in Appendix E.

Lemma 2 (Sawtooth Lemma)**.** *Let $x, y \in \mathbb{R}_+$. Let $\alpha \in [0, 1]$, $\beta \in [0, 4]$ be drawn uniformly and independently. The following properties hold:*

1. $\mathbb{E}_{\alpha,\beta} \left[g_t(\beta x + \alpha \cdot 2^{t+1}) \right] = 2^{t-1}$.
2. *If $|x - y| \leq 2^{t-1}$, then* $\mathbb{E}_{\alpha,\beta} \left[\left| g_t(\beta x + \alpha \cdot 2^{t+1}) - g_t(\beta y + \alpha \cdot 2^{t+1}) \right| \right] = \Omega(|x - y|)$.
3. *If $|x - y| > 2^{t-1}$, then* $\mathbb{E}_{\alpha,\beta} \left[\left| g_t(\beta x + \alpha \cdot 2^{t+1}) - g_t(\beta y + \alpha \cdot 2^{t+1}) \right| \right] = \Omega(2^t)$.

3 Shortest Path Decompositions

Our embeddings will crucially depend on the notion of shortest path decompositions. In the introduction we provided a recursive definition for SPD. Here we show an equivalent definition which will be more suitable for our purposes.

Definition 1 (Shortest Path Decomposition (SPD))**.** *Given a weighted graph $G = (V, E, w)$, a SPD of depth k is a pair $\{\mathcal{X}, \mathcal{P}\}$, where $\mathcal{X}$ is a collection $\mathcal{X}_1, \ldots, \mathcal{X}_k$ of partial partitions of V [2], and $\mathcal{P}$ is a collection of sets of paths $\mathcal{P}_1, \ldots, \mathcal{P}_k$, where $\mathcal{X}_1 = \{V\}$, $\mathcal{X}_k = \mathcal{P}_k$, and the following properties hold:*

1. *For every $1 \leq i \leq k$ and every subset $X \in \mathcal{X}_i$, there exist a unique path $P_X \in \mathcal{P}_i$ such that P_X is a shortest path in $G[X]$.*

[2] i.e. for every $X \in \mathcal{X}_i$, $X \subseteq V$, and for every different subsets $X, X' \in \mathcal{X}_i$, $X \cap X' = \emptyset$.

2. *For every $2 \le i \le k$, $\mathcal{X}_i$ consists of all connected components of $G[X \setminus P_X]$ over all $X \in \mathcal{X}_{i-1}$.*

In other words, $\bigcup_{i=1}^{k} \mathcal{P}_k$ is a partition of V into paths, where each path P_X is a shortest path in the component X it belongs to at the point it is deleted.

For a given graph G let $\mathsf{SPDdepth}(G)$ be the minimum k such that G admits an SPD of depth k. For a given family of graphs $\mathcal{G}$ let $\mathsf{SPDdepth}(\mathcal{G}) := \max_{G \in \mathcal{G}}\{\mathsf{SPDdepth}(G)\}$. In the following we consider the $\mathsf{SPDdepth}$ of some graph families.

3.1 The SPD Depth for Various Graph Families

One advantage of defining the shortest path decomposition is that several well-known graph families have bounded depth SPD.

- *Pathwidth.* Every graph $G = (V, E, w)$ with *pathwidth* k has an $\mathsf{SPDdepth}$ of $k + 1$. Indeed, let $\mathcal{T} = \langle \mathcal{B}_1, \dots, \mathcal{B}_s \rangle$ be a path decomposition of G, where $\mathcal{B}_1, \mathcal{B}_s$ are the two bags at the end of this path. Choose arbitrary vertices $x \in \mathcal{B}_1$ and $y \in \mathcal{B}_s$, and let P be a shortest path in G from x to y. By the definition of a path decomposition, the path P contains at least one vertex from every bag $\mathcal{B}_i$. Hence, deleting the vertices of P would reduce the size of each bag by one; consequently each connected component of $G \setminus P$ has pathwidth $k - 1$, and by induction $\mathsf{SPDdepth}$ k. Finally, a connected component of pathwidth 0 is necessarily a singleton, which has $\mathsf{SPDdepth}$ 1.

- *Treewidth.* Since every tree has pathwidth $O(\log n)$, we can show that an n-vertex treewidth-k graph has pathwidth $O(k \log n)$. Hence, treewidth-k graphs have $\mathsf{SPDdepth}$ $O(k \log n)$.

- *Planar.* Using cycle separators [Mil86] as in [Tho04, GKR04], every planar graph has $\mathsf{SPDdepth}$ $O(\log n)$; this follows as each cycle separator can be constructed as union of two shortest paths.

- *Minor-free.* Finally, every H-minor-free graph admits a balanced separator consisting of $g(H)$ shortest paths (for some function g) [AG06], and hence has an $\mathsf{SPDdepth}$ $O(g(H) \cdot \log n)$.

Combining these observation with Theorem 2, we get the following set of results:

Corollary 1. *Consider an n-vertex weighted graph G, Theorem 2 implies the following:*

- *If G has pathwidth k, it embeds into ℓ_p with distortion $O(k^{1/p})$.*
- *If G has treewidth k, it embeds into ℓ_p with distortion $O((k \log n)^{1/p})$.*
- *If G is planar, it embeds into ℓ_p with distortion $O(\log^{1/p} n)$.*
- *For every fixed H, if G excludes H as a minor, it embeds into ℓ_p with distortion $O(\log^{1/p} n)$, where the constant in the big-O depends on H.*

As mentioned in Section 1, we get a substantial improvement for the pathwidth case. Our result for treewidth improves upon that from [KLMN05] for $p > 2$; they got $O(k^{1-1/p}(\log n)^{1/p})$ distortion compared to our $O((k \log n)^{1/p})$. Our result appears to be closer to the truth, since the distortion tends to $O(1)$ as $p \to \infty$.

Finally, our results for planar graphs match the current state-of-the-art.

Our results for minor-free graphs depend on the Robertson-Seymour decomposition, and hence are currently better only for large values of p. (It remains an open question to improve the $\mathsf{SPDdepth}$ of H-minor-free graphs to have a $\text{poly}(|H|) \log n$ dependence, perhaps using the ideas from [AGG+14].) In general, we hope that our results will be useful in getting other embedding results, and will spur further work on understanding shortest path separators.

4 The Embedding Algorithm

Let $G = (V, E)$ be a weighted graph, and let $\{\mathcal{X}, \mathcal{P}\} = \{\{\mathcal{X}_1, , \dots, \mathcal{X}_k\}, \{\mathcal{P}_1, \dots, \mathcal{P}_k\}\}$ be an SPD of depth k for G. By scaling, we can assume that the minimum weight of an edge is 1; let $M \in \mathbb{N}$ be the minimal such that the diameter of G is strictly bounded by 2^M. Pick $\alpha \in [0, 1]$ and $\beta \in [0, 4]$ uniformly and independently.

For every $i \in [k]$, and $X \in \mathcal{X}_i$, we now construct an embedding $f_X : V \to \mathbb{R}^D$ (for some number of dimensions $D \in \mathbb{N}$). This map f_X consists of two parts.

<u>*First coordinate: Distance to the Path.*</u> The first coordinate of the embedding implements the distance to the path P_X, and is denoted by f_X^{path}. Let $X_1, \dots, X_s \in \mathcal{X}_{i+1}$ be the connected components of $G[X \setminus P_X]$ (note that it is also possible that $s = 0$). We use a separate coordinate for each X_j, and hence $f_X^{\text{path}} : V \to \mathbb{R}^s$. Moreover, for $v \in X$ we truncate at $2\, d_G(v, V \setminus X)$ in order to guarantee Lipschitz-ness. In particular, the coordinate corresponding to X_j is set to

$$\left(f_X^{\text{path}}\right)_{X_j}(v) = \begin{cases} \min\{d_X(v, P_X), 2d_G(v, V \setminus X)\} & \text{if } v \in X_j, \\ 0 & \text{otherwise.} \end{cases}$$

See Figure 5 (in Appendix A) for an illustration.

<u>*Second coordinate: Distance to the Root.*</u> The second part is denoted f_X^{root}, which is intended to capture the distance from the root r of the path. Again, to get the Lipschitz-ness, we would like to truncate the value at $2\, d_G(v, V \setminus X)$ as we did for f_X^{path}. However, a problem with this idea is that the root r can be arbitrarily far from some pair u, v that needs contribution from this coordinate. And hence, even if $|d_G(u, r) - d_G(v, r)| \approx d_G(u, v)$, there may be no contribution after the truncation. So we use the sawtooth function.

Specifically, we replace the ideal contribution $d_G(v, r)$ by the sawtooth function $g_t(d_G(v, r))$, where the scale t for the function is chosen such that $2^t \approx d_G(v, V \setminus X)$. To avoid the case that two nearby points use two different scales (and hence to guarantee Lipschitz-ness), we take an appropriate linear combination of the two distance scales closest to $2\, d_G(v, V \setminus X)$. Recall that the sawtooth function does not guarantee contribution for u, v due to its periodicity: we may be unlucky and have $g_t(d_G(v, r)) = g_t(d_G(u, r))$ even when $d_G(v, r)$ and $d_G(u, r)$ are very different. To guarantee a large enough contribution for all relevant pairs simultaneously, we add a random shift α, and apply a random "stretch" β to $d_G(v, r)$ before feeding it to g_t. Lemma 2 then shows that many of the choices of α and β give substantially different values for u, v.

Formally, the mapping is as follows. The function f_X^{root} consists of $M + 1$ coordinates, one for each distance scale $t \in \{0, 1, \dots, M\}$. The coordinate corresponding to t is denoted by $f_{X,t}^{\text{root}}$. Let r be an arbitrary endpoint of P_X; we will call r the "root" of P_X. Let $t_v \in \mathbb{N}$ be such that $2\, d_G(v, V \setminus X) \in [2^{t_v}, 2^{t_v+1})$. Set $\lambda_v = \frac{2\, d_G(v, V \setminus X) - 2^{t_v}}{2^{t_v}}$. Note that $0 \le \lambda_v < 1$. For $v \in X$, we define

$$f_{X,t}^{\text{root}}(v) = \begin{cases} \lambda_v \cdot g_t(\beta \cdot d_X(v, r) + \alpha \cdot 2^{t+1}) & \text{if } t = t_v + 1, \\ (1 - \lambda_v) \cdot g_t(\beta \cdot d_X(v, r) + \alpha \cdot 2^{t+1}) & \text{if } t = t_v, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

For all nodes $v \notin X$, we set $f_X^{\text{root}}(v) = \vec{0}$.

Define the map $f_X = f_X^{\text{path}} \oplus f_X^{\text{root}}$, and the final embedding is

$$f = \bigoplus_{i=1}^{k} \bigoplus_{X \in \mathcal{X}_i} f_X,$$

i.e., the concatenation of all the constructed embeddings. Before we start the analysis, let us record some simple observations.

Observation 3. *For the map f defined above, the following hold:*

- *The number of coordinates in f does not depend on α, β.*
- *For every $X \in \mathcal{X}_i$ and $v \notin X$, the map $f_X(v)$ is the constant vector $\vec{0}$.*
- *For every $X \in \mathcal{X}_i$ and $v \in X$, the map f_X is nonzero in at most 3 coordinates.*

Hence, since $\mathcal{X}_i$ is a partial partition of V and the depth of the SPD is k, we get that $f(v)$ is nonzero in at most $3k$ coordinates for each $v \in V$.

5 The Analysis

The main technical lemmas now show that the per-coordinate expansion is constant, and that for every pair, there exists a coordinate for which the *expected* contraction is constant.

Lemma 4 (Expansion Bound)**.** *For any vertices u, v, every coordinate j, and every choice of α, β,*

$$|f_j(v) - f_j(u)| = O(d_G(u, v)).$$

Lemma 5 (Contraction Bound)**.** *For any vertices u, v, there exists some coordinate j such that*

$$\mathbb{E}_{\alpha,\beta}\left[|f_j(v) - f_j(u)|\right] = \Omega(d_G(u, v)).$$

Given these two lemmas, we can combine them together to show that the entire embedding has small distortion. (Proof of the composition lemma can be found in Appendix D.)

Lemma 6 (Composition Lemma)**.** *Let (X, d) be a metric space. Suppose that there are constants ρ, τ and a function $f : X \to \mathbb{R}^s$, drawn from some probability space such that:*

1. *For every $u, v \in X$ and every $j \in [s]$, $|f_j(v) - f_j(u)| \leq \rho \cdot d(v, u)$.*
2. *For every $u, v \in X$, there exists $j \in [s]$ such that $\mathbb{E}[\,|f_j(v) - f_j(u)|\,] \geq \frac{1}{\tau} \cdot d(v, u)$.*
3. *For every $v \in X$, there is a subset of indices $I_v \subseteq [s]$ of size $|I_v| \leq k$, such that for every $j \notin I_v$, $f_j(v) = 0$. In other words, for every $v \in X$, $f(v)$ has support of size at most k.*

Then, for every $p \geq 1$, there is an embedding of (X, d) into ℓ_p with distortion $O(k^{1/p})$. Moreover, if there is an efficient algorithm for sampling such an f, then there is a randomized algorithm that constructs the embedding efficiently (in expectation).

Now Theorem 2, our embedding for graphs with low depth SPDs, immediately follows by applying the Composition Lemma (Lemma 6) to Lemma 4, Lemma 5, and Observation 3.

We defer the proof of Lemma 4 to Appendix B, and now give the proof of Lemma 5.

5.1 Bounding the Contraction: Proof of Lemma 5

Recall that we want to prove that for any pair u, v of vertices, the embedding has a large contribution between them. A natural proof idea is to show that vertices u, v would eventually be separated by the recursive procedure. When they are separated, either one of u, v is far from the separating path P, or they both lie close to the path. In the former case, the distance $d(v, P)$ gives a large contribution to the embedding distance, and in the latter case the distance from one end of the path (the "root") gives a large contribution.

However, there's a catch: the value of the v's embedding in any single coordinate cannot be more than v's distance to the boundary, and this causes problems. Indeed, if u, v fall very close to the path P at some step of the algorithm, they must get most of their contribution at this level, since future levels will not give much contribution. How can we do it, without assigning large values? This is where we use the sawtooth function: it gives a good contribution between points without assigning any vertex too large a value in any coordinate.

Formally, to bound the contraction and prove Lemma 5, for nodes u, v we need to show that there exists a coordinate j such that $\mathbb{E}_{\alpha,\beta}[|f_j(v) - f_j(u)|] = \Omega(d_G(u, v))$. For brevity, define

$$\Delta_{uv} := d_G(u, v). \tag{2}$$

Fix $c = 12$. Let i be the minimal index such that there exists $X \in \mathcal{X}_i$ with $u, v \in X$, and at least one of the following holds:

1. $\min\{d_X(v, P_X), d_X(u, P_X)\} \le \Delta_{uv}/c$ (i.e., we choose a path close to $\{v, u\}$).
2. v and u are in different components of $X \setminus P_X$.

Note that such an index i indeed exists: if v and u are separated by the SPD then condition (2) holds. The only other possibility that v and u are never separated is when at least one of them lies on one of the shortest paths. In such a case, surely condition (1) holds. By the minimality of i, for every $X' \in \mathcal{X}_{i'}$ such that $i' < i$ and $u, v \in X'$, necessarily $\min\{d_{X'}(v, P_{X'}), d_{X'}(u, P_{X'})\} > \Delta_{uv}/c$. Therefore, the ball with radius Δ_{uv}/c around each of v, u is contained in X. In particular, $\min\{d_G(v, V \setminus X), d_G(u, V \setminus X)\} > \Delta_{uv}/c$.

Suppose first that (2) occurs but not (1). Let j be the coordinate in f_X^{path} created for the connected component of v in $X \setminus P_X$. Then

$$\left|(f_X^{\text{path}})_j(v) - (f_X^{\text{path}})_j(u)\right| = \min\{d_X(v, P_X), 2d_G(v, V \setminus X)\} - 0 \ge \min\left\{\frac{\Delta_{uv}}{c}, 2\frac{\Delta_{uv}}{c}\right\} = \frac{\Delta_{uv}}{c}.$$

Next assume that (1) occurs. W.l.o.g., $d_X(v, P_X) \le d_X(u, P_X)$, so that $d_X(v, P_X) \le \Delta_{uv}/c$. Suppose first that $d_X(u, P_X) \ge 2\Delta_{uv}/c$. Then in the coordinate j in f_X^{path} created for the connected component of u in $X \setminus P_X$, we have

$$\begin{aligned}\left|(f_X^{\text{path}})_j(v) - (f_X^{\text{path}})_j(u)\right| &\ge |\min\{d_X(u, P_X), 2d_G(u, V \setminus X)\} - \min\{d_X(v, P_X), 2d_G(v, V \setminus X)\}| \\ &\ge \min\left\{2\frac{\Delta_{uv}}{c}, 2\frac{\Delta_{uv}}{c}\right\} - \frac{\Delta_{uv}}{c} = \frac{\Delta_{uv}}{c}.\end{aligned}$$

(It does not matter whether v, u are in the same connected component or not.) Thus it remains to consider the case $d_X(u, P_X) < 2\Delta_{uv}/c$. Let r be the root of P_X. Let v' (resp. u') be the closest vertex on P_X to v (resp. u) in $G[X]$. Then by the triangle inequality

$$d_X(v', u') \geq d_X(v, u) - d_X(v, v') - d_X(u, u') \geq \frac{c-3}{c}\Delta_{uv} \ .$$

In particular,

$$|d_X(v, r) - d_X(u, r)| \geq \left|d_X(v', r) - d_X(u', r)\right| - d_X(v, v') - d_X(u, u') \geq \frac{c-6}{c}\Delta_{uv} = \frac{1}{2}\Delta_{uv} \ , \qquad (3)$$

where we used that P_X is a shortest path in $G[X]$ (implying $|d_X(v', r) - d_X(u', r)| = d_X(v', u')$). See Figure 4 for illustration.

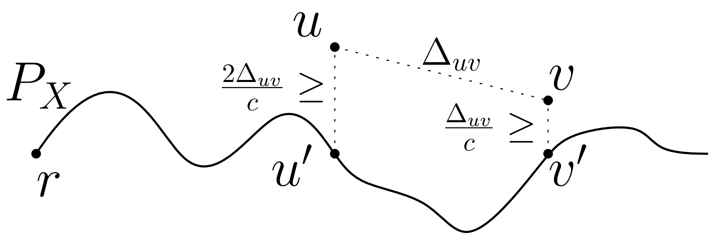


Figure 4: *P_X is a shortest path with root r. v (resp. u) is at distance at most $\frac{\Delta_{uv}}{c}$ (resp. $\frac{2\Delta_{uv}}{c}$) from v' (resp u'), it's closest vertex on P_X. By triangle inequality $d_X(v', u') \geq (1 - \frac{3}{c})\Delta_{uv}$. As u', v' lay on the same shortest path starting at r, $|d_X(v', r) - d_X(u', r)| = d_X(v', u')$. Using the triangle inequality again we conclude $|d_X(v, r) - d_X(u, r)| \geq |d_X(v', r) - d_X(u', r)| - \frac{3}{c}\Delta_{uv} \geq (1 - \frac{6}{c})\Delta_{uv}$.*

Set $x = d_X(v, r)$ and $y = d_X(u, r)$. Assume first that $d_G(v, V \setminus X) \geq d_G(u, V \setminus X)$. In particular, $t_v \geq t_u$. By the definition of t_v, $2d_G(v, V \setminus X) \leq 2^{t_v+1}$. Thus

$$2^{t_v} \geq \frac{\Delta_{uv}}{c} = \Omega(\Delta_{uv}) \qquad (4)$$

Claim 1. *Let $t \geq t_v$, then there is a constant ϕ such that*

$$\mathbb{E}_{\alpha,\beta}\left[\left|g_t(\beta \cdot x + \alpha \cdot 2^{t+1}) - g_t(\beta \cdot y + \alpha \cdot 2^{t+1})\right|\right] \geq \Delta_{uv}/\phi.$$

Proof. If $|x - y| \leq 2^{t-1}$, then using Property 2 of Lemma 2

$$\mathbb{E}_{\alpha,\beta}\left[\left|g_t(\beta \cdot x + \alpha \cdot 2^{t+1}) - g_t(\beta \cdot y + \alpha \cdot 2^{t+1})\right|\right] = \Omega\left(|x - y|\right) \overset{(3)}{=} \Omega(\Delta_{uv}) \ .$$

Otherwise, using Property 3 of Lemma 2

$$\mathbb{E}_{\alpha,\beta}\left[\left|g_t(\beta \cdot x + \alpha \cdot 2^{t+1}) - g_t(\beta \cdot y + \alpha \cdot 2^{t+1})\right|\right] = \Omega\left(2^t\right) \overset{(4)}{\geq} \Omega(\Delta_{uv}) \ . \qquad \square$$

Set $S = \max\left\{8\phi, \frac{8c}{2}\right\}$. We consider two cases:

- If for some $t \in \{0, 1, \ldots, M\}$, $|p_t - q_t| \cdot 2^t > \frac{\Delta_{uv}}{S}$, then

$$\begin{aligned}
\mathbb{E}_{\alpha,\beta}\left[\left|f^{\text{root}}_{X,t}(v) - f^{\text{root}}_{X,t}(u)\right|\right] &= \mathbb{E}_{\alpha,\beta}\left[\left|p_t \cdot g_t(\beta \cdot x + \alpha \cdot 2^{t+1}) - q_t \cdot g_t(\beta \cdot y + \alpha \cdot 2^{t+1})\right|\right] \\
&\geq \left|p_t \cdot \mathbb{E}_{\alpha,\beta}\left[g_t(\beta \cdot x + \alpha \cdot 2^{t+1})\right] - q_t \cdot \mathbb{E}_{\alpha,\beta}\left[g_t(\beta \cdot y + \alpha \cdot 2^{t+1})\right]\right| \\
&= |p_t - q_t| \cdot 2^t = \Omega(\Delta_{uv}) \ .
\end{aligned}$$

  Where the equality follows by Property 1 of Lemma 2.

- The other case is that for every t, $|p_t - q_t| \cdot 2^t \le \frac{\Delta_{uv}}{S}$. Note that $p_{t_v} + p_{t_v+1} = (1-\lambda_v) + \lambda_v = 1$. Let $t \in \{t_v, t_v+1\}$ be such that $p_t \ge \frac{1}{2}$. Using (4), $q_t \ge p_t - \frac{\Delta_{uv}}{2^{t_v} \cdot S} \ge \frac{1}{2} - \frac{2c}{2\Delta_{uv}} \cdot \frac{\Delta_{uv}}{S} \ge \frac{1}{4}$. In particular,

$$\begin{aligned}\mathbb{E}_{\alpha,\beta}\left[\left|f_{X,t}^{\mathrm{root}}(v) - f_{X,t}^{\mathrm{root}}(u)\right|\right] &= \mathbb{E}_{\alpha,\beta}\left[\left|p_t \cdot g_t(\beta \cdot x + \alpha \cdot 2^{t+1}) - q_t \cdot g_t(\beta \cdot y + \alpha \cdot 2^{t+1})\right|\right] \\ &\ge \min\{p_t, q_t\} \cdot \mathbb{E}_{\alpha,\beta}\left[\left|g_t(\beta \cdot x + \alpha \cdot 2^{t+1}) - g_t(\beta \cdot y + \alpha \cdot 2^{t+1})\right|\right] - |p_t - q_t| \cdot 2^t \\ &\ge \frac{1}{4} \cdot \frac{\Delta_{uv}}{\phi} - \frac{\Delta_{uv}}{S} = \Omega(\Delta_{uv}) ,\end{aligned}$$

  where in the first inequality we used Property 1 of Lemma 2, and in the second inequality we used Claim 1.

Finally, recall that we assumed $d_G(v, V \setminus X) \ge d_G(u, V \setminus X)$ for the proof above. The other case ($d_G(v, V \setminus X) < d_G(u, V \setminus X)$) is completely symmetric.

6 Conclusions

In this paper we introduced the notion of shortest path decompositions with low depth. We showed how these can be used to give embeddings into ℓ_p spaces. Our techniques give optimal embeddings of bounded pathwidth graphs into ℓ_2, and also new embeddings for graphs with bounded treewidth, as well as planar and excluded-minor families of graphs. Our embedding into ℓ_p admits a tight lower bound for fixed $p > 1$. We hope that our techniques will be useful for other embedding results.

Our work raises several open questions. While our embeddings are tight for fixed $p > 1$, can we improve the bounds for ℓ_1 embedding of bounded pathwidth graphs? Can we give better results for the $\mathsf{SPDdepth}$ of H-minor-free graphs? Our approach gives a $O(\sqrt{\log n})$-distortion embedding of planar graphs into ℓ_1, which is quite different from the previous known results via padded decompositions: can these ideas be used to make progress towards the planar graph embedding conjecture?

References

[ABN11] Ittai Abraham, Yair Bartal, and Ofer Neiman. Advances in metric embedding theory. *Advances in Mathematics*, 228(6):3026 – 3126, 2011.

[ACKW15] Ittai Abraham, Shiri Chechik, Robert Krauthgamer, and Udi Wieder. Approximate nearest neighbor search in metrics of planar graphs. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM 2015, August 24-26, 2015, Princeton, NJ, USA*, pages 20–42, 2015.

[AG06] Ittai Abraham and Cyril Gavoille. Object location using path separators. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing, PODC 2006, Denver, CO, USA, July 23-26, 2006*, pages 188–197, 2006.

[AGG+14] Ittai Abraham, Cyril Gavoille, Anupam Gupta, Ofer Neiman, and Kunal Talwar. Cops, robbers, and threatening skeletons: padded decomposition for minor-free graphs. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 - June 03, 2014*, pages 79–88, 2014.

[And86] Thomas Andreae. On a pursuit game played on graphs for which a minor is excluded. *Journal of Combinatorial Theory, Series B*, 41(1):37 – 47, 1986.

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

[BBMN11] Nikhil Bansal, Niv Buchbinder, Aleksander Madry, and Joseph Naor. A polylogarithmic-competitive algorithm for the k-server problem. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on*, pages 267 –276, oct. 2011.

[BCMN13] Yair Bartal, Douglas E. Carroll, Adam Meyerson, and Ofer Neiman. Bandwidth and low dimensional embedding. *Theor. Comput. Sci.*, 500:44–56, 2013.

[BGHK92] Hans L. Bodlaender, John R. Gilbert, Hjálmtýr Hafsteinsson, and Ton Kloks. *Approximating treewidth, pathwidth, and minimum elimination tree height*, pages 1–12. Springer Berlin Heidelberg, Berlin, Heidelberg, 1992.

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

[Car04] N. L. Carothers. *A Short Course on Banach Space Theory.* London Mathematical Society Student Texts. Cambridge University Press, 2004.

[CGN+06] Chandra Chekuri, Anupam Gupta, Ilan Newman, Yuri Rabinovich, and Alistair Sinclair. Embedding k-outerplanar graphs into ℓ_1. *SIAM J. Discrete Math.*, 20(1):119–136, 2006.

[CJLV08] Amit Chakrabarti, Alexander Jaffe, James R. Lee, and Justin Vincent. Embeddings of topological graphs: Lossy invariants, linearization, and 2-sums. In *49th Annual IEEE Symposium on Foundations of Computer Science, FOCS 2008, October 25-28, 2008, Philadelphia, PA, USA*, pages 761–770, 2008.

[GKL03] Anupam Gupta, Robert Krauthgamer, and James R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *FOCS '03: Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, page 534, Washington, DC, USA, 2003. IEEE Computer Society.

[GKR04] Anupam Gupta, Amit Kumar, and Rajeev Rastogi. Traveling with a Pez dispenser (or, routing issues in MPLS). *SIAM J. Comput.*, 34(2):453–474, 2004.

[GNRS04] Anupam Gupta, Ilan Newman, Yuri Rabinovich, and Alistair Sinclair. Cuts, trees and ℓ_1s-embeddings of graphs. *Combinatorica*, 24(2):233–269, 2004.

[GTW13] Anupam Gupta, Kunal Talwar, and David Witmer. On the non-uniform sparsest cut problem on bounded treewidth graphs. In *Annual ACM Symposium on Theory of Computing*, pages 281–290, Jun 2013.

[JLM11] Alexander Jaffe, James R. Lee, and Mohammad Moharrami. On the optimality of gluing over scales. *Discrete & Computational Geometry*, 46(2):270–282, Sep 2011.

[JS09] William b. Johnson and Gideon Schechtman. Diamond graphs and super-reflexivity. *Journal of Topology and Analysis*, 01(02):177–189, 2009.

[KK16] Lior Kamma and Robert Krauthgamer. Metric decompositions of path-separable graphs. *Algorithmica*, pages 1–9, 2016.

[KKM+12] Maleq Khan, Fabian Kuhn, Dahlia Malkhi, Gopal Pandurangan, and Kunal Talwar. Efficient distributed approximation algorithms via probabilistic tree embeddings. *Distributed Computing*, 25(3):189–205, 2012.

[KLMN05] Robert Krauthgamer, James R. Lee, Manor Mendel, and Assaf Naor. Measured descent: a new embedding method for finite metrics. *Geometric and Functional Analysis*, 15(4):839–858, 2005.

[LLR95] N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

[LN04] James R. Lee and Assaf Naor. Embedding the diamond graph in ℓ_p and dimension reduction in ℓ_1. *Geometric & Functional Analysis GAFA*, 14(4):745–747, 2004.

[LR10] James R. Lee and Prasad Raghavendra. Coarse differentiation and multi-flows in planar graphs. *Discrete & Computational Geometry*, 43(2):346–362, 2010.

[LS10] James R. Lee and Anastasios Sidiropoulos. Genus and the geometry of the cut graph. In *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 193–201, 2010.

[LS13] James R. Lee and Anastasios Sidiropoulos. Pathwidth, trees, and random embeddings. *Combinatorica*, 33(3):349–374, 2013.

[Mil86] Gary L. Miller. Finding small simple cycle separators for 2-connected planar graphs. *J. Comput. Syst. Sci.*, 32(3):265–279, 1986.

[MN13] Manor Mendel and Assaf Naor. Markov convexity and local rigidity of distorted metrics. *Journal of the European Mathematical Society*, 15(1):287–337, 2013.

[NR03] Ilan Newman and Yuri Rabinovich. A lower bound on the distortion of embedding planar metrics into Euclidean space. *Discrete Comput. Geom.*, 29(1):77–81, 2003.

[OS81] Haruko Okamura and P.D. Seymour. Multicommodity flows in planar graphs. *Journal of Combinatorial Theory, Series B*, 31(1):75 – 81, 1981.

[Rao99] Satish Rao. Small distortion and volume preserving embeddings for planar and Euclidean metrics. In *Proceedings of the Fifteenth Annual Symposium on Computational Geometry, Miami Beach, Florida, USA, June 13-16, 1999*, pages 300–306, 1999.

[ST04] Yuval Shavitt and Tomer Tankel. Big-bang simulation for embedding network distances in Euclidean space. *IEEE/ACM Trans. Netw.*, 12(6):993–1006, 2004.

[Tho04] Mikkel Thorup. Compact oracles for reachability and approximate distances in planar digraphs. *J. ACM*, 51(6):993–1024, November 2004.

A An Illustation of the Embedding

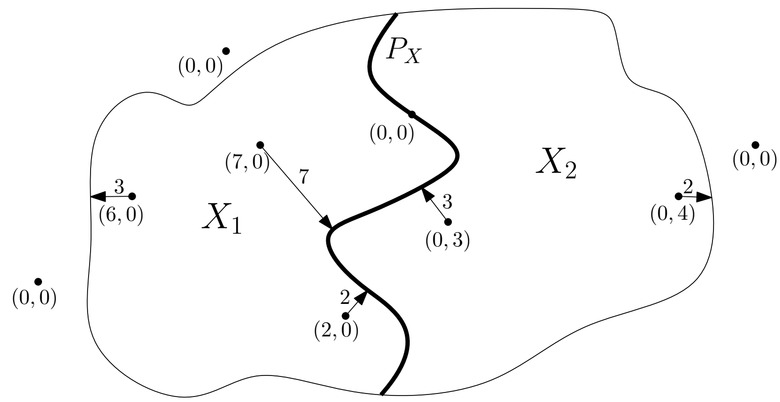


Figure 5: *The set $X \in \mathcal{X}_i$ surrounded by a closed curve. The path P_X partitions X into $X_1, X_2 \in \mathcal{X}_{i+1}$. The embedding f_X^{path} consists of two coordinates, and represented in the figure by a horizontal vector next to each vertex, where the first entry is w.r.t. X_1 and the second w.r.t. X_2. Each point on P_X, or not in X maps to* 0 *in both the coordinates. Each point in X_1 maps to* $\min\{d_X(v, P_X), 2d_G(v, V \setminus X)\}$ *in the first coordinate and to* 0 *in the second.*

B Proof of Lemma 4

In this section we bound the expansion of any coordinate in our embedding. Recall that the embedding of v lying in some component X consists of two sets of coordinates: its distance from the path, and its distance from the root. As mentioned in the introduction, since points outside X are mapped to zero, maintaining Lipschitz-ness requires us to truncate the contribution of v of any coordinate to its distance from the boundary. This truncation (either via taking a minimum with $d_G(v, V \setminus X)$, or via the sawtooth function), means that our proofs of expansion require more care. Let us now give the details.

Consider any level i, any set $X \in \mathcal{X}_i$, and any pair of vertices u, v. It suffices to show that $\|f_X(v) - f_X(u)\|_\infty = O(d_G(u, v))$. To begin, we may assume that both $u, v \in X$. Indeed, if both $u, v \notin X$, then $f_X(v) = f_X(u) = \vec{0}$ and we are done. If one of them, say v belongs to X while the other $u \notin X$, then $f_X(u) = \vec{0}$ while $f_X(v)$ is bounded by $2^{t_v+1} \leq 4d_G(v, V \setminus X) \leq 4\, d_G(u, v)$ in each coordinate.

Moreover, we may also assume that the shortest u-v path in G contains only vertices from X. Indeed, suppose their shortest path in G uses vertices from $V \setminus X$, then $d_G(u, V \setminus X) + d_G(v, V \setminus X) \leq d_G(u, v)$. But since both $f_X(v), f_X(u)$ are bounded in each coordinate by $4 \cdot \max\{d_G(u, V \setminus X), d_G(v, V \setminus X)\}$, we have constant expansion. Henceforth, we can assume that $d_G(u, v) = d_X(u, v)$. We now bound the expansion in each of the two parts of f_X separately.

Expansion of f_X^{path}. Let X_v, X_u be the connected components in $G[X \setminus P_X]$ such that $v \in X_v$ and $u \in X_u$. Consider the first case $X_v \neq X_u$, then P_X intersects the shortest path between v and u. In particular,

$$\begin{aligned}\|f_X^{\text{path}}(v) - f_X^{\text{path}}(u)\|_\infty &\leq \min\{d_X(v, P_X), 2d_G(v, V \setminus X)\} + \min\{d_X(u, P_X), 2d_G(u, V \setminus X)\} \\ &\leq d_X(v, P_X) + d_X(u, P_X) \leq d_X(v, u) = d_G(v, u) \ .\end{aligned}$$

Otherwise, $X_v = X_u$ and the two vertices lie in the same component. Now $\|f_X^{\text{path}}(v) - f_X^{\text{path}}(u)\|_\infty$

equals

$$\left| \min \left\{ d_X \left(v, P_X \right), 2 d_G \left(v, V \setminus X \right) \right\} - \min \left\{ d_X \left(u, P_X \right), 2 d_G \left(u, V \setminus X \right) \right\} \right|$$

Assuming (without loss of generality) that the first term is at least the second, we can drop the absolute value signs. Now the bound on the expansion follows from a simple case analysis. Indeed, suppose $d_X(u, P_X) \leq 2d_G(u, V \setminus X)$. Then we get

$$\begin{aligned} \| f_X^{\text{path}}(v) - f_X^{\text{path}}(u) \|_\infty &= \min \left\{ d_X \left(v, P_X \right), 2 d_G \left(v, V \setminus X \right) \right\} - d_X \left(u, P_X \right) \\ &\leq d_X \left(v, P_X \right) - d_X \left(u, P_X \right) \leq d_X(u, v) = d_G(u, v). \end{aligned}$$

The other case is that $d_X(u, P_X) > 2d_G(u, V \setminus X)$, and then

$$\begin{aligned} \| f_X^{\text{path}}(v) - f_X^{\text{path}}(u) \|_\infty &= \min \left\{ d_X \left(v, P_X \right), 2 d_G \left(v, V \setminus X \right) \right\} - 2 d_G \left(u, V \setminus X \right) \\ &\leq 2 d_G \left(v, V \setminus X \right) - 2 d_G \left(u, V \setminus X \right) \leq 2 d_G(u, v). \end{aligned}$$

Hence the expansion is bounded by 2.

Expansion of f_X^{root}. Let r be the root of P_X. For $t \in \{0, 1, \ldots, M\}$, let p_t (respectively, q_t) be the "weight" of v (respectively, u) on g_t—in other words, p_t is the constant in (1) such that $f_{X,t}^{\text{root}}(v) = p_t \cdot g_t(\beta \cdot d_X(v, r) + \alpha \cdot 2^{t+1})$. Note that $p_t \in \{0, \lambda_v, 1 - \lambda_v\}$ is chosen deterministically, and is nonzero for at most two indices t.

First, observe that for every t,

$$\begin{aligned} \left| f_{X,t}^{\text{root}}(v) - f_{X,t}^{\text{root}}(u) \right| &= \left| p_t \cdot g_t(\beta \cdot d_X(v, r) + \alpha \cdot 2^{t+1}) - q_t \cdot g_t(\beta \cdot d_X(u, r) + \alpha \cdot 2^{t+1}) \right| \\ &\leq \min \{p_t, q_t\} \cdot \left| g_t(\beta \cdot d_X(v, r) + \alpha \cdot 2^{t+1}) - g_t(\beta \cdot d_X(u, r) + \alpha \cdot 2^{t+1}) \right| + |p_t - q_t| \cdot 2^t \\ &\leq \min \{p_t, q_t\} \cdot \beta \cdot |d_X(v, r) - d_X(u, r)| + |p_t - q_t| \cdot 2^t \\ &\leq O(d_G(u, v)) + |p_t - q_t| \cdot 2^t \ . \end{aligned} \tag{5}$$

The first inequality used that g_t is bounded by 2^t, and the second inequality that g_t is 1-Lipschitz; both follow from Observation 1. The last inequality follows by the triangle inequality (since we assumed that the shortest path from v to u is contained within X).

Hence, it suffices to show that $|p_t - q_t| = O(d_G(u, v)/2^t)$. Indeed, for indices $t \notin \{t_u, t_u + 1, t_v, t_v + 1\}$, $p_t = q_t = 0$, hence $|p_t - q_t| = 0$. Let us consider the other cases. W.l.o.g., assume that $d_G(v, V \setminus X) \geq d_G(u, V \setminus X)$ and hence $t_v \geq t_u$.

- $\boldsymbol{t_u = t_v}$ **:** In this case, $|p_{t_v} - q_{t_v}| = |(1 - \lambda_v) - (1 - \lambda_u)| = \lambda_v - \lambda_u = |p_{t_v+1} - q_{t_v+1}|$. Moreover, this quantity is

$$\begin{aligned} \lambda_v - \lambda_u &= \frac{2 d_G \left(v, V \setminus X \right) - 2^t}{2^t} - \frac{2 d_G \left(u, V \setminus X \right) - 2^t}{2^t} \\ &= \frac{2 \left(d_G \left(v, V \setminus X \right) - d_G \left(u, V \setminus X \right) \right)}{2^t} \quad \leq \quad \frac{2 d_G(u, v)}{2^t} \ . \end{aligned}$$

  Hence, we get that $|p_t - q_t| = O(d_G(u, v)/2^t)$ for both $t \in \{t_v, t_v + 1\}$.

- $\boldsymbol{t_u = t_v - 1}$: It holds that

$$\begin{aligned} \lambda_v + (1 - \lambda_u) &\leq 2 \cdot \frac{2 d_G \left(v, V \setminus X \right) - 2^{t_v}}{2^{t_v}} + \frac{2^{t_u+1} - 2 d_G \left(u, V \setminus X \right)}{2^{t_u}} \\ &= \frac{2 d_G \left(v, V \setminus X \right) - 2 d_G \left(u, V \setminus X \right)}{2^{t_u}} \leq \frac{2 d_G(u, v)}{2^{t_u}} \ . \end{aligned}$$

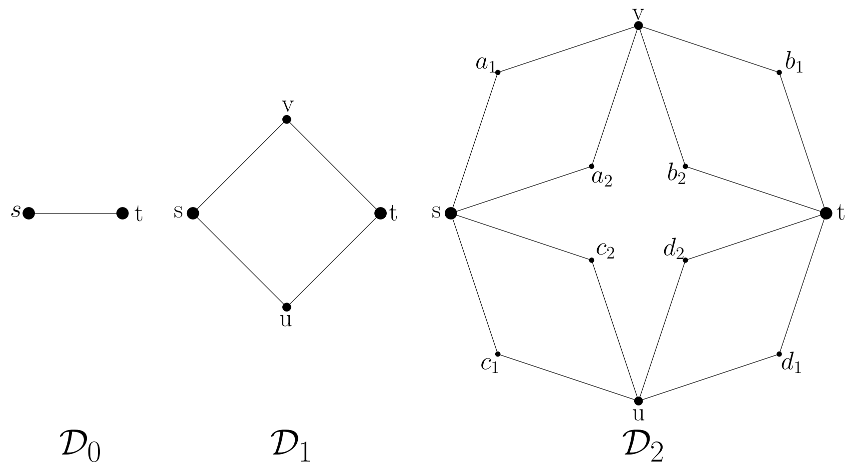


Figure 6: The first 3 diamond graphs. $\{s,t\}$ is the level 0 diagonal, $\{u,v\}$ is the level 1 diagonal, $\{a_1,a_2\}$, $\{b_1,b_2\}$, $\{c_1,c_2\}$, $\{d_1,d_2\}$ are the level 2 diagonals. $E_0=\{\{s,t\}\}$, $E_1=\{\{s,u\},\{t,u\},\{t,v\},\{s,v\}\}$.

If we define $\chi:=\lambda_v+(1-\lambda_u)$, we conclude that

$$\begin{array}{lllllll}
|p_{t_v+1}-q_{t_v+1}| & = & \lambda_v & \le & \chi & = & O(d_G(u,v)/2^{t_v+1}) \\
|p_{t_v}-q_{t_v}| & = & |1-\lambda_v-\lambda_u| & \le & \chi & = & O(d_G(u,v)/2^{t_v}) \\
|p_{t_u}-q_{t_u}| & = & 1-\lambda_u & \le & \chi & = & O(d_G(u,v)/2^{t_u})\ .
\end{array}$$

- $\boldsymbol{t_u < t_v - 1}$: By the definition of t_v and t_u,

$$2d_G(v,u)\ge 2\left(d_G(v,V\setminus X)-d_G(u,V\setminus X)\right)\ge 2^{t_v}-2^{t_u+1}\ge 2^{t_v-1}.$$

In particular, for every $t\le t_v+1$, $|p_t-q_t|\le 1\le \frac{2d_G(u,v)}{2^{t_v-1}}=O\left(\frac{d_G(u,v)}{2^t}\right)$.

C Lower Bound: Proof of Theorem 3

We start with the definition of the diamond graphs D_k.

Definition 2 (Diamond Graphs)**.** *Let $\mathcal{D}_0,\mathcal{D}_1,\mathcal{D}_2,\dots$ be a sequence of graphs defined as follows: $\mathcal{D}_0$ is a single edge, and for $i\ge 1$, $\mathcal{D}_i$ is obtained from $\mathcal{D}_{i-1}$ by replacing every edge of $\mathcal{D}_{i-1}$ with a square with two new vertices. See Figure 6 for illustration.*

For each of the new squares created at level i, call the two new vertices a *diagonal at level-i*.

Consider the graph $\mathcal{D}_k$. For $1\le i\le k$, denote by D_i the set of all level-i diagonals, and let E_i be the set of pairs of vertices which were edges in $\mathcal{D}_i$. It holds that $|E_k|=4^k$ and $D_k=4^{k-1}$. Moreover, $\mathcal{D}_0$ has pathwidth 1, and it can be verified by induction that $\mathcal{D}_k$ has pathwidth $k+1$.

Newman and Rabinovich [NR03] proved that every embedding of $\mathcal{D}_k$ into ℓ_2 requires distortion $\sqrt{k+1}$. Lee and Naor [LN04] generalized it for $1<p\le 2$, by proving that every embedding of $\mathcal{D}_k$ into ℓ_p requires distortion $\sqrt{1+(p-1)k}$. We will prove that for $p\ge 2$, every embedding of $\mathcal{D}_k$ into ℓ_p requires distortion at least $\left(\frac{k+1}{2^{p-2}}\right)^{1/p}$.

The following claim will be essential.

Claim 2 (ℓ_p Quadrilateral Inequality). *For $p \geq 2$, and every four vectors $a, b, c, d \in \ell_p$, it holds that*

$$\|a-c\|_p^p + \|b-d\|_p^p \leq 2^{p-2}\left(\|a-b\|_p^p + \|b-c\|_p^p + \|c-d\|_p^p + \|d-a\|_p^p\right) \tag{6}$$

Proof. The proof of the following inequality can be found at [Car04, Theorem 11.12],

$$\forall x, y \in \ell_p, \quad \|x+y\|_p^p + \|x-y\|_p^p \leq 2^{p-1}\left(\|x\|_p^p + \|y\|_p^p\right) .$$

Define $x_1 = b-a$, $y_1 = a-d$ and $x_2 = b-c$, $y_2 = c-d$. We get

$$\|b-d\|_p^p + \|b-2a+d\|_p^p \leq 2^{p-1}\left(\|a-b\|_p^p + \|d-a\|_p^p\right) .$$
$$\|b-d\|_p^p + \|b-2c+d\|_p^p \leq 2^{p-1}\left(\|b-c\|_p^p + \|c-d\|_p^p\right) .$$

By summing up and dividing by 2,

$$\|b-d\|_p^p + \frac{\|b-2a+d\|_p^p + \|b-2c+d\|_p^p}{2} \leq 2^{p-2}\left(\|a-b\|_p^p + \|b-c\|_p^p + \|c-d\|_p^p + \|d-a\|_p^p\right) . \tag{7}$$

The claim now follows by convexity. □

Fix some $p \geq 2$, and embedding $f : \mathcal{D}_k \to \ell_p$. We will assume w.l.o.g that f is non-contractive, and denote by ρ its expansion, i.e. distortion. Set $\alpha_0 = \frac{1}{2^{k(p-2)}}$ and for $i > 0$, $\alpha_i = \frac{1}{2^{(k-i+1)(p-2)}}$. Note that for $i \geq 1$, $\alpha_i \cdot 2^{p-2} = \alpha_{i+1}$. Our proof will be based on the following Poincaré-type inequality

Claim 3 (Diamond graph ℓ_p Poincaré-type inequality).

$$\sum_{i=0}^{k} \alpha_i \cdot \sum_{\{x,y\} \in D_i} \|f(x) - f(y)\|_p^p \leq \alpha_{k+1} \cdot \sum_{\{x,y\} \in E_k} \|f(x) - f(y)\|_p^p . \tag{8}$$

Proof. For edge $\{x, y\} \in E_{i-1}$, denote by $\{x', y'\} \in D_i$ the diagonal created by it. We have

$$\begin{aligned}
&\sum_{\{x,y\} \in D_i} \|f(x) - f(y)\|_p^p + \sum_{\{x,y\} \in E_{i-1}} \|f(x) - f(y)\|_p^p \\
&= \sum_{\{x,y\} \in E_{i-1}} \left(\|f(x) - f(y)\|_p^p + \left\|f(x') - f(y')\right\|_p^p\right) \\
&\overset{(6)}{\leq} 2^{p-2} \cdot \left(\sum_{\{x,y\} \in E_{i-1}} \left(\left\|f(x) - f(x')\right\|_p^p + \left\|f(x') - f(y)\right\|_p^p + \left\|f(y) - f(y')\right\|_p^p + \left\|f(y') - f(x)\right\|_p^p\right)\right) \\
&= 2^{p-2} \cdot \sum_{\{x,y\} \in E_i} \|f(x) - f(y)\|_2^2 .
\end{aligned}$$

Summing over $1 \leq i \leq k$'s, with appropriate scaling,

$$\begin{aligned}
&\sum_{i=1}^{k} \alpha_i \cdot \left(\sum_{\{x,y\} \in D_i} \|f(x) - f(y)\|_p^p + \sum_{\{x,y\} \in E_{i-1}} \|f(x) - f(y)\|_p^p\right) \\
&\qquad \leq \sum_{i=1}^{k} \alpha_i \cdot 2^{p-2} \cdot \sum_{\{x,y\} \in E_i} \|f(x) - f(y)\|_p^p \quad = \sum_{i=1}^{k} \alpha_{i+1} \cdot \sum_{\{x,y\} \in E_i} \|f(x) - f(y)\|_p^p .
\end{aligned}$$

Hence

$$\sum_{i=1}^{k}\alpha_i\cdot\sum_{\{x,y\}\in D_i}\|f(x)-f(y)\|_p^p+\alpha_1\cdot\sum_{\{x,y\}\in E_0}\|f(x)-f(y)\|_p^p\le\alpha_{k+1}\cdot\sum_{\{x,y\}\in E_k}\|f(x)-f(y)\|_p^p\ .$$

As $E_0=D_0$ and $\alpha_0=\alpha_1$, the claim follows. □

Next, we calculate

$$\begin{aligned}\sum_{i=0}^{k}\alpha_i\cdot\sum_{\{x,y\}\in D_i}d(x,y)^p=&\alpha_0\cdot\left(2^k\right)^p+\sum_{i=1}^{k}|D_i|\cdot\alpha_i\cdot\left(2^{k-i+1}\right)^p\\ &=2^{2k}+\sum_{i=1}^{k}4^{i-1}\cdot\left(2^{k-i+1}\right)^2=(k+1)\cdot4^k\ .\end{aligned}\tag{9}$$

Recall that f is non-contractive and has expansion ρ. Consequently,

$$\begin{aligned}(k+1)\cdot4^k&\overset{(9)}{=}\sum_{i=0}^{k}\alpha_i\cdot\sum_{\{x,y\}\in D_i}d(x,y)^p\\ &\le\sum_{i=0}^{k}\alpha_i\cdot\sum_{\{x,y\}\in D_i}\|f(x)-f(y)\|_p^p\\ &\overset{(8)}{\le}\alpha_{k+1}\cdot\sum_{\{x,y\}\in E_k}\|f(x)-f(y)\|_p^p\\ &\le\alpha_{k+1}\cdot\sum_{\{x,y\}\in E_k}(\rho\cdot d(x,y))^p\quad=\alpha_{k+1}\cdot(\rho)^p\cdot|E_k|\ .\end{aligned}$$

We conclude

$$\rho\ge\left(\frac{4^k}{|E_k|}\cdot\frac{k+1}{\alpha_{k+1}}\right)^{1/p}=\left(\frac{k+1}{2^{p-2}}\right)^{1/p}=\Omega(k^{1/p})\ .$$

D Proof of Lemma 6

We restate the lemma for convenience:

Lemma 6 (Composition Lemma)**.** *Let (X,d) be a metric space. Suppose that there are constants ρ,τ and a function $f:X\to\mathbb{R}^s$, drawn from some probability space such that:*

1. *For every $u,v\in X$ and every $j\in[s]$, $|f_j(v)-f_j(u)|\le\rho\cdot d(v,u)$.*
2. *For every $u,v\in X$, there exists $j\in[s]$ such that $\mathbb{E}[\,|f_j(v)-f_j(u)|\,]\ge\frac{1}{\tau}\cdot d(v,u)$.*
3. *For every $v\in X$, there is a subset of indices $I_v\subseteq[s]$ of size $|I_v|\le k$, such that for every $j\notin I_v$, $f_j(v)=0$. In other words, for every $v\in X$, $f(v)$ has support of size at most k.*

Then, for every $p\ge1$, there is an embedding of (X,d) into ℓ_p with distortion $O(k^{1/p})$. Moreover, if there is an efficient algorithm for sampling such an f, then there is a randomized algorithm that constructs the embedding efficiently (in expectation).

Proof. Fix $n = |X|$, and set $m = 48\rho\tau \cdot \ln n$. Let $f^{(1)}, f^{(2)}, \ldots, f^{(m)} : X \to \mathbb{R}^s$ be functions chosen i.i.d according to the given distribution. Set $g = m^{-1/p} \bigoplus_{i=1}^{m} f^{(i)}$. We argue that with high probability, g has distortion $O(k^{1/p})$ in ℓ_p.

Fix some pair of vertices $v, u \in V$. Set $d(v,u) = \Delta$. The upper bound follows from Property 1 and Property 3 of the lemma:

$$\begin{aligned}\|g(v) - g(u)\|_p^p &= \sum_{i=1}^{m} \sum_{j \in I_v \cup I_u} \left(m^{-1/p} \cdot \left| f_j^{(i)}(v) - f_j^{(i)}(u) \right| \right)^p \\ &\leq \sum_{i=1}^{m} \sum_{j \in I_v \cup I_u} \frac{1}{m} \cdot (\rho \cdot \Delta)^p \leq 2k \cdot (\rho \cdot \Delta)^p ,\end{aligned}$$

thus $\|g(v) - g(u)\|_p \leq O(k^{1/p} \cdot \Delta)$.

Next, for the contraction bound, let j be the index of Property 2 w.r.t v, u. Set $\mathcal{F} = \{f \ : \ |f_j(v) - f_j(u)| \geq \Delta/2\tau\}$ to be the event that we draw a function with significant contribution to v, u. Then using Property 1 and Property 2,

$$\frac{\Delta}{\tau} \leq \mathbb{E}\left[|f_j(v) - f_j(u)|\right] \leq \Pr\left[\overline{\mathcal{F}}\right] \cdot \frac{\Delta}{2\tau} + \Pr\left[\mathcal{F}\right] \cdot \rho\Delta \leq \frac{\Delta}{2\tau} + \Pr\left[\mathcal{F}\right] \cdot \rho\Delta ,$$

which implies that $\Pr\left[\mathcal{F}\right] \geq \frac{1}{2\rho\tau}$. Let $Q_{u,v}^{(i)}$ be an indicator random variable for the event $f^{(i)} \in \mathcal{F}$, and set $Q_{u,v} = \sum_{i=1}^{m} Q_{u,v}^{(i)}$. By linearity of expectation, $\mathbb{E}[Q_{u,v}] \geq \frac{m}{2\rho\tau} = 24 \cdot \ln n$. By a Chernoff bound

$$\Pr\left[Q_{u,v} \leq 12 \cdot \ln n\right] \leq \Pr\left[Q_{u,v} \leq \frac{1}{2} \cdot \mathbb{E}\left[Q_{u,v}\right]\right] \leq \exp(-\frac{1}{8}\mathbb{E}\left[Q_{u,v}\right]) \leq \exp(-3\ln n) = n^{-3} .$$

By taking a union bound over the $\binom{n}{2}$ pairs, with probability at least $1 - \frac{1}{n}$, for every $u, v \in V$, $Q_{u,v} > 12\ln n = \Omega(m)$. (Recall that both ρ, τ are universal constants.) If this event indeed occurs, then the contraction is indeed bounded:

$$\begin{aligned}\|g(v) - g(u)\|_p^p &\geq \sum_{i=1}^{m} \left(m^{-1/p} \cdot \left| f_j^{(i)}(v) - f_j^{(i)}(u) \right| \right)^p \\ &\geq \frac{1}{m} \sum_{i : Q_{u,v}^{(i)} = 1} \left| f_j^{(i)}(v) - f_j^{(i)}(u) \right|^p \\ &\geq \frac{Q_{u,v}}{m} \cdot \left(\frac{\Delta}{2\rho\tau} \right)^p = \Omega\left(\left(\frac{\Delta}{2\rho\tau} \right)^p \right) .\end{aligned}$$

In particular, for every u, v, $\|g(v) - g(u)\|_p \geq \Omega(\Delta)$. $\square$

E Proof of the Sawtooth Lemma (Lemma 2)

We restate Lemma 2 for convenience:

Lemma 2 (Sawtooth Lemma)**.** *Let $x, y \in \mathbb{R}_+$. Let $\alpha \in [0,1]$, $\beta \in [0,4]$ be drawn uniformly and independently. The following properties hold:*

1. $\mathbb{E}_{\alpha,\beta}\left[g_t(\beta x + \alpha \cdot 2^{t+1})\right] = 2^{t-1}$.

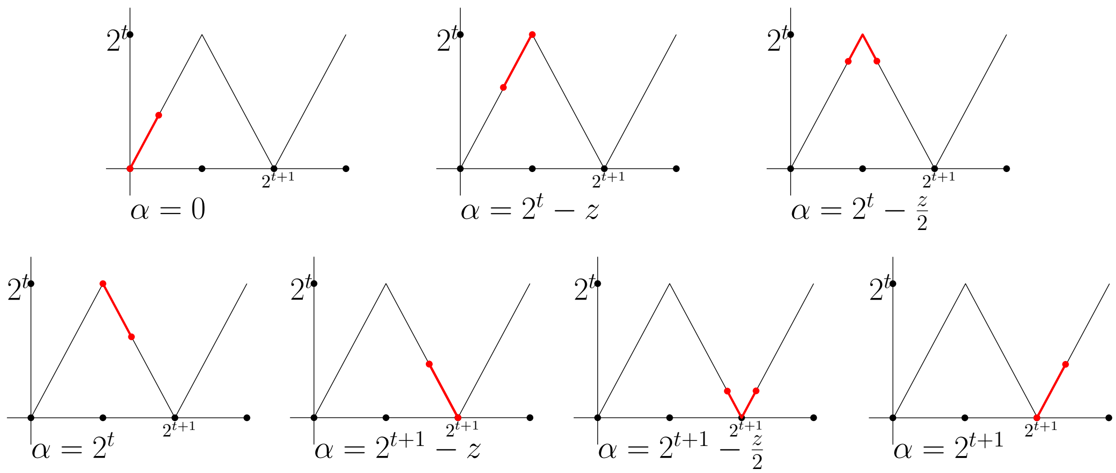


Figure 7: α is going from 0 to 2^{t+1}. $z \le 2^t$. In each of the figures the leftmost red point represents α while the rightmost red point represents $z+\alpha$. Each of the middle figures represent a moment when $g_t(z+\alpha) - g_t(z)$ changes its derivative.

2. *If* $|x-y| \le 2^{t-1}$*, then* $\mathbb{E}_{\alpha,\beta}\left[\left|g_t(\beta x + \alpha \cdot 2^{t+1}) - g_t(\beta y + \alpha \cdot 2^{t+1})\right|\right] = \Omega(|x-y|)$.
3. *If* $|x-y| > 2^{t-1}$*, then* $\mathbb{E}_{\alpha,\beta}\left[\left|g_t(\beta x + \alpha \cdot 2^{t+1}) - g_t(\beta y + \alpha \cdot 2^{t+1})\right|\right] = \Omega(2^t)$.

Property 1 is straightforward, as by Observation 1 g_t is periodic with period length 2^{t+1}. Indeed, for every fixed β, $\mathbb{E}_\alpha\left[g_t(\beta x + \alpha \cdot 2^{t+1})\right] = \mathbb{E}_\alpha\left[g_t(\alpha \cdot 2^{t+1})\right] = 2^{t-1}$. The following claim will be useful in the proofs of Property 2 and Property 3.

Claim 4. *For* $z \in [0, 2^{t+1}]$*,* $\mathbb{E}_{\alpha\in[0,1]}\left[\left|g_t(z + \alpha \cdot 2^{t+1}) - g_t(\alpha \cdot 2^{t+1})\right|\right] = \frac{(2^{t+1}-z)z}{2^{t+1}}$.

Proof. Set $(*) = \mathbb{E}_{\alpha\in[0,1]}\left[\left|g_t(z + \alpha \cdot 2^{t+1}) - g_t(\alpha \cdot 2^{t+1})\right|\right]$. By substituting the variable of integration, $(*) = \frac{1}{2^{t+1}} \cdot \int_0^{2^{t+1}} |g_t(z+\alpha) - g_t(\alpha)|\, d\alpha$. First assume that $z \le 2^t$, then there are 5 "phase changes" in $|g_t(z+\alpha) - g_t(\alpha)|$ from 0 to 2^{t+1}, at $2^t - z$, $2^t - \frac{z}{2}$, 2^t, $2^{t+1} - z$, $2^{t+1} - \frac{z}{2}$. (see Figure 7 for illustration). We calculate

$$\begin{aligned}
2^{t+1} \cdot (*) &= \int_0^{2^t - z} z d\alpha + \int_0^{\frac{z}{2}} (z - 2\alpha) d\alpha + \int_0^{\frac{z}{2}} 2\alpha d\alpha + \int_0^{2^t - z} z d\alpha + \int_0^{\frac{z}{2}} (z - 2\alpha) d\alpha + \int_0^{\frac{z}{2}} 2\alpha d\alpha \\
&= 2 \cdot \int_0^{2^t - z} z d\alpha + 2 \cdot \int_0^{\frac{z}{2}} z d\alpha = \left(2^{t+1} - z\right) z\,.
\end{aligned}$$

For $z > 2^t$, set $w = 2^{t+1} - z$. Then using that g^t is periodic,

$$\begin{aligned}
\mathbb{E}_{\alpha\in[0,1]}\left[\left|g_t(w + \alpha \cdot 2^{t+1}) - g_t(\alpha \cdot 2^{t+1})\right|\right] &= \mathbb{E}_{\alpha\in[0,1]}\left[\left|g_t(w + z + \alpha \cdot 2^{t+1}) - g_t(z + \alpha \cdot 2^{t+1})\right|\right] \\
&= \mathbb{E}_{\alpha\in[0,1]}\left[\left|g_t(2^{t+1} + \alpha \cdot 2^{t+1}) - g_t(z + \alpha \cdot 2^{t+1})\right|\right] \\
&= \mathbb{E}_{\alpha\in[0,1]}\left[\left|g_t(z + \alpha \cdot 2^{t+1}) - g_t(\alpha \cdot 2^{t+1})\right|\right]\,.
\end{aligned}$$

Hence by the first case, $(*) = \frac{(2^{t+1}-w)w}{2^{t+1}} = \frac{(2^{t+1}-z)z}{2^{t+1}}$. $\square$

For the proofs of Property 2 and Property 3 assume w.l.o.g $x > y$. Set $z = x - y$, and $(*) = \mathbb{E}_{\alpha,\beta}\left[\left|g_t(\beta x + \alpha \cdot 2^{t+1}) - g_t(\beta y + \alpha \cdot 2^{t+1})\right|\right]$. As g_t is a periodic function, we have that $(*) = \mathbb{E}_\beta\left[\mathbb{E}_\alpha\left[\left|g_t(\beta z + \alpha \cdot 2^{t+1}) - g_t(\alpha \cdot 2^{t+1})\right|\right]\right]$.

Proof of Property 2. Using Claim 4, we have

$$(*) = \mathbb{E}_\beta\left[\frac{\left(2^{t+1}-\beta z\right)\beta z}{2^{t+1}}\right] = \frac{1}{2^{t+1}}\cdot\frac{1}{4}\cdot\left(\frac{2^{t+1}z}{2}\beta^2 - \frac{z^2}{3}\cdot\beta^3 \,\Big|_0^4\right)$$
$$= \frac{1}{4}\cdot\left(\frac{16}{2}\cdot z - \frac{z^2}{2^{t+1}}\cdot\frac{2^6}{3}\right) \geq \frac{1}{4}\cdot\left(8-\frac{64}{3\cdot 4}\right)\cdot z = \frac{2}{3}\cdot|x-y| \ ,$$

where in the inequality we used that $z \leq 2^{t-1}$. □

Proof of Property 3. As g_t is periodic function, Claim 4 implies that for every $w \geq 0$ it holds that $\mathbb{E}_\alpha\left[\left|g_t(w+\alpha\cdot 2^{t+1}) - g_t(\alpha\cdot 2^{t+1})\right|\right] = \frac{\left(2^{t+1}-(w \mod 2^{t+1})\right)(w \mod 2^{t+1})}{2^{t+1}}$. Let $a \in [0,4]$ such that $a\cdot z = 2^{t+1}$ (such a exists as $|x-y| > 2^{t-1}$). The claim follows as,

$$\begin{aligned}
(*)\cdot 2^{t+1} &= \mathbb{E}_{\beta\in[0,4]}\left[\left(2^{t+1} - (\beta z \mod 2^{t+1})\right)(\beta z \mod 2^{t+1})\right] \\
&\geq \sum_{i=0}^{\lfloor\frac{4}{a}\rfloor-1}\frac{1}{4}\cdot\int\limits_{ia}^{(i+1)a}\left(2^{t+1} - (\beta\cdot\frac{2^{t+1}}{a} \mod 2^{t+1})\right)(\beta\cdot\frac{2^{t+1}}{a} \mod 2^{t+1})d\beta \\
&= \sum_{i=0}^{\lfloor\frac{4}{a}\rfloor-1}\frac{1}{4}\cdot\int\limits_0^a\left(2^{t+1}-\beta\cdot\frac{2^{t+1}}{a}\right)\cdot\beta\cdot\frac{2^{t+1}}{a}d\beta \\
&= \left\lfloor\frac{4}{a}\right\rfloor\cdot\frac{1}{4}\cdot\frac{a}{2^{t+1}}\cdot\int\limits_0^{2^{t+1}}\left(2^{t+1}-\gamma\right)\cdot\gamma d\gamma \\
&\geq \frac{4}{2a}\cdot\frac{1}{4}\cdot\frac{a}{2^{t+1}}\cdot\left(2^{t+1}\frac{\gamma^2}{2}-\frac{\gamma^3}{3}\,\Big|_0^{2^{t+1}}\right) = \frac{1}{2}\cdot\frac{1}{2^{t+1}}\cdot\frac{\left(2^{t+1}\right)^3}{6} = \frac{\left(2^{t+1}\right)^2}{12} \ .
\end{aligned}$$

□