Power of weak v/s strong triangle inequalities

Mohammad Moharrami *, Sushant Sachdeva †

July 5, 2011

Abstract

In a recent paper, Lee and Moharrami [13] construct a family of metrics (X_n, d_n) with $|X_n| = N$, such that $(X_n, \sqrt{d_n})$ embeds into L_2 with constant distortion, but embedding (X_n, d_n) into a metric of negative type requires distortion $\tilde{\Omega}(\log^{1/4} N)$. In this paper, we build on their analysis, and improve their result by showing an $\tilde{\Omega}(\log^{1/3} N)$ lower bound for embedding (X_n, d_n) into a metric of negative type. Moreover, we show that this analysis is essentially tight by constructing a map that has distortion $O(\log^{1/3} N)$.

This result implies a lower bound of $\tilde{\Omega}(\log^{1/3} N)$ for the integrality gap of the relaxed version of Goemans-Linial semidefinite program with weak triangle inequalities for Non-uniform Sparsest Cut.

1 Introduction

Let us recall the definition of a *negative-type* metric. A metric (X, d) is said to be of *negative-type* if there exists an embedding $f : X \to L_2$ such that $\forall x, y \in X : \; d(x, y) = \|f(x) - f(y)\|_2^2$. This is equivalent to saying that $(X, \sqrt{d})$ embeds into L_2 with distortion 1. By virtue of d being a metric, for all $x, y, z \in X$, f satisfies,

$$\|f(x) - f(y)\|_2^2 + \|f(y) - f(z)\|_2^2 \geq \|f(x) - f(z)\|_2^2. \tag{$*$}$$

The family of inequalities($*$) is known as *triangle inequalities*. Let $s = (v_0, \ldots, v_k)$ be a sequence of vertices, it is easy to verify that the triangle inequalities imply,

$$\sum_{i=1}^{k} \|f(v_i) - f(v_{i-1})\|_2^2 \geq \|f(v_k) - f(v_0)\|_2^2.$$

In fact, it is easy to see that these inequalities are equivalent to the triangle inequalities. We denote the class of all negative type metrics as NEG.

We can relax the inequalities to require that for every sequence of vertices $s = (v_0, \ldots, v_k)$,

$$C \sum_{i=1}^{k} \|f(v_i) - f(v_{i-1})\|_2^2 \geq \|f(v_k) - f(v_0)\|_2^2,$$

*Computer Science and Engineering, University of Washington. mohammad@cs.washington.edu
†Computer Science Department, Princeton University. sachdeva@cs.princeton.edu

where the constant $C > 1$ is independent of the sequence s. We call these inequalities *weak triangle inequalities* (Sometimes, to be unambiguous, we will refer to $(*)$ as *strong triangle inequalities*). It is easy to show that $f : X \to L_2$ satisfies weak triangle inequalities iff there is a metric (X, d) such that f is a constant distortion embedding for $(X, \sqrt{d})$ into L_2. We will call such metrics (X, d) as *weak negative type metrics* (Ideally, we should qualify it with the constant C, but we will drop it assuming $C = O(1)$).

Definition 1.1 *A metric (X, d) is a called a* weak negative type metric *if $(X, \sqrt{d})$ embeds into L_2 with constant distortion.*

The following question was asked by James Lee[17], [[]] Can every weak negative type metric be embedded into a negative type metric with O(1) distortion? ?
?

In the interest of space, we mention only a few reasons why the above question is interesting. We refer the interested reader to [13] for more details.

We need to recall the Sparsest Cut problem. Given a set of vertices V and two functions $\mathsf{cap}, \mathsf{dem} : V \times V \to \mathbb{R}^+$, the goal is to find a cut $(S, \bar{S})$ in V that minimizes the following ratio

$$\frac{\sum_{x \in S, y \in \bar{S}} \mathsf{cap}(x, y)}{\sum_{x \in S, y \in \bar{S}} \mathsf{dem}(x, y)}.$$

Goemans and Linial independently suggested the following SDP relaxation with triangle inequalities for Sparsest Cut.

$$\begin{aligned} \min \quad & \mathop{\mathbf{E}}_{x,y \in V} \left[\mathsf{cap}(x, y) \cdot \|v_x - v_y\|_2^2\right] & \\ \text{subject to} \quad & \mathop{\mathbf{E}}_{x,y \in V} \left[\mathsf{dem}(x, y) \cdot \|v_x - v_y\|_2^2\right] = 1 & \\ & \|v_x - v_y\|_2^2 + \|v_y - v_z\|_2^2 \geq \|v_x - v_z\|_2^2 & \forall x, y, z \in V(G). \end{aligned} \qquad (\dagger)$$

The integrality gap of $(\dagger)$ is the minimum distortion required for embedding an N-point negative type metric into L_1 (where $|V| = N$). Building on the work of [3], it was shown in [2] that this gap is at most $O(\sqrt{\log N} \log \log N)$ (following an earlier bound of [5]). SDP relaxations with triangle inequalities have also resulted in improved approximation ratios for several other problems [1, 4, 10].

Interestingly, for all the above results, the weak triangle inequalities are actually sufficient. A positive answer to question 1 would have implied that the optimal value (and hence the approximation guarantee) for $(\dagger)$ would change only by a constant factor if we replaced the triangle inequality constraints by weak triangle inequalities.

A positive answer would also have simplified the task of proving a lower bound on the integrality gap for $(\dagger)$ to proving a lower bound for the SDP with weak triangle inequalities (up to constants). Generally, proving lower bounds against strong triangle inequalities seems to be significantly more challenging than proving bounds for weak triangle inequalities[7, 6, 8, 9, 12, 11].

The above question was answered in the negative by Lee and Moharrami in [13] and a strong quantitative lower bound was exhibited.

Theorem 1.2 [13] *There is a family of weak negative type metrics (X_n, d_n) with $|X_n| = N$ such that embedding (X_n, d_n) into NEG requires distortion $\tilde{\Omega}(\log^{1/4} N)$.*

The $\tilde{\Omega}$ notation hides a factor of $\log \log N$.

1.1 Our Results and Contribution

In this paper, we build on the techniques of Lee and Moharrami[13] and give an improved analysis for embedding the family of metrics constructed in their paper into NEG and prove the following theorem.

Theorem 1.3 *There is a family of weak negative type metrics (X_n, d_n) with $|X_n| = N$ such that embedding (X_n, d_n) into* NEG *requires distortion $\tilde{\Omega}(\log^{1/3} N)$.*

Our result implies that the integrality gap for (†) with weak triangle inequalities (instead of strong triangle inequalities) is $\tilde{\Omega}(\log^{1/3} N)$. We note that this lower bound was the best in the literature until the very recent work of [16] that shows an almost tight lower bound of $\tilde{\Omega}(\sqrt{\log n})$. Still, our result is incomparable to [16] as they only show a lower bound for embedding a weak negative type metric into L_1 (Note that NEG does not embed isometrically into L_1 [11]).

The lower bound example in [13] is a certain graph G, composed with itself several times according to a well-defined graph product. It has been shown in [13] that the shortest path metric on the product graph is a weak negative type metric. The core of the lower bound result is an analysis of the distortion of an *'efficient'* embedding of G into NEG. We give an improved analysis of such an embedding of G to prove our result.

Moreover, we show that for the particular family of metrics constructed, the above analysis is tight up to a factor of $O(\log \log N)$.

Theorem 1.4 *For the metrics (X_n, d_n) from theorem 1.3, there is an embedding into* NEG *that has distortion $O(\log^{1/3} N)$.*

To construct these embeddings, we use some new and interesting ideas. In particular, we show that the embedding composition that was introduced in [13], to construct embeddings for product graphs, preserves strong triangle inequalities. Moreover, for the base graph G, we construct an embedding with an optimal trade-off between efficiency and distortion.

1.2 Techniques

Our work is built on the techniques from [13]. Below, we describe some of the main ideas in the paper.

Construction. We work with the same construction as in [13]. Here, we give a quick, informal description of their construction.

First let us define a graph product that we will require. Given a graph G with two vertices labeled s and t, we define $G^{\oslash k}$ inductively. $G^{\oslash 1} = G$, $G^{\oslash (k+1)}$ is obtained by replacing each edge in $G^{\oslash k}$ by a copy of G with s and t in place of the end points of the edge.

Let Q_m the m dimensional hypercube, and let B_m, R_m be the nodes with even and odd parity, respectively. Now construct the graph H_m as follows: We place m copies each of B_m, R_m alternatingly

$$B_m^{(1)} R_m^{(1)} B_m^{(2)} R_m^{(2)} \dots B_m^{(m)} R_m^{(m)}, \tag{1}$$

and add hypercube edges between adjacent pairs of layers. Furthermore, we add two vertices s and t and connect them to the vertices of the first and the last layer respectively using paths of length m. Denote this graph as H_m. This graph has been referred to as "string of cubes graph" in [13].

Our final metric is the shortest path metric on $H_m^{\oslash k}$ for some appropriate choice of k, m.

Efficiency. An important concept in the proof is the idea of efficiency [15, 13]. Consider a sequence of points $\{x_1, \ldots, x_k\}$ along with a distance function d. We say that the sequence $\{x_i\}$ is ε-efficient (with respect to d) if

$$\sum_{i=1}^{k-1} d(x_i, x_{i+1}) \leq (1+\varepsilon) d(x_1, x_k).$$

Note that if d satisfies the triangle inequality, the left side is at least $d(x_1, x_k)$. This notion of efficiency can be naturally extended to a distribution over sequences.

Lee and Moharrami[13] show that if an embedding of $G^{\oslash k}$ into any metric space has distortion D, it must contain a copy of the graph G with an embedding that is ε-efficient for $\varepsilon = O(\frac{1}{k}\log D))$. Taking k to be sufficiently large, we can assume that we have an ε-efficient embedding for ε small enough. It then suffices to show that an efficient embedding of H_m into NEG must have high distortion.

Strengthened Poincaré Inequality. The classical Enflo's Poincaré inequality for the hypercube $\{0,1\}^m$ states that, given $f : Q_m \to \mathbb{R}$, the following inequality holds:

$$\mathop{\mathbf{E}}_{x \in Q_m} (f(x) - f(\bar{x}))^2 \leq m \cdot \mathop{\mathbf{E}}_{x \in Q_m, k \in [m]} (f(x) - f(x \oplus e_k))^2,$$

where e_k denotes the bit string with all zeros except a one in the k^{th} position.

This inequality can be easily extended to functions $f : Q_m \to L_2$ by integrating,

$$\mathop{\mathbf{E}}_{x \in Q_m} \|f(x) - f(\bar{x})\|^2 \leq m \cdot \mathop{\mathbf{E}}_{x \in Q_m, k \in [m]} \|f(x) - f(x \oplus e_k)\|^2.$$

Given an efficient embedding $f : H_m \to \mathsf{NEG}$, we can restrict f to each copy of Q_m in H_m to obtain embeddings $f_i : Q_m \to L_2$. We will apply Poincaré's inequality to $F = \mathbf{E}_i f_i$ in order to obtain a lower bound on the distortion of the embedding.

2 Construction

2.1 Recursive Composition

We recall the definition of the $\oslash$ operation from [13] which is an extension of the definition in [15]). An *s-t graph* G is a graph which has two distinguished vertices $s(G), t(G) \in V(G)$. We define the length of an s-t graph G as $\mathsf{len}(G) = d_G(s,t)$. Throughout the paper, we will only be concerned with *symmetric* s-t graphs, i.e. graphs for which there is an automorphism which maps s to t. We assume that all s-t graphs are symmetric in the following definitions. A *marked graph* $G = (V, E)$ is one which carries an additional subset $E_M(G) \subseteq E$ of *marked edges.* Every graph is assumed to be equipped with the trivial marking $E_M(G) = E(G)$ unless a marking is otherwise specified.

Definition 2.1 (Composition of s-t graphs) *Given two marked s-t graphs H and G, define $H \oslash G$ to be the s-t graph obtained by replacing each* marked *edge $(u,v) \in E_M(H)$ by a copy of G. Formally,*

- $V(H \oslash G) = V(H) \cup (E_M(H) \times (V(G) \setminus \{s(G), t(G)\}))$.
- *For every edge $(u,v) \in E(H) \setminus E_M(H)$, there is a corresponding edge in $H \oslash G$.*

- *For every edge $e = (u,v) \in E_M(H)$, there are $|E(G)|$ edges,*

$$\Big\{\Big((e,v_1),(e,v_2)\Big) \mid (v_1,v_2) \in E(G) \text{ and } v_1,v_2 \notin \{s(G),t(G)\}\Big\} \cup \\ \Big\{\Big(u,(e,w)\Big) \mid (s(G),w) \in E(G)\Big\} \cup \Big\{\Big((e,w),v\Big) \mid (w,t(G)) \in E(G)\Big\}$$

- *Of all the edges introduced in the previous step, the ones corresponding to marked edges in G are precisely the marked edges in $H \oslash G$.*

- $s(H \oslash G) = s(H)$ *and* $t(H \oslash G) = t(H)$.

If H and G are equipped with length functions $\mathsf{len}_H, \mathsf{len}_G$, respectively, we define $\mathsf{len} = \mathsf{len}_{H\oslash G}$ as follows. Using the preceding notation, for every edge $e = (u,v) \in E_M(H)$,

$$\begin{aligned} \mathsf{len}\left((e,v_1),(e,v_2)\right) &= \frac{\mathsf{len}_H(e)}{d_{\mathsf{len}_G}(s(G),t(G))}\mathsf{len}_G(v_1,v_2) \\ \mathsf{len}\left(u,(e,w)\right) &= \frac{\mathsf{len}_H(e)}{d_{\mathsf{len}_G}(s(G),t(G))}\mathsf{len}_G(s(G),w) \\ \mathsf{len}\left((e,w),v\right) &= \frac{\mathsf{len}_H(e)}{d_{\mathsf{len}_G}(s(G),t(G))}\mathsf{len}_G(w,t(G)). \end{aligned}$$

This choice implies that $H \oslash G$ contains an isometric copy of $(V(H), d_{\mathsf{len}_H})$.

Definition 2.2 (Recursive composition) *For a marked s-t graph G and a number $k \in \mathbb{N}$, we define $G^{\oslash k}$ inductively by letting $G^{\oslash 0}$ be a single edge of unit length, and setting $G^{\oslash k} = G^{\oslash k-1} \oslash G$.*

The following result is straightforward.

Lemma 2.3 (Associativity of $\oslash$) *For any three graphs A, B, C, we have $(A \oslash B) \oslash C = A \oslash (B \oslash C)$, both graph-theoretically and as metric spaces.*

Definition 2.4 (Copy of a graph) *For two graphs G, H, a subset of vertices $X \subseteq V(H)$ is said to be a copy of G if there exists a bijection $f : V(G) \to X$ with distortion 1, i.e. $d_H(f(u), f(v)) = C \cdot d_G(u,v)$ for some constant $C > 0$.*

Now we make the following two simple observations about copies of H and G in $H \oslash G$.

Observation 2.5 *The graph $H \oslash G$ contains $|E_M(H)|$ copies of the graph G, one copy corresponding to each marked edge in H.*

Observation 2.6 *The subset of vertices $V(H) \subseteq V(H \oslash G)$ form an isometric copy of H.*

Let G be an s-t graph G, (X,d) a metric space, and consider a mapping $f : V(G) \to X$. Let $\mathcal{P}_{s,t}(G)$ be the set of all s-t shortest-paths in G and let μ be a measure on $\mathcal{P}_{s,t} = \mathcal{P}_{s,t}(G)$. We say that f is *ε-efficient with respect to μ* if it satisfies

$$\mathop{\mathbf{E}}_{\gamma\sim\mu} \sum_{uv\in\gamma} d(f(u),f(v)) \le (1+\varepsilon)\, d(f(s),f(t)).$$

For a marked s-t graph G, we define its *marked length* by

$$\mathsf{len}_M(G) = \min_{\gamma \in \mathcal{P}_{s,t}} \sum_{uv \in \gamma : (u,v) \in E_M(G)} \mathsf{len}_G(u, v).$$

Now we present, the coarse differentiation theorem that we need from [13]. A proof is available in the conference version[14].

Theorem 2.7 [13] *Let G be a marked s-t graph. Then for any $D \geq 1$ and $\varepsilon \geq 2D\left(1 - \frac{\mathsf{len}_M(G)}{\mathsf{len}(G)}\right)$, there exists a $k = O(\frac{1}{\varepsilon} \log D)$ such that the following holds. For every metric space (X, d), distribution μ on s-t shortest paths and a mapping $f : V(G^{\oslash k}) \to X$ with distortion D, there exists a copy of G in $G^{\oslash k}$ such that $f|_G$ is ε-efficient with respect to μ.*

2.2 Graph construction

The construction is identical to the one in [13]. We will refer to the graphs H_m described in Section 1.2. We use $[Q_m]_i$ to denote the i^{th} copy of Q_m in H_m, and for a vertex $x \in V(Q_m)$, we use $[x]_i$ to denote the copy of x in $[Q_m]_i$. For a directed edge $\vec{e} = (u, v)$, we define $f(\vec{e}) = f(u) - f(v)$.

For $m \in \mathbb{N}$, we define the graph I_m as follows. We begin with a copy of H_m where all the edges are marked. Then, we relabel the vertices s and t in H_m as s' and t'. Next, we add two distinguished vertices s and t and connect s to s' and t to t' by an edge of length m^{-2}. These new edges are unmarked. Finally we replace each path between s' and $[Q_m]_1$, and t' and $[Q_m]_m$ with an edge of length m.

We construct $\hat{I}_m$ from I_m by replacing each marked edge $e \in E_M(I_m)$ with a path of length $\tau m^2 \mathsf{len}(e)$, all of whose edges are marked, and have length $\frac{1}{\tau m^2}$, where $\tau \in \mathbb{N}$ is a universal constant. Our final construction is of the form $\hat{I}_m^{\oslash k}$ for appropriate choices of $m, k \in \mathbb{N}$. We equip these graphs with the shortest-path metric, which we denote $d_{m,k}$. In [13] the following theorem was proved.

Theorem 2.8 [13] *There exists a map $f : (\hat{I}_m^{\oslash k}, \sqrt{d_{m,k}}) \to L_2$, such that $\mathsf{dist}(f) \asymp 1$.*

This theorem was proved by taking an embedding of the base graph, $\hat{I}_m$, into L_2 with constant distortion and reducing the distance between all pairs of vertices except for the distance between s and t. Formally, they defined f^δ, the δ-contraction of the map f, to be

$$f^\delta(v) = \delta f(v) + (1 - \delta) \frac{(f(v) - f(s)) \cdot (f(t) - f(s))}{\|f(t) - f(s)\|_2} (f(t) - f(s)).$$

It was shown that, for some constant δ, the mapping f^δ is efficient and has constant distortion. They used this embedding of the base graph and constructed an embedding for the whole graph that satisfies weak triangle inequalities.

In this paper, we take an efficient embedding of the base graph and combine it with another embedding that has low distortion to construct an embedding that has both low distortion and low efficiency (Theorem 2.10). Then, we show that the recursive composition of embeddings into L_2 that was introduced in [13] preserves the strong triangle inequalities to prove Theorem 2.9.

Theorem 2.9 *For $m \geq 1$ and some $k = O(\sqrt{m} \log m)$, any embedding of $\hat{I}_m^{\oslash k}$ into NEG requires distortion $\tilde{\Omega}(\log^{1/3} N)$, where $N = |V(\hat{I}_m^{\oslash k})| = 2^{O(mk)}$.*

Theorem 2.10 *There exists a map $f : \hat{I}_m^{\oslash k} \to \mathsf{NEG}$, such that $\mathsf{dist}(f) \lesssim \min(k, \sqrt{m})$ $\lesssim \log^{1/3} N$, where $N = |V(\hat{I}_m^{\oslash k})| = 2^{O(mk)}$.*

3 Lower bound for NEG embedding

We will use Theorem 2.7 to reduce our task to proving lower bounds on efficient embeddings. Let μ_m be the uniform measure over s-t shortest paths in H_m of the form

$$(s, \ldots, [x]_1, [x \oplus e_k]_1, [x]_2, [x \oplus e_k]_2, \ldots, [x]_m, [x \oplus e_k]_m, \ldots, t),$$

where $k \in \{1, 2, \ldots, m\}$ and $x \in Q_m$ are chosen uniformly at random.

As in [13], we intend to apply Enflo's Poincaré inequality to the average of the embeddings for Q_m obtained by restricting our embedding for H_m to $[Q_m]_o$. Thus, we need to upper bound the average squared length of the edges and lower bound the average squared distance between anti-podal points, in the *average embedding.*

The following lemma, proved in [13], lower bounds the correlation between the embeddings for anti-podal points. A proof has been included in the Appendix for completeness (Section A.1).

Lemma 3.1 [13] *For any non-expanding map $f : V(H_m) \to \mathsf{NEG}$ with distortion at most D and $i, j \in [m]$ such that $|i - j| \leq \frac{m}{8D}$,*

$$\langle f([x]_i) - f([\bar{x}]_i), f([x]_j) - f([\bar{x}]_j) \rangle \geq \frac{m}{2D}.$$

The next lemma upper bounds the average correlation between the embeddings for edges of Q_m. This lemma is at the heart of the distortion lower bound and improves on a similar lemma from [13]. We defer the proof to the Appendix (Section A.1).

Lemma 3.2 *For any ε-efficient non-expanding map $f : V(H_m) \to \mathsf{NEG}$, and for any $l \leq \frac{m}{2}$, there exists an index $p \in [m]$ such that:*

$$\mathop{\mathbf{E}}_{\vec{e} \in E(Q_m)} \mathop{\mathbf{E}}_{i,j \in [p, p+l-1]} \langle f([\vec{e}]_i), f([\vec{e}]_j) \rangle = O\left(l^{-1} + \varepsilon\right).$$

Assuming the two lemmas above, we can now prove that an efficient embedding of H_m has high distortion.

Lemma 3.3 *There is some $\delta > 0$ such that every $\frac{\delta}{\sqrt{m}}$-efficient embedding of H_m into NEG has distortion $\Omega_\delta(m^{1/2})$.*

Proof: Let $l = \left\lfloor \frac{m}{8D} \right\rfloor$ in Lemma 3.2 above and use the index p to define $F(x) = \mathbf{E}_{i \in [p, p+l-1]} f_i(x)$. Thus, we get

$$\mathop{\mathbf{E}}_{\vec{e} \in E(Q_m)} \|F(\vec{e})\|^2 = O\left(\frac{D}{m} + \varepsilon\right).$$

From Lemma 3.1, we get $\forall i, j \in [p, p+l-1]$, $\langle f_i(x) - f_i(\bar{x}), f_j(x) - f_j(\bar{x}) \rangle \geq \frac{m}{2D}$. Averaging this over i, j and x, we get

$$\mathop{\mathbf{E}}_{x \in V(Q_m)} \|F(x) - F(\bar{x})\|^2 \geq \frac{m}{2D}.$$

Now we can use the strengthened Poincaré's inequality for the cube that states that $\mathbf{E}_{x \in V(Q_m)} \|F(x) - F(\bar{x})\|^2 \leq m\, \mathbf{E}_{\vec{e} \in E(Q_m)} \|F(\vec{e})\|^2$. Plugging in, we get $\frac{m}{D} \lesssim m \cdot \left(\frac{D}{m} + \varepsilon\right)$.

This implies that for some $\varepsilon = \frac{\delta}{\sqrt{m}}$, we must have $D = \Omega_\delta(\sqrt{m})$. ■

■

Since an efficient embedding of H_m (and hence $\hat{I}_m$) must have high distortion, we can now use Theorem 2.7 in order to conclude that any embedding of $\hat{I}_m^{\oslash k}$ must have high distortion.

Theorem 3.4 *For $m \geq 1$ and some $k = \Theta(\sqrt{m}\log m)$, any embedding of $\hat{I}_m^{\oslash k}$ into* NEG *requires distortion $\tilde{\Omega}(\log^{1/3} N)$ where $N = |V(\hat{I}_m^{\oslash k})| = 2^{O(mk)}$.*

Proof: First, suppose $f : \hat{I}_m^{\oslash k} \to \mathsf{NEG}$ has distortion at most $\sqrt{m}$. By Theorem 2.7, there must exist a copy of $\hat{I}_m$, which is $\frac{\delta}{2\sqrt{m}}$-efficient with respect to μ_m (we pick $k = \Theta(\sqrt{m}\log m)$ appropriately). Since $\hat{I}_m$ contains an isometric copy of H_m, we can further restrict this map to obtain an embedding of H_m into NEG. It is easy to verify that the map $f|_{H_m}$ is $\frac{\delta}{\sqrt{m}}$-efficient.

By Lemma 3.2, this embedding must have distortion $\Omega(\sqrt{m})$. Also, $\log N = O(m^{3/2}\log m)$ and hence f has distortion at least $\Omega(\sqrt{m}) = \Omega\left(\frac{\log^{1/3} N}{\log\log N}\right)$. ■

■

4 Upper Bound for Distortion

We refer to the exact construction of the graph as defined in Section 2.2. We present two different embeddings of the base graph, where one of them has low distortion with constant efficiency from [13], and the other one is efficient but has high distortion. Later we combine these two embeddings to obtain an embedding obtain a trade-off between efficiency and distortion. We use the $\oslash$-composition from [13] to construct an embedding for the iterated graph based on the embedding of $\hat{I}_m$. We show that if the embedding for the base graph satisfies strong triangle inequalities then the embedding for the iterated graph also satisfies strong triangle inequalities. Finally, we prove a bound on the total distortion of the resulting embedding. Our main contributions in this section are 1) proving that the composition preserves strong triangle inequalities, 2) constructing an embedding with high efficiency and distortion $O(\sqrt{m})$ for the base graph, and 3) combining this embedding with the embedding from [13] to obtain an embedding for the base graph with the optimal trade-off between efficiency and distortion. For completeness, we repeat some of the proofs and definitions from [13].

In the rest of this section, we often work with I_m instead of directly working with $\hat{I}_m$.

4.1 Embedding of the Base Graph into L_1

In this section we borrow the definition of $\hat{f}$ and extension to subdivision from [13]. We start by presenting one of the tools that we use to embed $\hat{I}_m$ into L_1.

Randomly extending to a subdivision. Let G be a metric graph, and for $h \in \mathbb{N}$, let G_h denote the metric graph where every edge $e \in E(G)$ is replaced by a path of h edges, each of length $\mathsf{len}(e)/h$. Orient each $e \in E(G)$, and denote the new path between the endpoint of $e = (u, v)$ by $\{u = P_e(0), P_e(1), \ldots, P_e(h) = v\}$.

Given any subset $S \subseteq V(G)$, we define a random subset $\mathsf{ext}_h(S) \subseteq V(G_h)$ as follows. Let $\{X_e\}_{e \in E(G)}$ be a family of i.i.d. uniform $[0,1]$ random variables, and put

$$\mathsf{ext}_h(S) = S \cup \bigcup_{(u,v)\in E(G)} \left\{ P_e(i) : X_e \leq \frac{h-i}{h}\mathbf{1}_S(u) + \frac{i}{h}\mathbf{1}_S(v) \right\}.$$

It is easy to check that the distribution of $\mathsf{ext}(S)$ does not depend on the orientation chosen for the edges of G. The preceding operation corresponds to taking a cut $S \subseteq V(G)$ in the original graph, and extending it to G_h in the following way: For every edge $(u,v) \in E(G)$ that is cut by S, we cut the new path from u to v in G_h uniformly at random.

Now, given a cut measure μ, we define the extension $\mathcal{E}_h\mu$ to be the cut measure on G_h defined by

$$\mathcal{E}_h(\mu)(S) = \mu(S \cap V(G)) \cdot \mathbf{Pr}(S = \mathsf{ext}_h(S \cap V(G))),$$

for every subset $S \subseteq V(G_h)$. By abuse of notation, given a mapping $f : V(G) \to L_1$, we will use $\mathcal{E}_h f : V(G) \to L_1$ to denote the mapping which arises from constructing a cut measure μ_f from f, applying $\mathcal{E}_h$, and then passing back to an L_1-valued mapping.

Lemma 4.1 [13] *For any graph G and $h \in \mathbb{N}$, the following holds. For every $f : V(G) \to L_1$, we have*

$$\mathsf{dist}(\mathcal{E}_s f) \leq 5 \cdot \mathsf{dist}(f).$$

A proof has been provided in the Appendix (Section A.2).

We continue by defining a cut measure μ_m on H_m, which is the sum of the following three measures.

μ_{ver}: These cuts are the only cuts that separate s from t. For any $0 \leq k < 4m-1$, we assign weight $\frac{1}{2}$ to the cut

$$S = \{v : v \in V(H_m), d_{H_m}(s,v) \leq (k+1) + m^{-2}\}.$$

We also put weight m^{-2} on the cuts $\{s\}$ and $\{t\}$.

μ_{hor}: These cuts are the hypercube cuts and they do not separate s from t. For any integer $1 \leq k \leq m$ and $b \in \{0,1\}$, we put weight $\frac{1}{4}$ on the cut

$$S = \left\{[x]_i : x \in \{0,1\}^m, x_k = b, i \in [m]\right\}.$$

μ_{st}: The cut measure puts weight $\frac{m}{4}$ on the single cut $\{s, t, s', t'\}$.

One can easily verify that for every edge $(u,v) \in E(H_m)$, we have

$$\mathsf{len}(u,v) = d_{\mu_{\mathsf{ver}}}(u,v) + d_{\mu_{\mathsf{hor}}}(u,v) + d_{\mathsf{st}}(u,v)\,. \tag{2}$$

Let $f_m : V(H_m) \to L_1$ be the embedding corresponding to the cut measure $\mu_{\mathsf{ver}} + \mu_{\mathsf{hor}} + \mu_{\mathsf{st}}$.

Let $h = \tau m^3$, and put $G = (H_m)_h$. Since $\hat{H}_m$ is isometric to a subset of G, we can be prove the statement of the lemma for G. Letting $\hat{f}_m : V(G) \to L_1$ be defined by $\hat{f}_m = \mathcal{E}_h f_m$.

Lemma 4.2 [13] *The map $\hat{f}_m$ has constant distortion.*

Efficient embedding of $\hat{I}_m$ into $\mathbf{L_1}$. First we define η_m the corresponding cut measure for efficient embedding of I_m as follows. Recall that, $V(Q_m) = B_m \cup R_m$, where B_m and R_m denote the nodes of even and odd parity, respectively. Then Q_m is bipartite with respect to the partition (B_m, R_m). H_m is the graph which consists of $2m$ layers of the form (1). For every $c \in Q_m$ and $k \in \{-m+1, -m+2, \ldots, m\}$, there exists a cut $\left(S_{c,k}, \overline{S_{c,k}}\right)$ of weight $\frac{m}{2} \cdot \frac{1}{2m \cdot 2^m}$, where

$$\begin{aligned} S_{c,k} =& \{[x]_i | d_{Q_m}(x,c) - 2i \leq k, x \in B_m, i \in [m]\} \\ & \cup \{[x]_i | d_{Q_m}(x,c) - 2i - 1 \leq k, x \in R_m, i \in [m]\} \cup \{t, t'\}. \end{aligned}$$

Let G be the graph obtained by subdividing each edge of I_m into τm^3 edges. The metric (G, len_G) contains an isometric copy of $(\hat{I}_m, \mathsf{len}_{\hat{I}_m})$. We define $\hat{\eta}_m$ to be the restriction of $\mathcal{E}_{\tau m^3}\eta_m$ to vertices of $\hat{I}_m$.

Furthermore, we define $g_m : I_m \to L_1$ the map corresponding to $\eta_m + \mu_{\mathsf{ver}}$ and $\hat{g}_m : I_m \to L_1$ as the map corresponding to $\hat{\eta}_m + \hat{\mu}_{\mathsf{ver}}$. Letting $\hat{g}_m : V(G) \to L_1$ be defined by $\hat{g}_m = \mathcal{E}_h f_m$. Let γ be a shortest path between s and t, all the cuts in $\hat{\eta}_m$ and $\hat{\mu}_{\mathsf{ver}}$ cut every shortest path between s and t exactly once, hence

$$\sum_{uv \in \gamma} \|\hat{g}_m(u) - \hat{g}_m(v)\|_1 \leq \|\hat{g}_m(s) - \hat{g}_m(t)\|.$$

The next lemma bounds the distortion for $\hat{g}_m$. Its proof has been included in the Appendix (Section A.2).

Lemma 4.3 *The map $\hat{g}_m$ is non-expanding and has distortion $O(\sqrt{m})$.*

All the edges between two consecutive layers in the image of $\hat{g}_m$ have exactly the same length, and the mapping is non-expanding. We construct $\hat{g}'_m$ from $\hat{g}_m$ by adding the cuts of the form $S = \{v : v \in V(I_m), \mathsf{len}_{I_m}(s, v) \leq (k+1)\}$ with appropriate weights to make the length of each edge $e \in E(\hat{I}_m)$ in the image of g' exactly $\mathsf{len}_{\hat{I}_m}(e)$. The map $\hat{g}'_m$ is non-expanding and for any two vertices in $\hat{I}_m$ their distance has increased compared to $\hat{g}$, therefore distortion of $\hat{g}'$ is at most $O(\sqrt{m})$.

4.2 Embedding Composition

In this section we first present a composition of embeddings which is equivalent to the composition that was defined in [13]. This composition is used to construct an embedding for $G^{\oslash k}$ from embedding of the graph G. We show that this composition preserves strong triangle inequalities.

Projection. For a point $x \in \mathbb{R}^n$, and a subspace $s \in \mathbb{R}^n$, $\mathsf{proj}_s(x)$ is the orthogonal projection of x onto s. Abusing the notation, for a line segment $(y, z) \in \mathbb{R}^n$, we define

$$\mathsf{proj}_{(y,z)}(x) = y + \frac{(x-y)\cdot(z-y)}{\|z-y\|_2^2}(z-y).$$

The point $\mathsf{proj}_{(y,z)}(x)$ is the orthogonal projection of x on the line that passes through y and z.

Composition of s-t maps. Two-sum of maps. For two graphs $G = (V, E)$ and $H = (W, F)$, the 2-sum of G and H is constructed by first taking the disjoint union of $V(G)$ and $V(H)$, and then choosing edges $\vec{e}_1 \in \vec{E}$ and $\vec{e}_2 \in \vec{F}$, identifying them, together with their endpoints. Let I be the 2-sum of G and H over $\vec{e}_1 = (u_1, v_1)$ and $\vec{e}_2 = (u_2, v_2)$. Furthermore, suppose $f_G : V(G) \to \mathbb{R}^n$ and $f_H : V(H) \to \mathbb{R}^m$ and $\|f_G(\vec{e}_1)\|_2 = \|f_H(\vec{e}_2)\|_2$. We define two sum of maps f_G and f_H, $f : V(I) \to \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m$, as follows,

$$f(x) = \begin{cases} (0,0,0) & \text{if } x = u_1 \\ (\|f_G(\vec{e}_1)\|_2, 0, 0) & \text{if } x = v_1 \\ (\|f_G(u_1) - \mathsf{proj}_{s_1}(f_G(x))\|_2, f_G(x) - \mathsf{proj}_{s_1}(f_G(x)), 0) & \\ \qquad\qquad \text{if } x \in V(G)\backslash\{u_1, v_1\}, \text{where } s_1 = & (f_G(u_1), f_G(v_1)), \\ (\|f_H(u_2) - \mathsf{proj}_{s_2}(f_H(x))\|_2, 0, f_H(x) - \mathsf{proj}_{s_2}(f_H(x))) & \\ \qquad\qquad \text{if } x \in V(H)\backslash\{u_2, v_2\}, \text{where } s_2 = & (f_H(u_2), f_H(v_2)). \end{cases}$$

$\oslash$-embedding. Let G and H be two s-t graphs, equipped with maps $f_G : V(G) \to L_2$ and $f_H : V(H) \to L_2$. We construct $f_{G \oslash H} : V(G \oslash H) \to L_2$ by applying the 2-sum composition on scaled copies of f_H for all edges in $E_M(G)$.

This construction is obtained by scaling (by the factor $\frac{\|(f_G(v_i) - f_G(u_i)\|_2}{\|(f_H(t) - f_H(s)\|_2}$), changing the basis for map f_H to a disjoint basis except for $f_H(s)$ and $f_H(t)$, and then translating it to attach on edge (u_i, v_i).

Observation 4.4 *Let $x, y \in G \oslash H$ be on two distinct copies of H on edges (u_x, v_x) and (u_y, v_y). Furthermore let $s_x = (f_{G \oslash H}(u_x), f_{G \oslash H}(v_x))$ and $s_y = (f_{G \oslash H}(u_y), f_{G \oslash H}(v_y))$, then*

$$\begin{aligned} \|f_{G \oslash H}(x) - f_{G \oslash H}(y))\|_2^2 &= \|\mathsf{proj}_{s_y}(f_{G \oslash H}(y)) - \mathsf{proj}_{s_x}(f_{G \oslash H}(x))\|_2^2 \\ &+ \|f_{G \oslash H}(x) - \mathsf{proj}_{s_x}(f_{G \oslash H}(x))\|_2^2 + \|f_{G \oslash H}(y) - \mathsf{proj}_{s_y}(f_{G \oslash H}(y))\|_2^2. \end{aligned}$$

Lemma 4.5 [13] *Let G, H and I be three marked s-t graphs. Let $x, y \in G \oslash H \oslash I$, and let $f_G : G \to L_2$, $f_H : H \to L_2$ and $f_I : I \to L_2$ be the maps from these graphs to L_2, then $f_{(G \oslash H) \oslash I} \sim f_{G \oslash (H \oslash I)}$ up to translation and change of basis.*

We prove that the $\oslash$-composition preserves triangle inequalities. We defer the proof of the theorem to the Appendix (Section A.3).

Theorem 4.6 *Let G and H be two graphs and suppose that f is the two-sum of the maps $g : G \to \mathsf{NEG}$ and $h : H \to \mathsf{NEG}$ on the edges $\vec{e}_1 = (u_1, v_1)$ and $\vec{e}_2 = (u_2, v_2)$, then image of f is also a negative type metric.*

By using the $\oslash$-composition we can construct an embedding for $G^{\oslash k}$ based on $f : V(G) \to L_2$.The map $\alpha \hat{f}_m \oplus (1 - \alpha)\hat{g}'_m$ is $O(\alpha)$-efficient, and has distortion $\min(1/\alpha, \sqrt{m})$. Every L_1 metric is also a NEG metric. We construct the map $f_{m,k} : \hat{I}_m^{\oslash k} \to \mathsf{NEG}$ using $\oslash$-embedding of this map, with $\alpha = \frac{1}{k}$.

Using Theorem 4.6, it follows that the resulting embedding is in NEG.

Distortion Bound We defer the analysis of the distortion of the embedding constructed above to the Appendix (Section A.4).

Acknowledgments The authors would like to thank James Lee for all the helpful discussions. The authors would also like to thank Sanjeev Arora and Madhur Tulsiani for their constructive comments.

References

[1] Amit Agarwal, Moses Charikar, Konstantin Makarychev, and Yury Makarychev. $O(\sqrt{\log n})$ approximation algorithms for Min UnCut, Min 2CNF deletion, and directed cut problems. In *STOC'05: Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 573–581. ACM, New York, 2005.

[2] Sanjeev Arora, James R. Lee, and Assaf Naor. Euclidean distortion and the sparsest cut. *J. Amer. Math. Soc.*, 21:1–21, 2008.

[3] Sanjeev Arora, Satish Rao, and Umesh Vazirani. Expander flows, geometric embeddings and graph partitioning. *J. ACM*, 56(2):Art. 5, 37, 2009.

[4] Moses Charikar, Konstantin Makarychev, and Yury Makarychev. Near-optimal algorithms for unique games (extended abstract). In *STOC'06: Proceedings of the 38th Annual ACM Symposium on Theory of Computing*, pages 205–214. ACM, New York, 2006.

[5] Shuchi Chawla, Anupam Gupta, and Harald Räcke. Embeddings of negative-type metrics and an improved approximation to generalized sparsest cut. *ACM Trans. Algorithms*, 4(2):Art. 22, 18, 2008.

[6] Jeff Cheeger and Bruce Kleiner. Generalized differential and bi-Lipschitz nonembedding in L_1. *C. R. Math. Acad. Sci. Paris*, 343(5):297–301, 2006.

[7] Jeff Cheeger and Bruce Kleiner. Differentiating maps into L^1, and the geometry of BV functions. *Ann. of Math. (2)*, 171(2):1347–1385, 2010.

[8] Jeff Cheeger and Bruce Kleiner. Metric differentiation, monotonicity and maps to L_1. *Inventiones Mathematicae*, 182:335–370, 2010. 10.1007/s00222-010-0264-9.

[9] Jeff Cheeger, Bruce Kleiner, and Assaf Naor. A $(\log n)^{\Omega(1)}$ integrality gap for the sparsest cut sdp. In *FOCS '09: Proceedings of the 2009 50th Annual IEEE Symposium on Foundations of Computer Science*, pages 555–564, Washington, DC, USA, 2009. IEEE Computer Society.

[10] George Karakostas. A better approximation ratio for the vertex cover problem. *ACM Trans. Algorithms*, 5(4):1–8, 2009.

[11] S.A. Khot and N.K. Vishnoi. The unique games conjecture, integrality gap for cut problems and embeddability of negative type metrics into ℓ_1. In *46th Annual Symposium on Foundations of Computer Science*, pages 53–62. IEEE Computer Soc., Los Alamitos, CA, 2005.

[12] Robert Krauthgamer and Yuval Rabani. Improved lower bounds for embeddings into L_1. In *SODA '06: Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1010–1017, New York, NY, USA, 2006. ACM Press.

[13] James R. Lee and Mohammad Moharrami. Bilipschitz snowflakes and metrics of negative type. 2010.

[14] James R. Lee and Mohammad Moharrami. Bilipschitz snowflakes and metrics of negative type. In *STOC '10: Proceedings of the 42nd ACM symposium on Theory of computing*, pages 621–630, New York, NY, USA, 2010. ACM.

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

[16] James R. Lee and Anastasios Sidiropoulos, Mohammad. Near-optimal distortion bounds for embedding doubling spaces into l_1. In *STOC '11: Proceedings of the 43nd ACM symposium on Theory of computing; to appear.*

[17] Jiri Matoušek. Open problems on embeddings of finite metric spaces. *Available at http://kam.mff.cuni.cz/∼matousek/haifaop.ps*, 2002.

A Proofs

A.1 Lower bound for NEG embedding

Lemma A.1 (Lemma 3.4 in [13]) *For $i, j \in [m]$ and a non-expanding embedding $f : V(H_m) \to \mathsf{NEG}$ with distortion at most D,*

$$\langle f([x]_i) - f([\bar{x}]_i), f([x]_j) - f([\bar{x}]_j)\rangle \geq \frac{m}{D} - 4|i-j|.$$

Proof: We have,

$$\begin{aligned}
\langle f([x]_i) - f([\bar{x}]_i), f([x]_j) - f([\bar{x}]_j)\rangle = & \ \frac{1}{2}(\|f([x]_i) - f([\bar{x}]_i)\|^2 + \|f([x]_j) - f([\bar{x}]_j)\|^2 \\
& -\|f([x]_i) - f([\bar{x}]_i) - (f([x]_j) - f([\bar{x}]_j))\|^2) \\
\geq & \ \frac{1}{2}\left(\frac{m}{D} + \frac{m}{D} - \|f([x]_i) - f([\bar{x}]_i) - (f([x]_j) - f([\bar{x}]_j))\|^2\right) \\
= & \ \frac{m}{D} - \frac{1}{2}\|f([x]_i) - f([x]_j) + (f([\bar{x}]_j) - f([\bar{x}]_i))\|^2 \\
\geq & \ \frac{m}{D} - \|f([x]_i) - f([x]_j)\|^2 - \|f([\bar{x}]_j) - f([\bar{x}]_i)\|^2 \\
\geq & \ \frac{m}{D} - 4|i-j|.
\end{aligned}$$

where the first inequality follows because f has distortion D and the last one follows because f is non-expanding. ■

■

An immediate corollary of the above lemma is Lemma 3.1, which has been restated below.

Lemma A.2 ([13], Lemma 3.1 restated) *For any non-expanding map $f : V(H_m) \to \mathsf{NEG}$ with distortion at most D and $i, j \in [m]$ such that $|i-j| \leq \frac{m}{8D}$,*

$$\langle f([x]_i) - f([\bar{x}]_i), f([x]_j) - f([\bar{x}]_j)\rangle \geq \frac{m}{2D}.$$

Lemma A.3 (Lemma 3.2 restated) *For any ε-efficient non-expanding map $f : V(H_m) \to \mathsf{NEG}$, and for any $l \leq \frac{m}{2}$, there exists an index $p \in [m]$ such that:*

$$\mathop{\mathbf{E}}_{\vec{e} \in E(Q_m)} \mathop{\mathbf{E}}_{i,j \in [p, p+l-1]} \langle f([\vec{e}]_i), f([\vec{e}]_j)\rangle = O\left(l^{-1} + \varepsilon\right).$$

Proof: Fix $x \in Q_m$ and $k \in [m]$. Consider a path $[x]_i, [x \oplus e_k]_i, [x]_{i+1}, \ldots, [x]_j$, and let $\vec{e} = (x, x \oplus e_k)$, $f([\vec{e}]_i) = f([x \oplus e_k]_i) - f([x]_i)$. We have the following,

$$\begin{aligned}
\|f([x]_i) - f([x]_j)\|^2 \ = & \ \|f([x]_i) - f([x \oplus e_k]_i)\|^2 + \|f([x \oplus e_k]_i) - f([x]_j)\|^2 \\
& +2\langle f([x]_i) - f([x \oplus e_k]_i), f([x \oplus e_k]_i) - f([x]_j)\rangle \\
\leq & \ \|f([x]_i) - f([x \oplus e_k]_i)\|^2 + \|f([x \oplus e_k]_i) - f([x]_j)\|^2 \\
& -2\langle f([x \oplus e_k]_i) - f([x]_i), f([x \oplus e_k]_j) - f([x]_j)\rangle \\
\leq & \ \|f([x]_i) - f([x \oplus e_k]_i)\|^2 + \|f([x \oplus e_k]_i) - f([x]_{i+1})\|^2 \\
& +\|f([x]_{i+1}) - f([x]_j)\|^2 - 2\langle f([\vec{e}]_i), f([\vec{e}]_j)\rangle.
\end{aligned}$$

where both the inequalities follow from triangle inequality. For a fixed j, we obtain the following inequality by induction,

$$\|f([x]_p) - f([x]_j)\|^2 \leq \sum_{i=p}^{j-1} \left(\|f([x]_i) - f([x \oplus e_k]_i)\|^2 + \|f([x \oplus e_k]_i) - f([x]_{i+1})\|^2 \right.$$
$$\left. -2 \left\langle f([\vec{e}]_i), f([\vec{e}]_j)\right\rangle\right).$$

A similar calculation yields the following inequality,

$$\|f([x \oplus e_k]_j) - f([x \oplus e_k]_{p+l-1})\|^2 \leq \sum_{i=j+1}^{p+l-1} (\|f([x \oplus e_k]_{i-1}) - f([x]_i)\|^2$$
$$+\|f([x]_i) - f([x \oplus e_k]_i)\|^2 - 2 \left\langle f([\vec{e}]_i), f([\vec{e}]_j)\right\rangle).$$

Using both inequalities, we get

$$\begin{aligned}
\|f([x]_p) - f([x \oplus e_k]_{p+l-1})\|^2 &\leq \|f([x]_p) - f([x]_j)\|^2 + \|f([x]_j) - f([x \oplus e_k]_j)\|^2 \\
&\quad + \|f([x \oplus e_k]_j) - f([x \oplus e_k]_{p+l-1})\|^2 \\
&\leq \sum_{i=p}^{p+l-1} \|f([x]_i) - f([x \oplus e_k]_i)\|^2 + \sum_{i=p}^{p+l-2} \|f([x \oplus e_k]_i) - f([x]_{i+1})\|^2 \\
&\quad -2 \sum_{i=p, i \neq j}^{p+l-1} \langle f([\vec{e}]_i), f([\vec{e}]_j)\rangle).
\end{aligned} \tag{3}$$

We partition the vertices of H_m into $\left\lceil \frac{m}{l} \right\rceil$ blocks, each of which consists of at most l consecutive hypercubes. Thus the starting indices for these blocks would be $P = \{1, l+1, 2l+1, \ldots, l \left\lceil \frac{m}{l} \right\rceil - 2l + 1\}$. Note that we ignore the last block which may contain fewer than l hypercubes. For the sake of brevity, let $r = l \left\lceil \frac{m}{l} \right\rceil - 2l + 1$.

Now, for an s-t shortest path from μ_m, we consider the expression $\|f(s) - f(t)\|^2$ and use triangle inequality to obtain a sum over terms $\|f([x]_p) - f([x \oplus e_k]_{p+l-1})\|^2$.

$$\|f(s) - f(t)\|^2 \leq \|f(s) - f([x]_1)\|^2 + \sum_{p \in P} \|f([x]_p) - f([x \oplus e_k]_{p+l-1})\|^2$$
$$+ \sum_{p \in P} \|f([x \oplus e_k]_{p+l-1}) - f([x]_{p+l})\|^2 + \|f([x \oplus e_k]_{r-1}) - f(t)\|^2.$$

We use the Inequality (3) to bound $\sum_{p \in P} \|f([x]_p) - f([x \oplus e_k]_{p+l-1})\|^2$. By averaging over all such shortest paths from μ_m, and using the ε-efficiency condition, we obtain the following inequality,

$$\mathop{\mathbf{E}}_{\gamma \sim \mu_m} \|f(s) - f(t)\|^2 \leq (1+\varepsilon) \mathop{\mathbf{E}}_{\gamma \sim \mu_m} \|f(s) - f(t)\|^2$$
$$- 2 \sum_{p \in P} \mathop{\mathbf{E}}_{\vec{e} \in \vec{E}(Q_m)} \sum_{i=p, i \neq p+j}^{p+l-1} \left\langle f([\vec{e}]_i), f([\vec{e}]_{p+j})\right\rangle.$$

Since the embedding is non-expanding, this implies,

$$\sum_{p \in P} \mathop{\mathbf{E}}_{\vec{e} \in \vec{E}(Q_m)} \sum_{i=p: i \neq p+j}^{p+l-1} \langle f([\vec{e}]_i), f([\vec{e}]_{p+j})\rangle \leq \varepsilon m.$$

Thus, for $l \leq \frac{m}{2}$, there exists a $p \in P$ such that

$$\mathop{\mathbf{E}}_{\vec{e} \in \vec{E}(Q_m)} \sum_{i=p, i \neq p+j}^{p+l-1} \langle f([\vec{e}]_i), f([\vec{e}]_{p+j}) \rangle \leq \frac{\varepsilon m}{\lceil \frac{m}{l} \rceil - 1} \leq 2\varepsilon l.$$

The embedding is non-expanding, therefore $\langle f[\vec{e}]_i, f[\vec{e}]_i \rangle \leq 1$. Adding the terms for $i = j$, we obtain

$$\mathop{\mathbf{E}}_{\vec{e} \in \vec{E}(Q_m)} \mathop{\mathbf{E}}_{i,j \in [p, p+l-1]} \langle f([\vec{e}]_i), f([\vec{e}]_j) \rangle \leq 2l^{-1} + 2\varepsilon.$$

■

■

A.2 Embedding the base graph

Lemma A.4 (Lemma 4.1) *Suppose that G is graph. For any non-expanding map $f : V(G) \to L_1$, we have*

$$\mathsf{dist}(\mathcal{E}_s f) \leq 5 \cdot \mathsf{dist}(f).$$

Proof: We first verify that $\mathcal{E}_s f$ is non-expanding. If u and $v \in P_{(x,y)}$ of G, then

$$\|\mathcal{E}_s f(u) - \mathcal{E}_s f(v)\|_1 = \|f(x) - f(y)\|_1 \frac{d_G(u,v)}{d_G(x,y)} \geq \mathsf{dist}(f) d_G(u,v),$$

where $P_{(x,y)}$ is the path on the edge (x, y). Now, we consider the case that u and v are on different edges of G. Let u' and v' be the closest vertices of G to u and v, respectively. We have,

$$\begin{aligned}\frac{1}{2}\|\mathcal{E}_s f(u) - \mathcal{E}_s f(v)\|_1 \geq & \frac{1}{2}\left(\|\mathcal{E}_s f(u') - \mathcal{E}_s f(v')\|_1 - \|\mathcal{E}_s f(u) - \mathcal{E}_s f(u')\|_1\right. \\ & \left. - \|\mathcal{E}_s f(v) - \mathcal{E}_s f(v')\|_1\right).\end{aligned}$$

On the other hand,

$$\begin{aligned}2\|\mathcal{E}_s f(u) - \mathcal{E}_s f(v)\|_1 \geq & \|\mathcal{E}_s f(u) - \mathcal{E}_s f(u')\|_1 + \|\mathcal{E}_s f(v) - \mathcal{E}_s f(v')\|_1 \\ \geq & \mathsf{dist}(f)\left(d(u', u) + d(v, v')\right),\end{aligned}$$

hence,

$$\begin{aligned}\frac{5}{2}\|f(u) - f(v)\|_1 \geq & \frac{1}{2}(\|\mathcal{E}_s f(u') - \mathcal{E}_s f(v')\|_1 + \|\mathcal{E}_s f(u) - \mathcal{E}_s f(u')\|_1 \\ & + \|\mathcal{E}_s f(v) - \mathcal{E}_s f(v')\|_1) \\ \geq & \frac{\mathsf{dist}(f)}{2}\Big(d_G(u, u') + d_G(u', v') + d_G(v', v)\Big) \geq \frac{\mathsf{dist}(f)}{2} d_G(u, v).\end{aligned}$$

■

■

Lemma A.5 (Lemma 4.3 restated,[13]) *The map $\hat{g}_m$ is non-expanding and has distortion $O(\sqrt{m})$.*

Proof: We first bound the distortion of g_m, and then use Lemma 4.1 to bound $\mathsf{dist}(\hat{g}_m)$.

To show that g_m is non-expanding, it is enough to show that the map is non-expanding on all of the edges. For any given edge $(u,v) \in E(I_m)$, we have $d_{\mu_{\mathsf{ver}}}(u,v) \leq \frac{1}{2}\mathsf{len}_{I_m}(u,v)$. For all the vertices v in the first layer $\Pr_{S\in\eta_m}[\mathbf{1}_S(v) \neq \mathbf{1}_S(s)] \leq \frac{1}{4}$, hence

$$d_{\eta_m}(s',v) + d_{\mu_{\mathsf{ver}}}(s',v) = \frac{1}{4}(\frac{m}{2}) + \frac{m}{2} \leq m.$$

The same analysis would give the same bound for the distance between the last layer and t'. For edges between s and s', and between t and t' we have

$$d_{\eta_m}(t',t) + d_{\mu_{\mathsf{ver}}}(t',t) = d_{\eta_m}(s',s) + d_{\mu_{\mathsf{ver}}}(s',s) = \frac{1}{m^2}.$$

For all other edges let u be at layer i and v be at layer $i+1$. Any set $S_{c,k}$ that contain u also contains v. The only cuts that can separate u from v in η_m are the cuts such that $i-k \leq d_{Q_m}(v,c) \leq i-k+1$. The total weight of these cuts is at most $\frac{2}{2m}(\frac{m}{2}) \leq \frac{1}{2}$, therefore

$$d_{\eta_m}(u,v) + d_{\mu_{\mathsf{ver}}}(u,v) \leq \frac{1}{2} + \frac{1}{2}.$$

Now, we have to bound the contraction of pairs. We can lower bound the distance between any vertex $x \in I_m$ and s by $\frac{d_{m,1}(x,s)}{2}$, using only μ_{ver}. Similarly we can bound the distance from s', t', and t to other vertices. For all other vertices $[x]_i$ and $[y]_j$. Without loss of generality assume that x is in layer i' and y is in layer j', where $i' \leq j'$. We have

$$\begin{aligned}
d_{m,1}([x]_i,[y]_j) &\leq 2\cdot\max(|i-j|, d_{Q_m}(x,y)) \\
&\leq 2\cdot(|i-j| + d_{Q_m}(x,y)) \\
&\leq \Big(4d_{\mu_{\mathsf{ver}}}([x]_i,[y]_j) + d_{Q_m}(x,y)\Big)
\end{aligned}$$

Now, we only need to bound $d_{Q_m}(x,y)$ by $d_{\eta_m}([x]_i,[y]_j)$ to bound the total distortion.

$$\begin{aligned}
d_{\eta_m}([x]_i,[y]_j) &\geq \sum_{S_{c,k}\in\eta_m : [y]_j\in S_{c,k}, [x]_i \notin S_{c,k}} \frac{m}{2}\cdot\frac{1}{2m\cdot 2^m} \\
&\geq \sum_{c\in Q_m} \sum_{\substack{d_{Q_m}(c,x)-i'\geq k,\\ d_{Q_m}(c,y)-j'\leq k}} \frac{1}{4\cdot 2^m} \\
&\geq \sum_{c\in Q_m} \max(0, d_{Q_m}(c,y) - d_{Q_m}d(c,x))\cdot\frac{1}{4\cdot 2^m} \\
&= \frac{1}{2}\mathbb{E}_{c\in Q_m}\left[\frac{|d_{Q_m}(c,y) - d_{Q_m}d(c,x)|}{4}\right] \\
&= \Theta(\sqrt{m})
\end{aligned}$$

■

■

A.3 Embedding Composition

Theorem A.6 (Theorem 4.6 restated) *Let G and H be two graphs and suppose that f is the two-sum of the maps $g : G \to \mathsf{NEG}$ and $h : H \to \mathsf{NEG}$ on the edges $\vec{e}_1 = (u_1, v_1)$ and $\vec{e}_2 = (u_2, v_2)$, then image of f is also a negative type metric.*

Before we prove this lemma we prove the following simple geometric lemma,

Lemma A.7 *Let x_0, x_1, y, and z be four points in $\mathbb{R}^n$ such that $\|x_0-y\|_2^2+\|x_0-z\|_2^2-\|y-z\|_2^2 \geq 0$ and $\|x_1 - y\|_2^2 + \|x_1 - z\|_2^2 - \|y - z\|_2^2 \geq 0$. We define x_α as $\alpha(x_1) + (1 - \alpha)x_0$. For $0 \leq \alpha \leq 1$, we have:*

$$(x_\alpha - y)(x_\alpha - z) + \alpha(1 - \alpha)(x_0 - x_1)^2 \geq 0.$$

Proof: Let $f(\alpha) = (x_\alpha - y)(x_\alpha - z) + \alpha(1 - \alpha)(x_0 - x_1)^2$. We have that $f(0)$ and $f(1)$ are both non-negative. To prove this Lemma we show that f is a linear function and it assumes its minimum on one of the end points.

$$\begin{aligned} f(\alpha) &= \langle x_\alpha - y, x_\alpha - z\rangle + \alpha(1-\alpha)\|x_0 - x_1\|_2^2 \\ &= \langle x_0 + \alpha(x_1 - x_0) - y, x_0 + \alpha(x_1 - x_0) - z\rangle + \alpha(1-\alpha)\|x_0 - x_1\|_2^2 \\ &= C + 2\alpha\langle x_0, x_1 - x_0\rangle + \|\alpha(x_1 - x_0)\|_2^2 - \langle y + z, \alpha(x_1 - x_0)\rangle \\ &\quad +\alpha(1-\alpha)\|x_0 - x_1\|_2^2 \\ &= C + 2\alpha\langle x_0, x_1 - x_0\rangle - \langle y + z, \alpha(x_1 - x_0)\rangle + \alpha\|x_0 - x_1\|_2^2. \end{aligned}$$

Note that, in the above equation C is the constant part of the function that does not depend on α. ■

■

■

Proof: [Proof of Theorem A.6] We have to show that for all possible values of x, y and z in domain of f the inequality

$$\|f(x) - f(z)\|_2^2 + \|f(z) - f(y)\|_2^2 \geq \|f(x) - f(z)\|_2^2$$

holds. Instead of proving this inequality, we prove the following equivalent inequality,

$$\langle f(x) - f(y), f(z) - f(y)\rangle \geq 0.$$

Without loss of generality assume that $y \in V(G)$. If both $x, y \in V(G)$, then since the two-sum operation does not change the distances between pairs in $V(G)$ the inequality holds. If one of $x \in V(G)$ and $z \in V(H)$ we have,

$$\langle f(x) - f(y), f(z) - f(y)\rangle = \left\langle f(x) - f(y), \mathsf{proj}_{(f(u_2),f(v_2))}(f(z)) - f(y)\right\rangle.$$

We can write $\mathsf{proj}_{(f(u_2),f(v_2))}(f(z))$ as $\alpha f(u_2) + (1 - \alpha)f(v_2)$ for some $\alpha \in [0, 1]$. Therefore,

$$\begin{aligned} \langle f(x) - f(y), f(z) - f(y)\rangle &= \langle f(x) - f(y), \alpha f(u_2) + (1-\alpha)f(v_2)\rangle \\ &= (f(x) - f(y)) \cdot (\alpha f(u_1) + (1-\alpha)f(v_1) - f(y)) \\ &= \langle f(x) - f(y), \alpha f(u_1) + (1-\alpha)f(v_1) - f(y)\rangle \\ &= \alpha\langle f(x) - f(y), f(u_1) - f(y)\rangle \\ &\qquad +(1-\alpha)\langle f(x) - f(y), f(v_1) - f(y)\rangle \\ &\geq 0. \end{aligned}$$

Finally, we consider the case where $x, z \in V(H)$. Let $y' = \mathsf{proj}_{(f(u_1), f(v_1))}(y)$,

$$\begin{aligned} \langle f(x) - f(y), f(z) - f(y) \rangle &= \left\langle f(x) - (f(y) - y') - y', f(z) - (f(y) - y') - y' \right\rangle \\ &= \left\langle f(x) - y', f(z) - y' \right\rangle + \|f(y) - y'\|_2^2, \end{aligned}$$

where the second equality holds because

$$\left\langle f(y) - y', f(z) - y' \right\rangle = \left\langle f(y) - y', f(x) - y' \right\rangle = 0.$$

By Lemma A.7

$$\left\langle f(x) - y', f(z) - y' \right\rangle \geq \left\langle f(u_1) - y', f(v_1) - y' \right\rangle.$$

On the other hand, the inequality $\langle f(v_1) - f(y), f(u_1) - f(y) \rangle \geq 0$ holds, thus

$$\|(f(y) - y')\|_2^2 \geq \left\langle f(u_1) - y', y' - f(v_1) \right\rangle.$$

Hence,

$$\begin{aligned} \langle f(x) - f(y), f(z) - f(y) \rangle &= \left\langle f(x) - y', f(z) - y' \right\rangle + \|f(y) - y'\|_2^2 \\ &\geq \left\langle f(v_1) - y', f(u_1) - y' \right\rangle + \|f(y) - y'\|_2^2 \\ &\geq 0. \end{aligned}$$

■

■

A.4 Distortion Bound

Theorem A.8 (Theorem 2.10 restated) *There exists a map $f : \hat{I}_m^{\oslash k} \to \mathsf{NEG}$, such that $\mathsf{dist}(f) \lesssim \min(k, \sqrt{m})$.*

As a corollary of this theorem is that for every k, there exists an embedding of $\hat{I}_m^{\oslash k}$ into negative type metrics with distortion at most $O(\log^{1/3} n)$, where n is the number of vertices in the graph.

Before we prove Theorem A.8, we need to prove Lemma A.9. Using this lemma we only need to go back to the common ancestor of two vertices in the analysis, we can bound the expansion of the map.

Lemma A.9 *For any edge $\vec{e} \in \vec{E}_M(\hat{I}_m^{\oslash k})$, we have $\mathsf{len}_{\hat{I}_m}(e) \asymp \|f_{m,k}(\vec{e})\|_2^2$.*

After finding the common ancestor we divide a path in $\hat{I}_m^{\oslash k}$ into three parts. The two parts that are not part of the common ancestor are bounded using Lemma A.10. The other part is completely inside the common ancestor. We bound this part by distortion bound on a single copy of $\hat{I}_m$ and the bounds obtained for the other two parts.

Lemma A.10 *Let $x, y, z \in \mathbb{R}^n$ such that square of their distances satisfy triangle inequality. then,*

$$\min(\|x - z\|_2^2, \|y - z\|_2^2) \asymp \|z - \mathsf{proj}_{(x,y)}(z)\|_2^2.$$

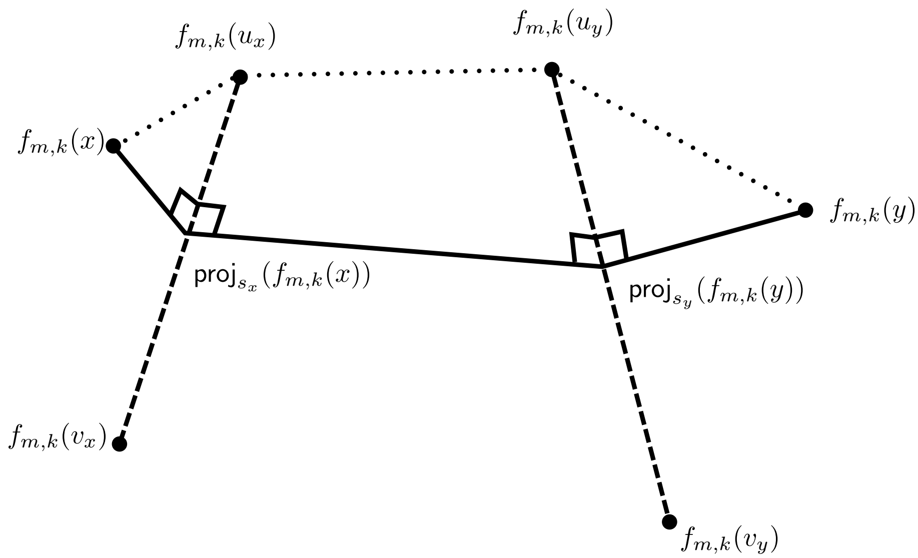

Figure 1: We bound $\|f_{m,k}(x) - \mathsf{proj}_{s_x}(f_{m,k}(x))\|_2$, $\|\mathsf{proj}_{s_y}(f_{m,k}(y)) - f_{m,k}(x)\|_2$, and $\|f_{m,k}(x) - f_{m,k}(y)\|_2$ separately.

Proof: [Proof of Theorem A.8]

First, we prove this Theorem for the case that one of the vertices is s. Let u_x be the closest vertex to x in G. We have,

$$\begin{aligned} d_{m,k}(x,s) \asymp d_{m,k}(u_x,s) &\gtrsim \min(k,\sqrt{m})\|f_{m,k}(u_x) - f_{m,k}(s)\|_2^2 \\ &\asymp \min(k,\sqrt{m})\|f_{m,k}(x) - f_{m,k}(s)\|_2^2. \end{aligned} \tag{4}$$

Now, we bound the contraction of $f_{m,k}$. For vertices $x, y \in \hat{I}_m^{\oslash k}$, Lemma A.9 shows that the common ancestor of x and y has a constant distortion. Let $\hat{I}_m$ be the common ancestor, and let x be on the edge (u_x, v_x) and y on the edge (u_y, v_y). Suppose that $\mathsf{len}_{\hat{I}_m^{\oslash k}}(u_y, y) \le \mathsf{len}_{\hat{I}_m^{\oslash k}}(v_y, y)$, and $\mathsf{len}_{\hat{I}_m^{\oslash k}}(u_x, x) \le \mathsf{len}_{\hat{I}_m^{\oslash k}}(v_x, x)$. Furthermore, let $s_x = (f_{m,k}(u_x), f_{m,k}(v_x))$, and $s_y = (f_{m,k}(u_y), f_{m,k}(v_y))$. We bound the distance between $f_{m,k}(x)$ and $f_{m,k}(y)$ by dividing it to three parts, and bound each part separately (see Figure 1).

$$\begin{aligned} \|f_{m,k}(x) - f_{m,k}(y)\|_2^2 &= \|f_{m,k}(x) - \mathsf{proj}_{s_x}(f_{m,k}(x))\|_2^2 \\ &+ \|f_{m,k}(y) - \mathsf{proj}_{s_y}(f_{m,k}(y))\|_2^2 + \|\mathsf{proj}_{s_y}(f_{m,k}(y)) - \mathsf{proj}_{s_x}(f_{m,k}(x))\|_2^2 \\ &\gtrsim \max(\|f_{m,k}(x) - \mathsf{proj}_{s_x}(f_{m,k}(x))\|_2^2 + \|f_{m,k}(y) - \mathsf{proj}_{s_y}(f_{m,k}(y))\|_2^2 \\ &\qquad , \|\mathsf{proj}_{s_y}(f_{m,k}(y)) - \mathsf{proj}_{s_x}(f_{m,k}(x))\|_2^2). \end{aligned}$$

We can bound $\|\mathsf{proj}_{s_y}(f_{m,k}(y)) - \mathsf{proj}_{s_x}(f_{m,k}(x))\|_2^2$ using the following inequality,

$$\begin{aligned} \|\mathsf{proj}_{s_y}(f_{m,k}(y)) - \mathsf{proj}_{s_x}(f_{m,k}(x))\|_2^2 &\ge \|f_{m,k}(u_y) - f_{m,k}(u_x)\|_2^2 \\ &- \|\mathsf{proj}_{s_y}(f_{m,k}(y)) - f_{m,k}(u_y)\|_2^2 - \|\mathsf{proj}_{s_x}(f_{m,k}(x)) - f_{m,k}(u_x)\|_2^2. \end{aligned}$$

We can bound the total length using Lemma A.10, and Inequality (4),

$$\begin{aligned}\min(k, \sqrt{m}) \cdot \|f_{m,k}(x) - f_{m,k}(y)\|_2^2 &\gtrsim \max\Big(d_{m,k}(u_x, x) + d_{m,k}(u_y, y),\\ &\qquad d_{m,k}(u_y, u_x) - O(d_{m,k}(u_x, x) + d_{m,k}(u_y, y))\Big)\\ &\gtrsim d_{m,k}(u_x, x) + d_{m,k}(u_y, y) + d_{m,k}(u_y, u_x).\end{aligned}$$

■

■

Now we present the proof for Lemma A.9.

Proof: [Proof of Lemma A.9] Let $f = \alpha \hat{f}_m \oplus (1-\alpha)\hat{g}'_m$, for all edges in $(u, v) \in E(\hat{I}_m)$, the following inequality holds,

$$\frac{\|f(u) - f(v)\|_1}{d_{m,1}(u, v)} = \left(1 + O(\frac{1}{k})\right) \frac{\|f(t) - f(s)\|_1}{d_{m,1}(s, t)}.$$

Therefore, we can bound expansion of each edge by

$$\left(1 + O(\frac{1}{k})\right)^k \asymp 1.$$

■

■

Proof: [Proof of Lemma A.10] Without loss of generality assume that $\|x - z\|_2^2 \leq \|y - z\|_2^2$. Since square of distances among x, y, z satisfy triangle inequality,

$$\begin{aligned}\|x - z\|_2^2 &= \|z - \mathsf{proj}_{(x,y)}(z)\|_2^2 + \|x - \mathsf{proj}_{(x,y)}(z)\|_2^2\\ &\leq \|z - \mathsf{proj}_{(x,y)}(z)\|_2^2 - \Big\langle x - \mathsf{proj}_{(x,y)}(z), y - \mathsf{proj}_{(x,y)}(z)\Big\rangle\\ &\leq 2\|x - \mathsf{proj}_{(x,y)}(z)\|_2^2.\end{aligned}$$

■

■