# Improved lower bounds for embeddings into $L_1$

[Extended Abstract]

Robert Krauthgamer* Yuval Rabani†

**Abstract**

We simplify and improve upon recent lower bounds on the minimum distortion of embedding certain finite metric spaces into $L_1$. In particular, we show that for infinitely many values of $n$ there are $n$-point metric spaces of negative type that require a distortion of $\Omega(\log\log n)$ for such an embedding, implying the same lower bound on the integrality gap of a well-known SDP relaxation for SPARSEST-CUT. This result builds upon and improves the recent lower bound of $(\log\log n)^{1/6-o(1)}$ due to Khot and Vishnoi [STOC 2005]. We also show that embedding the edit distance on $\{0,1\}^n$ into $L_1$ requires a distortion of $\Omega(\log n)$. This result simplifies and improves a very recent lower bound due to Khot and Naor [FOCS 2005].

## 1 Introduction

In recent years, low distortion embeddings of finite metric spaces into $L_1$ have become a powerful tool in an algorithm designer's arsenal. Such embeddings are extremely useful in two very different contexts, which we discuss below.

In combinatorial optimization, cuts in an $n$-vertex graph correspond to $n$-point cut semi-metrics.[1] These semi-metrics span the cone of $n$-point semi-metrics that are subsets of $L_1$. Polynomial-time computable relaxations of NP-hard cut problems are often expressed as optimization over larger sets of semi-metrics. Thus, using ("rounding") a relaxed solution to approximate the optimal solution to the original problem often boils down to embedding the relaxed solution into $L_1$.

In data analysis, proximity and classification problems are often easier to perform when data sets are subsets of $L_1$. In fact, for many such problems $L_1$ behaves as well as Euclidean space. Therefore, a common approach to solving such problems with other input families is first to embed the distance function into a well-behaved normed space such as $L_1$, and then to use known solutions for the chosen target space.

We consider in this paper the embedding into $L_1$ of two types of metrics that have attracted much attention recently, as we survey below. Firstly, we consider negative type metrics. A finite metric $d$ has *negative type* if and only if the metric that is derived by taking the square roots of the original distances is Euclidean (i.e., embeds isometrically into $L_2$). We show that such $n$-points metrics may require $\Omega(\log\log n)$ distortion to embed into $L_1$. Secondly, we consider the edit distance on $\{0,1\}^n$, which is the minimum number of character insert/delete/substitute operations required to transform one string into the other. We show that this metric requires $\Omega(\log n)$ distortion to embed into $L_1$. Both results are proved using simple tools from Fourier analysis of boolean functions. The lower bound for negative type metrics improves upon the $(\log\log n)^{1/6-o(1)}$ bound by Khot and Vishnoi [17].[2] The lower bound for edit distance improves upon the $(\log n)^{1/2-o(1)}$ lower bound by Khot and Naor [16]. Both previous bounds were proved very recently using Bourgain's deep result [8] on noise insensitive boolean functions. In contrast with the use of $\varepsilon$-noise in [17, 16], we use hypercube edges in both of our lower bound constructions. This is the main reason we can simplify the analysis considerably. As is turns out, our simpler constructions also lead to stronger bounds.

**Sparsest-cut and negative type metrics.** The most striking connection between $L_1$ embeddings and combinatorial approximation algorithms is exhibited by the sparsest-cut problem. In this problem the input is an undirected graph $G = (V, E)$ with positive edge capacities $c : E \to \mathbb{N}$, a set of pairs of vertices $D = \{s_i, t_i\}_{i=1}^k$ (called terminals or demand pairs), and a positive demand function $h : \{1, 2, \ldots, k\} \to \mathbb{N}$. The goal is to find a cut $(S, \overline{S})$ in $G$ (where $\emptyset \neq S \subsetneq V$,

*IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120, USA. Part of this work was done while visiting the Technion. Email: robi@almaden.ibm.com.

†Computer Science Department, Technion—Israel Institute of Technology, Haifa 32000, Israel. Part of this work was done while visiting IBM Almaden Research Center. Research was supported in part by ISF grant number 52/03 and BSF grant number 02-00282. Email: rabani@cs.technion.ac.il

[1] A semi-metric satisfies all the properties of a metric, except that points may be co-located.

[2] They also indicate that the bound improves to $(\log\log n)^{1/4-o(1)}$ using [21].

$S \cap \overline{S} = \emptyset$, and $S \cup \overline{S} = V$) that minimizes the ratio

$$\frac{\sum_{e \in E : |S \cap e| = 1} c(e)}{\sum_{i : |S \cap \{s_i, t_i\}| = 1} h(i)}.$$

In the uniform-demand case, $D$ consists of all pairs of vertices and $h(i) = 1$ for all $i$. Our lower bound requires a non-uniform choice of demands.

The ground-breaking work of [18] gave an $O(\log n)$-approximation for uniform demand sparsest-cut. The algorithm is based on a linear programming relaxation that produces a relaxed solution in the form of a (general) metric on the vertex set $V$, which is used to generate a cut by a region growing argument. Followup work applied this method to a similar relaxation for the non-uniform case. Eventually, tight $O(\log k)$ bounds (asymptotically matching the integrality gap) were derived in [4, 20] by relating the problem to embedding into $L_1$ and applying a theorem of Bourgain [7] that shows that every $n$-point metric space embeds into $L_1$ with distortion $O(\log n)$.

An obvious direction for improving the approximation guarantee is to use a more restricted set of metrics in the relaxation. A natural candidate for this is the set of $n$-point negative type metrics, which includes all $n$-point subsets of $L_1$ and can be optimized over in polynomial time using semidefinite programming. Formally, associate with every vertex $x \in V$ with a vector $v_x$, then the relaxation can be stated as given in Figure 1.

It has been conjectured that this relaxation can be used to get a constant factor approximation. A recent breakthrough result [3] obtained an $O(\sqrt{\log n})$ approximation for the uniform demand case. Followup work [9, 2] led to a bound of $O(\sqrt{\log k \log\log k})$ for the non-uniform case by showing an embedding of negative type metrics into $L_1$. (In fact, the embedding is into $L_2$, which embeds isometrically into $L_1$.)

On the other hand, two recent results [10, 17] independently proved that under Khot's unique games conjecture [15] the sparsest-cut problem is NP-hard to approximate within any constant factor. Quantitatively stronger versions of the conjecture imply inapproximability factors of up to $\Omega(\log\log n)$. The second paper [17] further showed that the integrality ratio of the above semidefinite relation can be as bad as $(\log\log N)^{1/6-o(1)}$, exhibiting in particular an $n$-point negative type metric that requires a distortion of $(\log\log N)^{1/6-o(1)}$ to embed into $L_1$. We prove the following theorem.

Theorem 1.1. *The semidefinite relaxation in Figure 1 has integrality ratio* $\Omega(\log\log n)$.

Using standard arguments (see [17]), this theorem implies the following corollary.

Corollary 1.1. *There exist infinitely many positive integers $n$ and $n$-point negative type metric spaces that require a distortion of* $\Omega(\log\log n)$ *to embed into* $L_1$.

Our proof of Theorem 1.1 is based on the argument in [17]. Our input instance differs in the choice of edges—we use the regular hypercube edges instead of the $\varepsilon$-noise model used in [17]. As a result, we are able to simplify some of the analysis, using Friedgut's approximation of low average sensitivity boolean functions by juntas [12]. We point out that our construction does *not* yield a strong integrality ratio for the related unique games semidefinite relaxation. In fact, the relaxation for the "underlying" unique games instance has (fractional) value of about $1 - O(1/\log n)$, whereas there is no integral solution of value $1 - \Omega(\log\log n / \log n)$. This fact is somewhat surprising, in view of the approach and motivation for the original argument in [17]. In particular, it indicates that our results might not be attained by applying the reduction in [10] to the unique games integrality ratio instance of [17].

**Edit distance.** The *edit distance* (a.k.a. Levenshtein distance [19]) between two strings is the minimum number of character insertions, deletions, and substitutions needed to transform one string to the other. Edit distance is a fundamental measure of similarity between strings and appropriately weighted variants of it play a central role in several domains of data analysis. In particular, efficient algorithms for dealing with variations of the edit distance (often referred to as sequence alignment) are among the most investigated computational problems in molecular biology, dating back to 1970 [22].

We restrict our discussion to strings in $\{0,1\}^n$, focusing on the asymptotic behavior as $n$ goes to infinity. For $x, y \in \{0,1\}^n$, let $\mathrm{ed}(x, y)$ denote their edit distance. It is a simple observation that $(\{0,1\}^n, \mathrm{ed})$ forms a metric space. Our concern is with low distortion embeddings of this metric space into $L_1$, motivated, as explained above, by applications in data analysis such as near-neighbor searching.

Until recently, no non-trivial bounds were known for embedding edit distance into $L_1$. In fact, isometric embedding was only recently ruled out by the $\frac{3}{2}$ lower bound of [1]. On the other hand, an $L_1$ embedding of edit distance with distortion bounded by $2^{O(\sqrt{\log n \log\log n})}$ was recently achieved in [23], and this indeed yields the best approximate nearest neighbor scheme currently known for edit distance. (An upper bound of $O(n^{2/3})$ can be derived from the results in [5].) Lastly, the lower bound was significantly strengthened to $(\log n)^{1/2-o(1)}$ in [16], the paper that motivated our work on edit distance. We show the following theorem.

Theorem 1.2. *The distortion of embedding the metric*

$$
\begin{array}{lll}
\text{Minimize} & \displaystyle\sum_{e=\{x,y\}\in E} c(e)(v_x - v_y)^2 & \\
\text{s.t.} & (v_x - v_z)^2 \le (v_x - v_y)^2 + (v_y - v_z)^2 & , \quad \forall x, y, z \in V \\
& \displaystyle\sum_{\{s_i,t_i\}\in D} h(i)(v_{s_i} - v_{t_i})^2 = 1 &
\end{array}
$$

Figure 1: A semidefinite programming relaxation for sparsest-cut

*space* $(\{0,1\}^n, \mathrm{ed})$ *into* $L_1$ *is* $\Omega(\log n)$.

Although inspired by the lower bound of [16], our proof of the theorem is shorter and simpler, and requires more basic machinery, namely, a result due to Kahn, Kalai, and Linial [14]. Technically, the simplification is obtained by analyzing *separately* the two different types of edit operations, namely, a random bit flip and a cyclic shift (with equal weights). For edit distance, this simplification leads to a stronger distortion lower bound because it can be applied also to a random bit flip (i.e. the hyberpcube edges) compared with the $\varepsilon$-noise used in [16]. More generally, it shows an effective way to exploit "soft" symmetry constraints to derive nonembeddability into $L_1$, compared with the hard constraints used in [17].

## 2 Integrality ratio for sparsest-cut

Our proof actually analyzes the (non-uniform demand) balanced cut problem instead of analyzing the sparsest cut problem. In the $B$-balanced cut problem, the input is similar to that of the sparsest cut problem with an additional parameter $0 < B < 1$. The goal is to find a minimum weight subset of the edges $F \subseteq E$ whose removal disconnects at least $B$-fraction of the demand. The above-mentioned semidefinite relaxation for sparsest cut can be adapted to $B$-balanced cut as given in Figure 2. (As we do not need the full generality of the problem, we restrict our attention to the case of unit edge capacities and unit, though non-uniform, demands.)

A standard argument, which can be found in [10, 17], shows that for all $0 < B' < B \le 1$, if there is an $n$-vertex instance in which the size of every $B'$-balanced cut exceeds by at least a $\rho$ factor the value of the $B$-balanced cut relaxation in Figure 2, then the integrality ratio of the sparsest cut relaxation in Figure 1 on $n$-vertex instances is at least $\Omega(\rho/(B - B'))$. Therefore, our goal in the rest of this section is to show an instance with a gap of $\rho = \Omega(\log \log N)$ with $B = 1/2$ and $B' = 1/4$.

We will first define a graph $G^*$ that is based on a folding of the $2n$-dimensional hypercube, and define for it a set of demand pairs $D$ to make it an instance of $\frac{1}{2}$-balanced cut. Next, we will demonstrate a solution of the semidefinite relaxation in Figure 2 for this instance. The solution vectors will be identical to those of [17], so the solution's feasibility follows from the proof in that paper, and we will only need to evaluate the value of this solution. (Notice that the set of feasible solutions is independent of the structure of the input graph, which is why we can use the same solution as in [17], even though we use a different graph on the same set of vertices.) Finally, we will lower bound the cost of a 1/4-balanced cut, which will conclude the proof of Theorem 1.1.

*Note*: We employ the notation used in an early version of [17]. It is substantially different than the newer version of [17].

**2.1 The instance** Fix an integer $k > 0$ and set $n = 2^k$. Let $\mathcal{F}$ be the family of all functions $f : \{-1,+1\}^k \mapsto \{-1,+1\}$, hence $|\mathcal{F}| = 2^n$. For $S \subseteq [k]$, define

$$\chi_S : \{-1,+1\}^k \mapsto \{-1,+1\}, \quad \chi_S(z) = \prod_{i\in S} z_i.$$

For $g, h \in \mathcal{F}$, let $gh$ denote the function that is the point-wise multiplication of $g$ and $h$. For $z \in \{-1,+1\}^k$ and $S \subseteq [k]$, let $z \circ S$ denote the vector obtained from $z$ by flipping the bits in $S$. For $g \in \mathcal{F}$ and $S \subseteq [k]$, define

$$g \circ S : \{-1,+1\}^k \mapsto \{-1,+1\}, \quad (g \circ S)(z) = g(z \circ S).$$

Notice that for every $S \subseteq [k]$, both $f \mapsto f\chi_S$ and $g \mapsto g \circ S$ are one-to-one functions.

For $f, g, f', g' \in \mathcal{F}$, write $(f, g) \equiv (f', g')$ if there exists $S \subseteq [k]$ such that $f' = f\chi_S$ and $g' = g \circ S$. It is easy to see that $\equiv$ is an equivalence relation on $\mathcal{F} \times \mathcal{F}$, and the size of every equivalence class is exactly $2^k = n$. Let $[(f, g)]$ denote the equivalence class of $(f, g)$.

Let $d_{\mathcal{H}}(f, f')$ denote the Hamming distance between $f, f' \in \mathcal{F}$, i.e., the number of $z \in \{-1,+1\}^k$ such that $f(z) \ne f'(z)$. Notice that for all $f \in \mathcal{F}$ and $\emptyset \ne S \subseteq [k]$, we have $d_{\mathcal{H}}(f, f\chi_S) = n/2$.

We are now ready to define the graph $G^* = (V^*, E^*)$. The vertex set is

$$V(G^*) = \{[(f, g)] : f, g \in \mathcal{F}\},$$

$$\begin{array}{lll} \text{Minimize} & \frac{1}{4}\sum_{(i,j)\in E}(v_i - v_j)^2 & \\ \text{s.t.} & (v_i)^2 = 1 & , \quad \forall i \in V \\ & (v_i - v_k)^2 \le (v_i - v_j)^2 + (v_j - v_k)^2 & , \quad \forall i,j,k \in V \\ & \frac{1}{4}\sum_{\{i,j\}\in D}(v_i - v_j)^2 \ge B \cdot |D| & \end{array}$$

Figure 2: A semidefinite programming relaxation for $B$-balanced cut

i.e., the collection of equivalence classes in $\mathcal{F}\times\mathcal{F}$, hence $|V(G^*)| = 2^{2n}/n$. The edge set is

$$E(G^*) = \Big\{(u,u') : \exists (f,g) \in u, \exists (f',g') \in u', \text{ s.t. } d_{\mathcal{H}}(f,f') + d_{\mathcal{H}}(g,g') = 1\Big\}$$

i.e., two vertices $u, v \in V(G^*)$ are connected by an edge if there exists a representation $u = [(f,g)]$ and $u' = [(f',g')]$ such that either (i) $f = f'$ and the functions $g, g'$ differ in exactly one coordinate, or (ii) $g = g'$ and the functions $f, f'$ differ in exactly one coordinate. It is easy to verify that for all $f, f' \in \mathcal{F}$ and $S \subseteq [k]$, we have $d_{\mathcal{H}}(f,f') = d_{\mathcal{H}}(f\chi_S, f'\chi_S)$ and $d_{\mathcal{H}}(f,f') = d_{\mathcal{H}}(f\circ S, f'\circ S)$. Thus, all vertices in $G^*$ have degree $2n$.

Similarly to [17], we define the demands to be uniform inside every $n$-dimensional cube induced by a single $f \in \mathcal{F}$, i.e., $D = \{(u,u') : \exists[(f,g)] \in u, \exists[(f',g')] \in u', f = f'\}$. We note that we could have also defined them similarly to [10] to be $D = \{(u,u') : \exists f, g \in \mathcal{F}, u = [(f,g)], u' = [(f,-g)]\}$.

Remark: The graph $G^*$ is the $2n$-dimensional hypercube with the standard edges (i.e., connecting vertices at Hamming distance 1), folded by merging together every vertex $(f,g)$ with $(f\chi_S, g\circ S)$ for every $S \subseteq [k]$. It is instructive to think of $f \in \mathcal{F}$ and $g \in \mathcal{F}$ as vectors in $\{-1,+1\}^n$. The map $f \mapsto f\chi_S$ then corresponds to flipping $n/2$ bits in $f$ (as determined by $S$), and the $n$ flip patterns generate the Hadamard code. The map $g \mapsto g\circ S$ corresponds to permuting the coordinates of $g$ (as determined by $S$), and the $n$ permutations form a 1-transitive group, meaning that for every $i, j \in [n]$ there is $S \subseteq [k]$ such that $g_j = (g\circ S)_i$.

### 2.2 The solution to the semidefinite relaxation

For a function $f \in \mathcal{F}$, define $\psi(f) = \left(\frac{f(z)}{\sqrt{n}}\right)_{z\in\{-1,+1\}^n}$ to be the truth table of $f$ viewed as a unit-length vector in $\mathbb{R}^n$. We associate every $f, g \in \mathcal{F}$ with the following vector in $\mathbb{R}^n$

$$\varphi(f,g) = \left(\frac{1}{\sqrt{n}}\sum_{S\subseteq[k]} g(1\circ S)\psi(f\chi_S)^{\otimes s}\right)^{\otimes t}.$$

where $s = 8$ and $t = 2^{240} + 1$. It can be easily verified that for all $f, g \in \mathcal{F}$ and $S \subseteq [k]$, we have $\varphi(f,g) = \varphi(f\chi_S, g\circ S)$, because $(g\circ T)(1\circ S) = g((1\circ S)\circ T) = g(1\circ(S\triangle T))$ and $(f\chi_T)\chi_S = f\chi_{S\triangle T}$. Therefore, $\varphi$ can be viewed as a function from $V(G^*)$ to $\mathbb{R}^n$, and we can take $(\varphi(u))_{u\in V(G^*)}$ as a solution to the semidefinite relaxation in Figure 2. Since these vectors are exactly the vectors used as the semidefinite solution in [17], it follows from that paper that they satisfy all the relaxation's constraints. However, our choice of edges is different and thus the value of this semidefinite solution is different.

Lemma 2.1. *Let $f, g, f', g \in \mathcal{F}$ and suppose that $d_{\mathcal{H}}(f,f') + d_{\mathcal{H}}(g,g') = 1$. Then*

$$\frac{1}{n}\sum_{S,S'\subseteq[k]} g(1\circ S)g'(1\circ S')(\psi(f\chi_S)\cdot\psi(f'\chi_{S'}))^s \ge 1 - O(s/2^n).$$

*Proof.* It is easily seen that $\{\psi(f\chi_S)\}_{S\subseteq[k]}$ is an orthonormal basis of $\mathbb{R}^n$. For every $h \in \mathcal{F}$, we also have $\psi(h) \in \mathbb{R}^n$, and thus $\sum_{S\subseteq[k]}(\psi(f\chi_S)\cdot\psi(h))^2 = \|\psi(h)\|^2 = 1$.

Suppose first that $g = g'$ and $d_{\mathcal{H}}(f,f') = 1$, and fix $S \subseteq [k]$. We then have $d_{\mathcal{H}}(f\chi_S, f'\chi_S) = 1$, and thus $(\psi(f\chi_S)\cdot\psi(f'\chi_S)) = 1 - 2/n$. It follows that

$$\begin{aligned} \sum_{S'\subseteq[k], S'\neq S}(\psi(f\chi_S\cdot\psi(f'\chi_{S'}))^s &\le (1-(1-2/n)^2)^{s/2} \\ &\le (4/n + 4/n^2)^{s/2}. \end{aligned}$$

Thus, the LHS in the statement of the lemma is lower bounded by $(1-2/n)^2 - (4/n + 4/n^2)^{s/2} \ge 1 - O(s/n)$.

Suppose next that $f = f'$ and $d_{\mathcal{H}}(g,g') = 1$. It follows that for all $S', S \subseteq [k]$ with $S' \neq S$ we have $(\psi(f\chi_S)\cdot\psi(f'\chi_{S'}))=0$. Thus, the LHS in the statement of the lemma is given by $\frac{1}{n}\sum_{S\subseteq[k]} g(1\circ S)g'(1\circ S)=1-2/n$.

LEMMA 2.2. *The solution* $(\varphi(u))_{u\in V(G^*)}$ *for the semidefinite relaxation in Figure 2 has value* $O(st/n)\cdot|E(G^*)|$.

*Proof.* Let $(u,u')\in E(G^*)$. Let $f,g,f',g'\in\mathcal{F}$ be such that $u=[(f,g)]$, $u'=[(f',g')]$, and $d_{\mathcal{H}}(f,f')+d_{\mathcal{H}}(g,g')=1$. By Lemma 2.1, $\|\varphi(u)-\varphi(u')\|^2=2-2(\varphi(u)\cdot\varphi(u'))\le 2-2(1-O(s/n))^t\le O(st/n)$, and we conclude that $\sum_{(u,u')\in E(G^*)}\|\varphi(u)-\varphi(u')\|^2\le O(st/n)\cdot|E(G^*)|$.

**2.3 Balanced cuts** We use the following theorem of Friedgut. (The *average sensitivity* (aka total influence) of a boolean function $b:\{0,1\}^n\to\{0,1\}$ is $\sum_{i=1}^n \Pr_{x\in\{0,1\}^n}[f(x)\neq f(x^i)]$, where $x^i\in\{0,1\}$ is derived from $x$ by flipping the $i$-th bit.)

THEOREM 2.1. (FRIEDGUT [12]) *Let* $b:\{0,1\}^n\to\{0,1\}$ *be a boolean function with average sensitivity* $k$. *Let* $\epsilon>0$ *and let* $M=\frac{k}{\epsilon}$. *Then, there exists a boolean function* $b':\{0,1\}^n\to\{0,1\}$ *depending only on* $2^{O((2+\sqrt{2\log(4M)/M})M)}$ *variables, such that* $b'$ *differs from* $b$ *on at most* $\epsilon 2^n$ *inputs.*

The following lemma shows that every $\frac{1}{4}$-balanced cut in $G^*$ cuts $\Omega(\frac{\log n}{n})|E(G^*)|$ edges.

LEMMA 2.3. *Let* $A:V(G^*)\mapsto\{0,1\}$ *be a function such that* $\Pr_{(u,u')\in D}[A(u)\neq A(v)]\ge\frac{1}{4}$. *Then* $\Pr_{(u,u')\in E(G^*)}[A(u)\neq A(v)]\ge\Omega(\frac{\log n}{n})$.

*Proof.* Let $G$ be the $2n$-dimensional hypercube whose vertex set is $V(G)=\mathcal{F}\times\mathcal{F}$ and edge set is $E(G)=\{((f,g),(f',g')) : d_{\mathcal{H}}(f,f')+d_{\mathcal{H}}(g,g')=1\}$. To see that this is a hypercube, write every function $f\in\mathcal{F}$ as a vector in $\{-1,+1\}^n$. Let us set up the terminology for the $2n$ variables (dimensions) of this hypercube. In particular, we associate each of the last $n$ variables (i.e., the $g$-variables when a vertex in $V(G)$ is written as $(f,g)$) with a distinct set $S\subseteq[k]$, as follows: Fix a function $\sigma_0:\{-1,+1\}^k\mapsto\{-1,+1\}$ that takes the value $-1$ exactly once, and for every $S\subset[k]$, let $\sigma_S:\{-1,+1\}^k\mapsto\{-1,+1\}$ be the function $\sigma_S=\sigma_0\circ S$. Notice that $\{\sigma_S\}_{S\subseteq[k]}$ is the family of functions in $\mathcal{F}$ that take the value $-1$ exactly once, and thus $d_{\mathcal{H}}((f,g),(f,g\sigma_S))=1$.

We extend the function $A:V(G^*)\mapsto\{0,1\}$ in a straightforward way to a function $B:V(G)\mapsto\{0,1\}$ by defining $B(u)=A([u])$. By definition, for all $f\in\mathcal{F},g\in\mathcal{F},S\subseteq[k]$ we have $B(f,g)=B(f\chi_S,g\circ S)$. It is easily seen that

$$\Pr_{(u,u')\in E(G)}[B(u)\neq B(u')]=\Pr_{(u,u')\in E(G^*)}[A(u)\neq A(u')].$$

Assume, towards contradiction, that this quantity is at most $\frac{c\log n}{n}$, where $c>0$ is a constant that will be determined shortly. Observe that $B$ is $\frac{1}{4}$-balanced because

$$\begin{aligned}\Pr_{u,u'\in V(G)}[B(u)\neq B(u')] &\ge \Pr_{f,g,g'\in\mathcal{F}}[B(f,g)\neq B(f,g')]\\ &= \Pr_{(u,u')\in D}[A(u)\neq A(u')]\\ &\ge \frac{1}{4}.\end{aligned}$$

Thus, using Theorem 2.1 we get that $B$ can be $\varepsilon$-approximated by a junta function $\hat{B}$ that depends only on variables in $J$ for $|J|\le 2^{O(c\log n/\varepsilon)}$. Fixing $\varepsilon=\frac{1}{20}$, we see that if $c>0$ is a sufficiently small constant then $|J|\le n^{1/3}$. Let $J_g$ be the set of $g$-variables in $J$. For $T\subseteq[k]$, define $J_g\circ T$ to contain the variable indexed by $S\triangle T$ whenever $J_g$ contains the variable indexed by $S$.

We first claim that there exists $S^*\in[k]$ such that $J_g\cap(J_g\circ S)=\emptyset$. To prove the claim, observe that for two variables indexed by $T_1,T_2$ in $J_g$, there is only one $S\subset[k]$, namely $S=T_1\triangle T_2$, such that $T_2\in J\cap(J\circ S)$. Hence, at most $n^{2/3}$ choices of $S\in[k]$ may yields $J\circ S$ that has a non-empty intersection with $J$. The claim follows since the number of choices for $S$ is $2^k=n$.

We now wish to upper bound $\Pr_{(u,u')\in D}[A(u)\neq A(u')]=\Pr_{f,g,h\in\mathcal{F}}[B(f,g)\neq B(f,gh)]$. To this end, notice that every $h\in\mathcal{F}$ can be written as $h=h_Jh_{\bar{J}}$, where $h_J\in\mathcal{F}$ is negative only on $g$-coordinates in $J$, and $h_{\bar{J}}\in\mathcal{F}$ is negative only on $g$-coordinates in $\bar{J}$. Thus,

$$\begin{aligned}&\Pr_{f,g,h\in\mathcal{F}}[B(f,g)\neq B(f,gh)]\\ \le\ &\Pr_{f,g,h_{\bar{J}}}[B(f,g)\neq B(f,gh_{\bar{J}})]\\ &\quad+\Pr_{f,g,h_{\bar{J}},h_J}[B(f,gh_{\bar{J}})\neq B(f,gh)]\\ \le\ & 2\varepsilon+\Pr_{f,g,h_J}[B(f,g)\neq B(f,gh_J)]\\ =\ & 2\varepsilon+\Pr_{f,g,h_J}[B(f\chi_S,g\circ S)\neq B(f\chi_S,(gh_J)\circ S)]\\ \le\ & 4\varepsilon;\end{aligned}$$

the second inequality holds because $h_{\bar{J}}$ is negative only in variables that $\hat{B}$ does not depend on; the third inequality holds because for all $f\in\mathcal{F},g\in\mathcal{F},S\subseteq[k]$, we have by definition $B(f,g)=B(f\chi_S,g\circ S)$, the last inequality follows by noticing that $(gh_J)\circ S=(g\circ S)h_{J\circ S}$ differs from $g\circ S$ in coordinates that are all not in $J$.

We conclude that the fraction of demands separated by this cut is $\Pr_{(u,u')\in D}[A(u) \neq A(u')] \leq 4\varepsilon = \frac{1}{5}$, contradicting our assumption that the cut is $\frac{1}{4}$-balanced.

Combining Lemmas 2.2 and 2.3 we see that every $\frac{1}{4}$-balanced cut exceeds by at least a $\Omega(\log\log N)$ factor the value of the $\frac{1}{2}$-balanced cut semidefinite relaxation in Figure 2. We conclude that the integrality ratio of the semidefinite relaxation in Figure 1 is $\Omega(\log\log N)$, completing the proof of Theorem 1.1.

## 3 Lower bound for embedding edit distance into $L_1$

To prove Theorem 1.2, fix an integer $n$ and let $V = \{0,1\}^n$. We will need a few weight functions over $V \times V$; for simplicity, we define them as probability distributions. First, let

$$E_H = \{(x,y) : x, y \in V,\ \|x-y\|_1 = 1\},$$

and let $\tau_H$ be a probability distribution over $V \times V$ that has a uniform support over $E_H$ and assigns probability 0 for pairs not in $E_H$. Next, define $S : \{0,1\}^n \to \{0,1\}^n$ to be the cyclic left-shift operation, i.e. $S(x_1, \ldots, x_n) = (x_2, \ldots, x_n, x_1)$, let

$$E_S = \{(x, S(x)) : x \in V\},$$

and let $\tau_S$ be a probability distribution over $V \times V$ that has a uniform support over $E_S$ and assigns probability 0 for pairs not in $E_S$. Let the distribution $\tau$ be the average of $\tau_H$ and $\tau_S$, i.e. $\tau(x,y) = (\tau_H(x,y) + \tau_S(x,y))/2$. Let $\sigma$ be the uniform distribution over $V$. For $A \subseteq V$, let $\bar{A} = V \setminus A$ and let $\Lambda(A) = (A \times \bar{A}) \cup (\bar{A} \times A)$ be the collection of pairs "crossing" the cut $(A, \bar{A})$. A straightforward counting argument (see [6, Lemma 8] and [16, Lemma 4.4]) shows that for two strings $x, y \in V$ drawn independently at random, $\mathrm{ed}(x,y) \geq \Omega(n)$ with probability $\Omega(1)$. We thus get

$$\frac{\mathbb{E}_{(x,y)\in\tau}[\mathrm{ed}(x,y)]}{\mathbb{E}_{x\in\sigma,y\in\sigma}[\mathrm{ed}(x,y)]} \leq \frac{2}{\Omega(n)} \leq O(1/n).$$

Using the cut cone representation of $L_1$ metrics (see [4, 20, 11]), if $(V, \mathrm{ed})$ embeds into $L_1$ with distortion $D > 0$ then there must exist $A \subseteq V$ such that

$$\frac{\tau(\Lambda(A))}{\sigma(A)\sigma(\bar{A})} \leq O(D/n).$$

The following key lemma would then complete the proof of Theorem 1.2.

Lemma 3.1. *For every $A \subseteq V$,*

$$\tau(\Lambda(A)) \geq \Omega\left(\frac{\log n}{n}\right)\sigma(A)\sigma(\bar{A}).$$

*Proof.* Fix $A \subseteq V$, and assume without loss of generality that $|A| \leq |\bar{A}|$, i.e. $\sigma(A) \leq 1/2$. Define accordingly a boolean function $f : \{0,1\}^n \to \{0,1\}$ by $f(x) = 1_{\{x\in A\}}$. Let $I_j$ be the influence of the $j$-th variable in $f$, i.e., $I_j = \Pr_{x\in V}[f(x) \neq f(x \oplus e_j)]$, where $e_j$ is the $j$-th unit vector and $\oplus$ represents coordinate-wise addition modulo 2. We shall soon require the following bound that is implicit in [14]; see also [13, Lemma 3.4] or [10, Lemma 2.3] for details.

Lemma 3.2. (Kahn, Kalai, and Linial [14]) *Let $f : \{0,1\}^n \to \{0,1\}$ be a boolean function with balance $p = \Pr_{x\in\{0,1\}^n}[f(x) = 1] \leq 1/2$, and let $I_j = \Pr_{x\in\{0,1\}^n}[f(x) \neq f(x \oplus e_j)]$ be the influence of the $j$-th variable. Then for all $\delta > 0$,*

$$\max_{j\in[n]} I_j \leq \delta \quad \Rightarrow \quad \sum_{j\in[n]} I_j \geq \Omega(p)\log(1/\delta).$$

Let $c > 0$ be a constant to be determined later, and assume towards contradiction that $\tau(\Lambda(A)) < \frac{c\log n}{n} \cdot \sigma(A)$. Observe that the total influence of $f$ is

$$\sum_{j\in[n]} I_j = n\tau_H(\Lambda(A)) \leq 2n\tau(\Lambda(A)) < 2c\log n \cdot \sigma(A),$$

and its balance is $\sigma(A)$. We thus get from Lemma 3.2 that there exists $l \in [n]$ such that $I_l \geq 1/n^{1/8}$, if only $c > 0$ is chosen to be a sufficiently small (depending only on the hidden constant in Lemma 3.2).

We now claim that for every $k \in \{1, \ldots, n^{1/4}\}$, $I_{l+k} \geq 1/(2n^{1/8})$. Indeed, by our assumption above,

$$\begin{aligned}
\Pr_{x\in V}[f(x) \neq f(S(x))] &= \tau_S(\Lambda(A)) \\
&\leq 2\tau(\Lambda(A)) \\
&\leq \tfrac{2c\log n}{n} \cdot \sigma(A) \\
&\leq \tfrac{c\log n}{n}.
\end{aligned}$$

Observe that if $x$ is chosen uniformly at random from $V$ then $S(x)$ is also uniformly distributed over $V$. For every $k \in [n^{1/4}]$ (the constant 1/4 is somewhat arbitrary), it thus follows by a union bound that

$$\begin{aligned}
\Pr_{x\in V}[f(x) \neq f(S^k(x))] &\leq \sum_{i=0}^{k-1} \Pr_{x\in V}[f(S^i(x)) \neq f(S^{i+1}(x))] \\
&\leq \tfrac{ck\log n}{n} \\
&\ll n^{-1/2}.
\end{aligned}$$

Observe that $S^k(x) \oplus e_l = S^k(x \oplus e_{l+k})$, and that if $x$ has uniform distribution over $V$ then so does $x + e_{l+k}$;

thus, the last inequality implies that

$$\begin{aligned} I_j &= \Pr_{x\in V}[f(S^k(x)) \neq f(S^k(x) \oplus e_j)] \\ &\leq \Pr_{x\in V}[f(S^k(x)) \neq f(x)] + \Pr_{x\in V}[f(x) \neq f(x \oplus e_{l+k})] \\ &\quad + \Pr_{x\in V}[f(x \oplus e_{l+k}) \neq f(S^k(x \oplus e_{l+k}))] \\ &\leq I_{l+k} + 2/n^{1/2}. \end{aligned}$$

The claim now follows from the bound $I_l \geq 1/n^{1/8}$.

Finally, notice that by the above claim, the total influence is at least $\sum_{k=1}^{n^{1/4}} I_{l+k} \geq \frac{1}{2} n^{1/8} \gg \frac{c \log n}{n}$, which contradicts our assumption above and completes the proof of Lemma 3.1.

## 4 Discussion

One obvious challenge left open from Section 3 is to determine $c_1(\{0,1\}^n, \mathrm{ed})$, as there is still a large gap between the upper bound of [23] and the lower bound of Theorem 1.2. We note however that Lemma 3.1 is tight: There exists $A \subseteq V$ such that $1/4 \leq \sigma(A) \leq 3/4$ and $\tau(\Lambda(A)) \leq O(\frac{\log n}{n})$. It is also interesting to note that there exists a collection $V'$ containing $1 - o(1)$ fraction of the strings in $\{0,1\}^n$, such that $c_1(V', \mathrm{ed}) = \Theta(\log n)$. The lower bound follows from our proof above; details are omitted from this version.

It was pointed out by Assaf Naor that Lemma 3.1 can be cast as the following Poincaré inequality: For every $f : \{0,1\}^n \to L_1$,

$$\sum_j \mathop{\mathbb{E}}_{x\in\sigma} \|f(x) - f(x \oplus e_j)\|_1 + n \mathop{\mathbb{E}}_{x\in\sigma} ||f(x) - f(S(x))||_1 \geq \Omega(\log n) \mathop{\mathbb{E}}_{x\in\sigma, y\in\sigma} \|f(x) - f(y)\|_1.$$

Furthermore, it can be generalized to some other operations on the coordinates (other than the cyclic shift).

Another intriguing question is the least distortion for embedding $(\{0,1\}^n, \mathrm{ed})$ into $L_2$-squared (equivalently, embedding $(\{0,1\}^n, \sqrt{\mathrm{ed}})$ into $L_2$). While a squared-$L_2$ embedding is technically weaker than an $L_1$-embedding (i.e., it follows from but generally does not imply an $L_1$-embedding), it is as useful for many applications like Nearest Neighbor Search. We remark that an exact analogue to Lemma 3.1 (based on same $\tau$ and $\sigma$) is not true, because letting $h(x)$ be the Hamming weight of $x$ we obtain $\mathbb{E}_{(x,y)\in\tau} |h(x) - h(y)|^2 \leq O(1/n)\, \mathbb{E}_{x\in\sigma, y\in\sigma} |h(x) - h(y)|^2$.

Finally, we note that our results do not give a lower bound for several related problems, such as embedding into $L_1$ of the Ulam metric or of the edit distance with moves.

**Acknowledgments.** We thank James R. Lee, Elchanan Mossel, Assaf Naor and the anonymous reviewers for useful comments. The first author also wishes to thank the participants of the Spring 2005 Berkeley theory reading group for stimulating discussions that partly led to this research.

## References

[1] A. Andoni, M. Deza, A. Gupta, P. Indyk, and S. Raskhodnikova. Lower bounds for embedding edit distance into normed spaces. In *Proceedings of the 14th annual ACM-SIAM Symposium on Discrete Algorithms*, pages 523–526, 2003.

[2] S. Arora, J. R. Lee, and A. Naor. Euclidean distortion and the sparsest cut. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, 2005.

[3] S. Arora, S. Rao, and U. Vazirani. Expander flows, geometric embeddings, and graph partitionings. In *36th Annual Symposium on the Theory of Computing*, pages 222–231, May 2004.

[4] Y. Aumann and Y. Rabani. An $O(\log k)$ approximate min-cut max-flow theorem and approximation algorithm. *SIAM J. Comput.*, 27(1):291–301, 1998.

[5] Z. Bar-Yossef, T. S. Jayram, R. Krauthgamer, and R. Kumar. Approximating edit distance efficiently. In *45th IEEE Symposium on Foundations of Computer Science*, pages 550–559. IEEE Computer Soc., 2004.

[6] T. Batu, F. Ergün, J. Kilian, A. Magen, S. Raskhodnikova, R. Rubinfeld, and R. Sami. A sublinear algorithm for weakly approximating edit distance. In *35th Annual ACM Symposium on Theory of Computing*, pages 316–324, 2003.

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

[8] J. Bourgain. On the distributions of the Fourier spectrum of Boolean functions. *Israel J. Math.*, 131:269–276, 2002.

[9] S. Chawla, A. Gupta, and H. Räcke. Improved approximations to sparsest cut. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 102–111, 2005.

[10] S. Chawla, R. Krauthgamer, R. Kumar, Y. Rabani, and D. Sivakumar. On the hardness of approximating multicut and sparsest-cut. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, pages 144–153, June 2005.

[11] M. M. Deza and M. Laurent. *Geometry of cuts and metrics.* Springer-Verlag, Berlin, 1997.

[12] E. Friedgut. Boolean functions with low average sensitivity depend on few coordinates. *Combinatorica*, 18(1):27–35, 1998.

[13] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124(10):2993–3002, 1996.

[14] J. Kahn, G. Kalai, and N. Linial. The influence of variables on boolean functions. In *Proceedings of the 29th IEEE Symposium on Foundations of Computer Science*, pages 68–80, 1988.

[15] S. Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 767–775, 2002.

[16] S. Khot and A. Naor. Nonembeddability theorems via Fourier Analysis. In *46th IEEE Symposium on Foundations of Computer Science*, 2005. To appear.

[17] S. Khot and N. K. Vishnoi. The unique games conjecture, integrality gap for cut problems and the embeddability of negative type metrics into $l_1$. In *46th IEEE Symposium on Foundations of Computer Science*, 2005. To appear.

[18] T. Leighton and S. Rao. Multicommodity max-flow min-cut theorems and their use in designing approximation algorithms. *J. ACM*, 46(6):787–832, 1999.

[19] V. I. Levenshtein. Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Dokl.*, 10:707–710, 1965.

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

[21] E. Mossel, R. O'Donnell, and K. Oleszkiewicz. Noise stability of functions with low influences: Invariance and optimality. In *Proceedings of the 46th IEEE Symposium on Foundations of Computer Science*, 2005. To appear.

[22] S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48(3):443–453, 1970.

[23] R. Ostrovsky and Y. Rabani. Low distortion embeddings for edit distance. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing*, pages 218–226, 2005.