Near-Neighbor Preserving Dimension Reduction for Doubling Subsets of L1

Ioannis Z. Emiris, Vasilis Margonis, Ioannis Psarros

Near-Neighbor Preserving Dimension Reduction for Doubling Subsets of ℓ_1

Ioannis Z. Emiris
Department of Informatics & Telecommunications,
National & Kapodistrian University of Athens, Greece
ATHENA Research & Innovation Center, Greece
emiris@di.uoa.gr

Vasilis Margonis
Department of Informatics & Telecommunications,
National & Kapodistrian University of Athens, Greece
basilis.math@gmail.com

Ioannis Psarros[1]
Institute of Computer Science, University of Bonn, Germany
ipsarros@uni-bonn.de

Abstract

Randomized dimensionality reduction has been recognized as one of the fundamental techniques in handling high-dimensional data. Starting with the celebrated Johnson-Lindenstrauss Lemma, such reductions have been studied in depth for the Euclidean (ℓ_2) metric, but much less for the Manhattan (ℓ_1) metric. Our primary motivation is the approximate nearest neighbor problem in ℓ_1. We exploit its reduction to the decision-with-witness version, called approximate *near* neighbor, which incurs a roughly logarithmic overhead. In 2007, Indyk and Naor, in the context of approximate nearest neighbors, introduced the notion of nearest neighbor-preserving embeddings. These are randomized embeddings between two metric spaces with guaranteed bounded distortion only for the distances between a query point and a point set. Such embeddings are known to exist for both ℓ_2 and ℓ_1 metrics, as well as for doubling subsets of ℓ_2. The case that remained open were doubling subsets of ℓ_1. In this paper, we propose a dimension reduction by means of a *near* neighbor-preserving embedding for doubling subsets of ℓ_1. Our approach is to represent the pointset with a carefully chosen covering set, then randomly project the latter. We study two types of covering sets: c-approximate r-nets and randomly shifted grids, and we discuss the tradeoff between them in terms of preprocessing time and target dimension. We employ Cauchy variables: certain concentration bounds derived should be of independent interest.

2012 ACM Subject Classification Theory of computation → Nearest neighbor algorithms; Mathematics of computing → Dimensionality reduction

Keywords and phrases Approximate nearest neighbor, Manhattan metric, randomized embedding

Digital Object Identifier 10.4230/LIPIcs.APPROX-RANDOM.2019.47

Category RANDOM

Related Version A preliminary version is available at `https://arxiv.org/abs/1902.08815`.

Funding *Ioannis Z. Emiris*: Partially supported by the European Union's H2020 research and innovation programme under grant agreement No. 734242 (LAMBDA).
Ioannis Psarros: Generously supported by the Hausdorff Center for Mathematics.

Acknowledgements IZE is member of team AROMATH, joint between INRIA Sophia-Antipolis and NKUA. IP thanks Robert Krauthgamer for useful discussions on the topic.

[1] This work was done while the third author was a PhD candidate in the Department of Informatics & Telecommunications, National & Kapodistrian University of Athens, Greece.

1 Introduction

Proximity search is a fundamental computational problem with several applications in Computer Science and beyond. Proximity problems in metric spaces of low dimension have been typically handled by methods which discretize the space and therefore are affected by the curse of dimensionality, making them unfit for high-dimensional spaces. In the past two decades, the increasing need for analyzing high-dimensional data led researchers to devise randomized and approximation algorithms with polynomial dependence on the dimension.

A fundamental proximity problem is Approximate Nearest Neighbor search. By known reductions [11], one can (up to polylogarithmic factors) focus on the decision version with witness, namely the (c, R)-Approximate Near Neighbor problem:

▶ **Definition 1** (Approximate Near Neighbor). *Let (X, d_X) be a metric space. Given $P \subseteq X$ and reals $R > 0$, $c \geq 1$, build a data structure $\mathcal{S}$ that, given a query point $q \in X$, performs as follows:*

- *If the nearest neighbor of q lies within distance at most R, then $\mathcal{S}$ is allowed to report any point $p^* \in P$ such that $d_X(q, p^*) \leq cR$.*
- *If all points lie at distance more than cR from q, then $\mathcal{S}$ should return $\perp$.*

In general, $\mathcal{S}$ returns either a point at distance $\leq cR$ or $\perp$, even when none of the above two cases occurs.

From now on, we assume $R = 1$ because we can re-scale the data set, and we refer to this problem as c-ANN, or simply ANN. We focus on subsets of ℓ_1^d: the input dataset consists of n vectors in $\mathbb{R}^d$ and the distance function is the standard ℓ_1 norm $\|\cdot\|_1$. Note that all logarithms are base 2.

Previous work. Some highlights in the study of data structures for high-dimensional normed spaces are the various variants, proofs, and applications of the Johnson Lindenstrauss Lemma (e.g. [1, 2, 3]), sketches based on p-stable distributions [14], and Locality Sensitive Hashing (e.g. [15, 4, 5]). In the core of most high-dimensional solutions lies the fact that for certain metric spaces e.g. $\ell_p, p \in [1, 2]$, the distance can be efficiently sketched. Spaces which are considered to be harder in this context, such as ℓ_∞, can also be treated [13], and are very interesting since they can be used as host spaces for various norms [6].

Significant amount of work has been undertaken for pointsets of low doubling dimension, since it is today one of the primary paradigms for capturing input structure (formal definitions in the next section). For any finite metric space X of doubling dimension $\dim(X)$, there exists a data structure [12, 9] with expected preprocessing time $O(2^{\dim(X)} n \log n)$, space usage $O(2^{\dim(X)} n)$ (or even $O(n)$) and query time $O(2^{\dim(X)} \log n + \varepsilon^{-O(\dim(X))})$.

In [16], they introduced the notion of nearest-neighbor preserving embeddings, and it was proven that in this context one can achieve dimension reduction for doubling subsets of ℓ_2, with the target dimension depending only on the dataset's doubling dimension. Even before, Indyk [14] had introduced a randomized embedding for dimension reduction in ℓ_1, which is suitable for proximity search purposes, and it achieves target dimension polylogarithmic in the size of the pointset. Naturally, such approaches can be easily combined with any known data structure to be used in the projection space. Randomized embeddings have been recently used in the ANN context [8], for doubling subsets of ℓ_p, $2 < p < \infty$.

It is known that dimension reduction in ℓ_1 cannot be achieved in the same generality as in ℓ_2, even assuming that the pointset is of low doubling dimension [18]: there are arbitrarily large n-point subsets $P \subseteq \ell_1$ which are doubling with constant 6, such that every embedding

with distortion D of P into ℓ_1^k requires dimension $n^{\Omega(1/D^2)}$. Aiming for more restrictive guarantees, e.g. preserving distances within some pre-defined range, is a relevant workaround. Then, dimension reduction techniques for doubling subsets of ℓ_p, $p \in [1,2]$, exist [7], but they rely on partition algorithms which require the whole pointset to be known in advance. Hence, applicability of such techniques is quite limited and, specifically, it is not clear whether they can be used in an online setting where query points are not known beforehand.

Contribution. In this paper, we establish two non-linear *near* neighbor-preserving embeddings for doubling subsets of ℓ_1^d. We use a definition which is essentially a modified version of the nearest neighbor preserving embedding of [16]: the guarantees which are required are weaker since we consider the decision version of the problem, therefore the embedding depends on some range parameter $R > 0$.

▶ **Definition 2** (Near-neighbor preserving embedding). *Let (Y, d_Y), (Z, d_Z) be metric spaces and $X \subseteq Y$. A distribution over mappings $f : Y \to Z$ is a* near-neighbor preserving embedding *with range $R > 0$, distortion $D \geq 1$ and probability of correctness $\mathcal{P} \in [0,1]$ if for every $\alpha \geq D$ and any $q \in Y$, if $x \in X$ is such that $d_Y(x,q) \leq R$, then with probability at least $\mathcal{P}$,*

- $d_Z(f(x), f(q)) \leq D \cdot R$,
- $\forall p \in X : \ d_Y(p,q) > D \cdot \alpha \cdot R \implies d_Z(f(p), f(q)) > \alpha \cdot R$.

Considering a pointset $P \subset \ell_1^d$ of cardinality n, our results concern ℓ_1^k as the target space, where k depends on the doubling dimension of P. We assume that $R = 1$, since we can rescale the dataset. More specifically:

1. In Theorem 10, we prove that for every $\varepsilon \in (0, 1/2)$ and $c \geq 1$, there is a randomized mapping $h : \ell_1^d \to \ell_1^k$ that can be computed in time $\tilde{O}(dn^{1+1/\Omega(c)})$ and is *near* neighbor-preserving for P with distortion $1+6\varepsilon$ and probability of correctness $\Omega(\varepsilon)$, where

$$k = (\log \lambda_P \cdot \log(c/\varepsilon))^{\Theta(1/\varepsilon)} / \zeta(\varepsilon),$$

for a function $\zeta(\varepsilon) > 0$ depending only on ε. Although the mapping h depends on the pointset, the parameter c is user-defined and therefore provides a trade-off between preprocessing time and target dimension.

2. In Theorem 13, we show that for every $\varepsilon \in (0, 1/2)$, there is a randomized mapping $h' : \ell_1^d \to \ell_1^k$ that can be computed in time $O(dkn)$ and is *near* neighbor-preserving for P with distortion $1+6\varepsilon$ and probability of correctness $\Omega(\varepsilon)$, where

$$k = (\log \lambda_P \cdot \log(d/\varepsilon))^{\Theta(1/\varepsilon)} / \zeta(\varepsilon),$$

for a function $\zeta(\varepsilon) > 0$ depending only on ε. In this case, the function h' is oblivious to P and well-defined over the whole space, but the target dimension depends on d.

On the low-preprocessing-time extreme, one can embed the dataset in near-linear time, but the target dimension is polynomial in $\log \log n$. This is to be juxtaposed to the analogous result by Indyk [14], which provides with target dimension polynomial in $\log n$, without any assumption on the doubling dimension of the dataset. On the other hand, one can obtain a preprocessing time of $dn^{1+\delta}$ for any constant $\delta > 0$, and target dimension which depends solely on the doubling dimension.

Techniques. Both embeddings consist of two basic components. First, we represent the pointset P with an ε-covering set, and then we apply a random linear projection à la Indyk [14] to that set, using Cauchy variables.

The role of the covering set is to exploit the doubling dimension of P. In the analogous result for ℓ_2 [16], no representative sets were used; the mapping was just a random linear projection of P. In the case of ℓ_1 however, a similar analysis of a linear projection with Cauchy variables without these representative sets seems to be impossible, since the Cauchy distribution is heavy tailed.

In Theorem 10, we consider c-approximate r-nets as a covering set. Inspired by the algorithm of [10] for ℓ_2, we design an algorithm that computes a c-approximate r-net in ℓ_1 in subquadratic –but superlinear– time. On the other hand, Theorem 13 relies on randomly shifted grids, which can be computed in linear time, but are inferior to nets in terms of capturing the doubling dimension of the pointset.

To bound the distortion incurred by the randomized projection, we exploit the 1-stability property of the Cauchy distribution. To this end, we prove a concentration bound for sums of independent Cauchy variables that should be of interest beyond the scope of this paper. To overcome the technical difficulties associated with the heavy tails of the Cauchy distribution, we study sums of *square roots* of Cauchy variables, where in [14], Indyk considers sums of *truncated* Cauchy variables instead. Although our concentration bound is rather weak, it is sufficient for our purposes and its analysis is much simpler compared to Indyk's.

Algorithmic implications. Our results show that efficient dimension reduction for doubling subsets of ℓ_1 is possible, in the context of ANN. In particular, these results imply efficient sketches, meaning that one can solve ANN with minimal storage per point. Dimension reduction also serves as a problem reduction from a high-dimensional hard instance to a low-dimensional easy instance. Since the algorithms presented in this paper are quite simple, they should also be of practical interest: they easily extend the scope of any implementation which has been optimized to solve the problem in low dimension, so that it may handle high-dimensional data.

Our embedding can be combined with the bucketing method of [11] for the $(1+\varepsilon)$-ANN problem in ℓ_1^d. For instance, setting $c = \log n$ in Theorem 10, yields preprocessing time $dn^{1+o(1)}$, space $n^{1+o(1)}$ and query time $O(d)\cdot(\log \lambda_P \cdot \log\log n)^{O(1/\varepsilon)}$ assuming that the doubling dimension is a fixed constant. This improves upon existing results: the query time of [17] depends on the aspect ratio of the dataset, while the data structures of [12, 9] support queries with time complexity which depends exponentially on the doubling dimension. However, it is worth noting that one could potentially improve the results of [17, 12, 9] in the special case of ℓ_1, by employing ANN data structures with fast query time, in order to accelerate the traversal of the net-tree. Hence, while our result gives a simple framework for exploiting the intrinsic dimension of doubling subsets of ℓ_1, it is unlikely that it shall improve upon simple variants of previous results in terms of complexity bounds.

Organization. The next section introduces basic concepts and some relevant existing results. Section 3 establishes a concentration bound on sums of independent Cauchy variables. Section 4, achieves dimensionality reduction by means of representing the pointset by a carefully chosen net, while Section 5 employs randomly shifted grids for the same task. We conclude with discussion of results and potential improvements.

2 Preliminaries

In this section, we define basic notions about doubling metrics and present useful previous results.

▶ **Definition 3.** *Consider any metric space (X, d_X) and let $B(p,r) = \{x \in X \mid \mathrm{d}_X(x,p) \leq r\}$. The* doubling constant *of X, denoted λ_X, is the smallest integer λ_X such that for any $p \in X$ and $r > 0$, the ball $B(p,r)$ can be covered by λ_X balls of radius $r/2$ centered at points in X.*

The *doubling dimension* of (X, d_X) is defined as $\log \lambda_X$. Nets play an important role in the study of embeddings, as well as in designing efficient data structures for doubling metrics.

▶ **Definition 4.** *For $c \geq 1$, $r > 0$ and metric space (V, d_V), a c-approximate r-net of V is a subset $\mathcal{N} \subseteq V$ such that no two points of $\mathcal{N}$ are within distance r of each other, and every point of V lies within distance at most $c \cdot r$ from some point of $\mathcal{N}$.*

▶ **Theorem 5.** *Let $P \subset \ell_1^d$ such that $|P| = n$. Then, for any $c > 0$, $r > 0$, one can compute a c-approximate r-net of P in time $\tilde{O}(dn^{1+1/c'})$, where $c' = \Omega(c)$. The result is correct with high probability. The algorithm also returns the assignment of each point of P to the point of the net which covers it.*

Proof. We employ some basic ideas from [11]. An analogous result for ℓ_2 is stated in [10]. First, we assume $r = 1$, since we are able to re-scale the point set. Now, we consider a randomly shifted grid with side-length 2. The probability that two points $p, q \in P$ fall into the same grid cell, is at least $1 - \|p-q\|_1/2$. For each non-empty grid cell we snap points to a grid: each coordinate is rounded to the nearest multiple of $\delta = 1/10dc$. Then, coordinates are multiplied by $1/\delta$ and each point $x = (x_1, \ldots, x_d) \in [2\delta]^d$ is mapped to $\{0,1\}^{2d/\delta}$ by a function G as follows: $G(x) = (g(x_1), \ldots, g(x_d))$, where $g(z)$ is a binary string of z ones followed by $2/\delta - z$ zeros. For any two points p, q in the same grid cell, let $f(p)$,$f(q)$ be the two binary strings obtained by the above mapping. Notice that,

$$\|f(p) - f(q)\|_1 \in (2/\delta) \cdot \|p-q\|_1 \pm 1.$$

Hence,

$$\|p-q\|_1 \leq 1 \implies \|f(p) - f(q)\|_1 \leq (2/\delta) + 1,$$

$$\|p-q\|_1 \geq c \implies \|f(p) - f(q)\|_1 \geq (2/\delta) \cdot c - 1.$$

Now, we employ the LSH family of [11], for the Hamming space. After standard concatenation, we can assume that the family is $(\rho, c'\rho, n^{-1/c'}, n^{-1})$-sensitive, where $\rho = (2/\delta) + 1$ and $c' = \Omega(c)$. Let $\alpha = n^{-1/c'}$ and $\beta = n^{-1}$.

Notice that for the above two-level hashing table we obtain the following guarantees. Any two points $p, q \in P$, such that $\|p-q\|_1 \leq 1$, fall into the same bucket with probability $\geq \alpha/2$. Any two points $p, q \in P$, such that $\|p-q\|_1 \geq c$, fall into the same bucket with probability $\leq \beta$.

Finally, we independently build $k = \Theta(n^{1/c'} \log n)$ hashtables as above, where the random hash function is defined as a concatenation of the function which maps points to their grid cell id and one LSH function. We pick an arbitrary ordering $p_1, \ldots, p_n \in P$. We follow a greedy strategy in order to compute the approximate net. We start with point p_1, and we add it to the net. We mark all (unmarked) points which fall at the same bucket with p_1, in one of the k hashtables, and are at distance $\leq cr$. Then, we proceed with point p_2. If p_2 is unmarked, then we repeat the above. Otherwise, we proceed with p_3. The above iteration stops when all points have been marked. Throughout the procedure, we are able to store one pointer for each point, indicating the center which covered it.

Correctness. The probability that a good pair p, q does not fall into the same bucket for any of the k hashtables is $\leq (1-\alpha/2)^k \leq n^{-10}$. Hence, with high probability, the packing property holds, and the covering property holds because the above algorithm stops when all points are marked.

Running time. The time to build the k hashtables is $k \cdot n = \tilde{O}(n^{1+1/c'})$. Then, at most n queries are performed: for each query, we investigate k buckets and the expected number of false positives is $\leq k \cdot n^2 \cdot \beta = \tilde{O}(n^{1+1/c'})$. Hence, if we stop after having seen a sufficient amount of false positives, we obtain time complexity $\tilde{O}(n^{1+1/c'})$ and the covering property holds with constant probability. We can repeat the above procedure $O(\log n)$ times to obtain high probability of success. ◀

The main result in the context of randomized embeddings for dimension reduction in ℓ_1^d is the following theorem, which exploits the 1-stability property of Cauchy random variables and provides with an asymmetric guarantee: The probability of non-contraction is high, but the probability of non-expansion is constant. Nevertheless, this asymmetric property is sufficient for proximity search.

▶ **Theorem 6** (Thm 5, [14]). *For any $\varepsilon \leq 1/2$, $\delta > 0$, $\varepsilon > \gamma > 0$ there is a probability space over linear mappings $f : \ell_1^d \to \ell_1^k$, where $k = (\ln(1/\delta))^{1/(\varepsilon-\gamma)}/\zeta(\gamma)$, for a function $\zeta(\gamma) > 0$ depending only on γ, such that for any pair of points $p, q \in \ell_1^d$:*

$$\mathbf{Pr}\left[\, \|f(p) - f(q)\|_1 \leq (1-\varepsilon)\, \|p-q\|_1 \,\right] \leq \delta,$$

$$\mathbf{Pr}\left[\, \|f(p) - f(q)\|_1 \geq (1+\varepsilon)\, \|p-q\|_1 \,\right] \leq \frac{1+\gamma}{1+\varepsilon}.$$

Note that the embedding is defined as $f(u) = Au/T$, where A is a $k \times d$ matrix with each element being an i.i.d. Cauchy random variable. In addition, T is a scaling factor defined as the expectation of a sum of truncated Cauchy variables, such that $T = \Theta(k \log(k/\varepsilon))$ (see Lemma 5 in [14]).

One key observation here is that given a pointset P in a space of bounded aspect ratio Φ, one can directly employ Theorem 6. The number of points can be upper bounded by a function of λ_P and Φ, and hence the new dimension, k, depends only on these parameters. This paper proves better bounds than the ones of Theorem 6 for doubling subsets of ℓ_1^d, without any assumption on the aspect ratio.

3 Concentration bounds for Cauchy variables

In this section, we prove some basic properties of the Cauchy distribution, which serves as our main embedding tool.

Let $C_{\mathcal{D}}$ denote the Cauchy distribution with density $c(x) = (1/\pi)/(1+x^2)$. One key property of the Cauchy distribution is the so-called 1-stability property: Let $v = (v_1, \ldots, v_k) \in \mathbb{R}^k$ and $X_1, \ldots, X_k$ be i.i.d. random variables following $C_{\mathcal{D}}$, then $\sum_{j=1}^k X_i v_i$ is distributed as $X \cdot \|v\|_1$, where $X \sim C_{\mathcal{D}}$.

The Cauchy distribution has undefined mean. However, for $0 < q < 1$, the mean of the q-th power of a Cauchy random variable can be defined. More specifically, for some $X \sim C_{\mathcal{D}}$ we have

$$\mathbb{E}\left[|X|^{1/2}\right] = \frac{2}{\pi}\int_0^\infty \frac{\sqrt{x}}{1+x^2}\,\mathrm{d}x = \frac{2}{\pi}\frac{\pi}{\sqrt{2}} = \sqrt{2}.$$

The following lemma provides a bound for the moment-generating function of $|X|^{1/2}$.

▶ **Lemma 7.** *Let $X \sim C_{\mathcal{D}}$. Then for any $\beta > 1$:*

$$\mathbb{E}\left[\exp\left(-\beta|X|^{1/2}\right)\right] \leq \frac{2}{\beta}.$$

Proof. For any constant β,

$$\int_0^1 \mathrm{e}^{-\beta x^{1/2}}\,\mathrm{d}x = \frac{2}{\beta^2}\left(1 - \frac{\beta+1}{\mathrm{e}^{\beta}}\right).$$

Then, for any $\beta > 1$,

$$\begin{aligned}
\mathbb{E}\left[\exp\left(-\beta|X|^{1/2}\right)\right] &= \int_{-\infty}^{\infty} \mathrm{e}^{-\beta|x|^{1/2}} \cdot c(x)\,\mathrm{d}x = \frac{2}{\pi}\int_0^{\infty} \mathrm{e}^{-\beta x^{1/2}} \cdot \frac{1}{1+x^2}\,\mathrm{d}x \\
&= \frac{2}{\pi}\int_0^{1} \mathrm{e}^{-\beta x^{1/2}} \cdot \frac{1}{1+x^2}\,\mathrm{d}x + \frac{2}{\pi}\int_1^{\infty} \mathrm{e}^{-\beta x^{1/2}} \cdot \frac{1}{1+x^2}\,\mathrm{d}x \\
&\leq \frac{2}{\pi}\int_0^{1} \mathrm{e}^{-\beta x^{1/2}}\,\mathrm{d}x + \frac{2}{\pi}\int_1^{\infty} \mathrm{e}^{-\beta} \cdot \frac{1}{1+x^2}\,\mathrm{d}x \\
&= \frac{2}{\pi}\cdot\frac{2}{\beta^2}\left(1 - \frac{\beta+1}{\mathrm{e}^{\beta}}\right) + \frac{1}{2\mathrm{e}^{\beta}} \\
&\leq \frac{4}{\pi\beta^2} + \frac{1}{2\mathrm{e}^{\beta}} \\
&\leq \frac{2}{\beta}.
\end{aligned}$$

◀

Let $S := \sum_{j=1}^{k} |X_j|$ where each X_j is an i.i.d. Cauchy variable. To prove concentration bounds for S, we study the sum $\tilde{S} := \sum_{j=1}^{k} |X_j|^{1/2}$. By Hölder's Inequality, for any $x \in \mathbb{R}^d$ and $p > q > 0$,

$$\|x\|_p \leq \|x\|_q \leq d^{1/q-1/p}\,\|x\|_p\,.$$

Consequently, for $x = (X_1, \ldots, X_k) \in \mathbb{R}^k$, $p = 1$ and $q = 1/2$ we have that $S \leq \tilde{S}^2 \leq k \cdot S$, hence for any $t > 0$,

$$\mathbf{Pr}[S \leq t] \leq \mathbf{Pr}[\tilde{S} \leq \sqrt{tk}]. \tag{1}$$

We use the bound on the moment-generating function, to prove a Chernoff-type concentration bound for $\tilde{S}$, which by Eq. (1) translates into a concentration bound for S.

▶ **Lemma 8.** *For every $D > 1$,*

$$\mathbf{Pr}\left[\tilde{S} \leq \frac{\mathbb{E}[\tilde{S}]}{D}\right] \leq \left(\frac{10}{D}\right)^k.$$

Proof. Since X_j's are independent, $\mathbb{E}[\tilde{S}] = \sqrt{2}k$. Then, by Lemma 7 and Markov's inequality, for any $\beta > 1$, it follows that

$$\begin{aligned}
\mathbf{Pr}\left[\tilde{S} \leq \frac{\mathbb{E}[\tilde{S}]}{D}\right] &= \mathbf{Pr}\left[\exp(-\beta\tilde{S}) \geq \exp\left(-\beta \cdot \frac{\mathbb{E}[\tilde{S}]}{D}\right)\right] \\
&\leq \frac{\mathbb{E}[\exp(-\beta\tilde{S})]}{\exp(-\beta\,\mathbb{E}[\tilde{S}]/D)} \\
&= \frac{\mathbb{E}[\exp(-\beta|X_j|^{1/2})]^k}{\exp(-\beta\sqrt{2}k/D)} \\
&\leq \left(\frac{2}{\beta}\right)^k \cdot \mathrm{e}^{\sqrt{2}\beta k/D}.
\end{aligned}$$

Setting $\beta = D$ completes the proof. ◀

4 Net-based dimension reduction

In this section we describe the dimension reduction mapping for ℓ_1 via r-nets. Let $P \subset \ell_1^d$ be a set of n points with doubling constant λ_P. For some point $x \in \mathbb{R}^d$ and $r > 0$, we denote by $B_1(x, r)$ the ℓ_1-ball of radius r around x. The embedding is non-linear and is carried out in two steps.

First, we compute a c-approximate (ε/c)-net $\mathcal{N}$ of P with the algorithm of Theorem 5. Moreover, the algorithm assigns each point of P to the point of $\mathcal{N}$ which covered it. Let $g : P \to \mathcal{N}$ be this assignment. In the second step, for every $s \in \mathcal{N}$ and any query point $q \in \ell_1^d$, we apply the linear map of Theorem 6. That is, $f(s) = As/T$, where A is a $k \times d$ matrix with each element being an i.i.d. Cauchy random variable. Recall that value $T = \Theta(k \log(k/\varepsilon))$. By the 1-stability property of the Cauchy distribution, $f(s)$ is distributed as $\|s\|_1 \cdot (Y_1, \ldots, Y_k)$, where each Y_j is i.i.d. and $Y_j \sim C_{\mathcal{D}}$. Hence, $\|f(s)\|_1 = \|s\|_1 \cdot S$ where $S := \sum_j |Y_j|$.

We define the embedding to be $h = f \circ g$. We apply h to every point in P, and f to any query point q. It is clear from the properties of the net that g incurs an additive error of $\pm\varepsilon$ on the distance between q and any point in P, so it is sufficient to consider the distortion of f.

Our analysis consists of studying separately the following disjoint subsets of $\mathcal{N}$: Points that lie at distance at most D_0 from the query and points that lie at distance at least D_0, for some $D_0 > 1$ chosen appropriately. For the former set, we directly apply Theorem 6, as it has bounded diameter.

The next lemma guarantees the low distortion for points of the latter set, namely those that are sufficiently far from the query. We consider the sum of the square roots of each $|Y_j|$, i.e., $\tilde{S} = \sum_j |Y_j|^{1/2}$, in order to employ the tools of Section 3.

▶ **Lemma 9.** *Fix a query point $q \in \ell_1^d$. For any $\varepsilon \le 1/2$, $c \ge 1$, $\delta \in (0,1)$, there exists $D_0 = O(\log(k/\varepsilon))$ such that for $k = \Theta\left(\log^2 \lambda_P \cdot \log(c/\varepsilon) + \log(1/\delta)\right)$, with probability at least $1 - \delta$,*

$$\forall s \in \mathcal{N} : \ \|s - q\|_1 \ge D_0 \implies \|f(s) - f(q))\|_1 \ge 4.$$

Proof. Assume wlog that the query point is the origin $(0, \ldots, 0)$. For some $D_0 > 1$, we define the following subsets of $\mathcal{N}$:

$$N_i := \{s \in \mathcal{N} \mid D_i \le \|s\|_1 < D_{i+1}\}, \ D_i = 2^{2i} D_0, \ i = 0, 1, 2, \ldots$$

By the definition of doubling constant and the fact that two points of $\mathcal{N}$ lie at distance at least ε,

$$|N_i| \le \lambda_P^{\lceil \log(4cD_{i+1}/\varepsilon) \rceil} \le \lambda_P^{4\log(cD_{i+1}/\varepsilon)}.$$

Therefore, by the union bound, and Eq. (1):

$$\begin{aligned}
\mathbf{Pr}\left[\exists i \exists s \in N_i : \|f(s)\|_1 \le \frac{4\|s\|_1}{D_i}\right] &= \mathbf{Pr}\left[\exists i \exists s \in N_i : S \le \frac{4T}{D_i}\right] \\
&\le \sum_{i=0}^{\infty} |N_i| \, \mathbf{Pr}\left[\tilde{S} \le \frac{\sqrt{4kT}}{\sqrt{D_i}}\right] \\
&= \sum_{i=0}^{\infty} |N_i| \, \mathbf{Pr}\left[\tilde{S} \le \mathbb{E}[\tilde{S}] \cdot \sqrt{\frac{2T}{k 2^{2i} D_0}}\right].
\end{aligned}$$

By Lemma 8, for $D_0{=}\lceil 800T/k\rceil{=}\Theta(\log(k/\varepsilon))$ and $k > 4{\cdot}\log\lambda_P{\cdot}\log(cD_0/\varepsilon) + 2\log(2\lambda_P/\delta)$:

$$\begin{aligned}
\sum_{i=0}^{\infty} |N_i|\,\mathbf{Pr}\left[\tilde{S} \le \frac{\mathbb{E}[\tilde{S}]}{10\cdot 2^{i+1}}\right] &\le \sum_{i=0}^{\infty} \lambda_P^{4\log(cD_{i+1}/\varepsilon)}\left(\frac{1}{2^{i+1}}\right)^k \\
&= \sum_{i=0}^{\infty} \frac{2^{\log(\lambda_P)(4\log(cD_0/\varepsilon)+2i+2)}}{2^{k(i+1)}} \\
&\le \sum_{i=0}^{\infty} \frac{2^{\log(\lambda_P)\cdot 4\log(cD_0/\varepsilon)}\cdot 2^{2\log(\lambda_P)(i+1)}}{2^{(4\cdot\log\lambda_P\cdot\log(cD_0/\varepsilon))(i+1)}\cdot 2^{2\log(2\lambda_P/\delta))(i+1)}} \\
&\le \sum_{i=0}^{\infty} 2^{-2\log(2/\delta))(i+1)} \\
&= \sum_{i=0}^{\infty}\left(\frac{\delta^2}{4}\right)^i - 1 \\
&= \frac{\delta^2}{4-\delta^2} \\
&\le \delta.
\end{aligned}$$

Finally, for some large enough constant C, we demand that

$$k > C\left(\log\lambda_P\cdot\log(c\log k/\varepsilon) + \log(1/\delta)\right) > 4\cdot\log\lambda_P\cdot\log(cD_0/\varepsilon) + 2\log(2\lambda_P/\delta)$$

which is satisfied for $k = \Theta\left(\log^2\lambda_P\cdot\log(c/\varepsilon) + \log(1/\delta)\right)$. ◀

▶ **Theorem 10.** *Let $P \subset \ell_1^d$ such that $|P| = n$. For any $\varepsilon \in (0,1/2)$ and $c \ge 1$, there is a non-linear randomized embedding $h = f\circ g : \ell_1^d \to \ell_1^k$, where $k = (\log\lambda_P\cdot\log(c/\varepsilon))^{\Theta(1/\varepsilon)}/\zeta(\varepsilon)$, for a function $\zeta(\varepsilon) > 0$ depending only on ε, such that, for any $q \in \ell_1^d$, if there exists $p^* \in P$ such that $\|p^* - q\|_1 \le 1$, then, with probability $\Omega(\varepsilon)$:*

$$\begin{aligned}
&\|h(p^*) - f(q)\|_1 \le 1 + 3\varepsilon, \\
&\forall p \in P:\ \|p-q\|_1 > 1+9\varepsilon \implies \|h(p) - f(q)\|_1 > 1 + 3\varepsilon.
\end{aligned}$$

Set P can be embedded in time $\tilde{O}(dn^{1+1/\Omega(c)})$, and any query $q \in \ell_1^d$ can be embedded in time $O(dk)$.

Proof. Let f, g be the mappings defined in the beginning of the section and $D_0 = \Theta(\log(k/\varepsilon))$. Assume wlog for simplicity that $q = 0^d$. Then, by Lemma 9 for $k = \Theta\left(\log^2\lambda_P\cdot\log(c/\varepsilon)\right)$, with probability at least $1-\varepsilon/5$, we have:

$$\forall p \in P:\ \|p-q\|_1 \ge D_0 + \varepsilon \implies \|h(p) - f(q)\|_1 \ge 4.$$

By Theorem 6, for $\gamma = \varepsilon/10$ and $\delta = \varepsilon/(5\lambda_P^{8\log(cD_0/\varepsilon)})$, with probability at least $1-\varepsilon/5$, we get:

$$\forall p \in P:\ \|p-q\|_1 \in (1+9\varepsilon, D_0+\varepsilon) \implies \|h(p) - f(q)\|_1 > (1+8\varepsilon)(1-\varepsilon) \ge 1 + 3\varepsilon.$$

Moreover,

$$\mathbf{Pr}\left[\,\|h(p^*) - f(q)\|_1 \le 1 + 3\varepsilon\right] \ge 1 - \frac{1+\varepsilon/10}{1+\varepsilon} \ge 1 - (1-\varepsilon/2).$$

Then, the target dimension needs to satisfy the following inequality:

$$k \geq \frac{\left(\ln\left(5\lambda_P^{8\log(cD_0/\varepsilon)}/\varepsilon\right)\right)^{2/\varepsilon}}{\zeta(\varepsilon)} = \frac{\left(\Theta(\log\log k \cdot \log\lambda_P + \log\lambda_P \cdot \ln(c/\varepsilon))\right)^{2/\varepsilon}}{\zeta(\varepsilon)}.$$

Hence, for $k = (\log\lambda_P \cdot \log(c/\varepsilon))^{\Theta(1/\varepsilon)}/\zeta(\varepsilon)$, we achieve a total probability of success in $\Omega(\varepsilon)$, which completes the proof. ◀

5 Dimension reduction based on randomly shifted grids

In this section, we explore some properties of randomly shifted grids, and we present a simplified embedding which consists of a first step of snapping points to a grid, and a second step of randomly projecting grid points.

Let $w > 0$ and t be chosen uniformly at random from the interval $[0, w]$. The function

$$h_{w,t}(x) = \left\lfloor \frac{x-t}{w} \right\rfloor$$

induces a random partition of the real line into segments of length w. Hence, the function

$$g_w(x) = (h_{w,t_1}(x_1), ..., h_{w,t_d}(x_d)),$$

for $t_1, \ldots, t_d$ independent uniform random variables in the interval $[0, w]$, induces a randomly shifted grid in $\mathbb{R}^d$. For a set $X \subseteq \mathbb{R}^d$, we denote by $g_w(X)$, the image of X on the randomly shifted grid points defined by g_w. For some $x \in \mathbb{R}^d$ and $r > 0$, the number of grid cells of $g_w(\ell_1^d)$ that $B_1(x, r)$ intersects per axis is independent, and in expectation is $1+2r/w$. Then, the expected total number of grid cells that $B_1(x, r)$ intersects is at most $(1+2r/w)^d$.

Now let $P \subset \ell_1^d$ be a set of n points with doubling constant λ_P and $q \in \ell_1^d$ a query point. For $w = \varepsilon/d$, the ℓ_1-diameter of each cell is ε and therefore $g_w(P)$ is an ε-covering set of P.

▶ **Lemma 11.** *Let $\mathcal{R} > 1$ and $P' := B_1(q, \mathcal{R}) \cap P$. Then, for $w = \varepsilon/d$*

$$\mathbb{E}\left[|g_w(P')|\right] \leq 8\lambda_P^{2\log(d\mathcal{R}/\varepsilon)}.$$

Proof. By the doubling constant definition, there exists a set of balls of radius ε/d^2 centered at points in P', of cardinality at most $\lambda_P^{2\log(d\mathcal{R}/\varepsilon)}$ which covers P'. For each ball of radius ε/d^2, the expected number of intersecting grid cells is $(1+2/d)^d \leq \mathrm{e}^2$. The lemma follows by linearity of expectation. ◀

The next lemma shows that, with constant probability, the growth on the number of representatives, as we move away from q, is bounded.

▶ **Lemma 12.** *Let $\{D_i\}_{i\in\mathbb{N}}$ be a sequence of radii such that, for any i, $D_{i+1} = 4D_i$. Let A_i be the points of $g_w(P)$ within distance $D_{i+1} = 2^{2(i+1)}D_0$ from q. Then, with probability at least $1/3$,*

$$\forall i \in \{-1, 0, \ldots\} : |A_i| \leq 4^{i+3}\lambda_P^{2\log(dD_{i+1}/\varepsilon)}.$$

Proof. By Lemma 11, $\mathbb{E}[|A_i|] \leq 8\lambda_P^{2\log(dD_{i+1}/\varepsilon)}$ for every $i \in \{-1, 0, \ldots\}$. Then, a union bound followed by Markov's inequality yields

$$\mathbf{Pr}\left[\exists i \in \{0, 1, \ldots\} : |A_i| \geq 4^{i+1}\,\mathbb{E}[|A_i|]\right] \leq 1/3.$$

In addition,

$$\mathbf{Pr}\left[|A_{-1}| \geq 4\,\mathbb{E}[|A_i|]\right] \leq 1/4.$$ ◀

▶ **Theorem 13.** *Let $P \subset \ell_1^d$ such that $|P| = n$. For any $\varepsilon \in (0, 1/2)$, there is a non-linear randomized embedding $h' : \ell_1^d \to \ell_1^k$, where $k = (\log \lambda_P \cdot \log(d/\varepsilon))^{\Theta(1/\varepsilon)} / \zeta(\varepsilon)$, for a function $\zeta(\varepsilon) > 0$ depending only on ε, such that for any $q \in \ell_1^d$, if there exists $p^* \in P$ such that $\|p^* - q\|_1 \leq 1$, then with probability $\Omega(\varepsilon)$,*

$$\|h'(p^*) - f(q)\|_1 \leq 1 + 3\varepsilon,$$
$$\forall p \in P: \ \|p - q\|_1 > 1 + 9\varepsilon \implies \|h'(p) - f(q)\|_1 > 1 + 3\varepsilon.$$

Any point can be embedded in time $O(dk)$.

Proof. We follow the same reasoning as in the proof of Theorem 10. The embedding is $h' = f \circ g_{\varepsilon/d}$, where f is the randomized linear map defined in Section 4. As before, we apply h' to every point in P, and only f to queries. The randomly shifted grid incurs an additive error of ε in the distances between q and P.

Assume wlog that $q = 0^d$ and let A_i be the points of $g_{\varepsilon/d}(P)$ within distance $D_{i+1} = 2^{2(i+1)} D_0$ from q. Hence, by Lemma 12,

$$\begin{aligned}\mathbf{Pr}\left[\exists i \exists s \in A_i : \|f(s)\|_1 \leq \frac{4\|s\|_1}{D_i}\right] &\leq \sum_{i=0}^{\infty} |A_i| \, \mathbf{Pr}\left[S \leq \frac{4T}{D_i}\right] \\ &\leq \sum_{i=0}^{\infty} 4^{i+3} \lambda_P^{2\log(dD_{i+1}/\varepsilon)} \, \mathbf{Pr}\left[\tilde{S} \leq \frac{\sqrt{4kT}}{\sqrt{D_i}}\right].\end{aligned}$$

As in Lemma 9, for $D_0 = \lceil 800T/k \rceil = \Theta(\log(k/\varepsilon))$, $k \geq 20 \log \lambda_P \cdot \log\left(\frac{dD_0}{\varepsilon\delta}\right)$ and $\delta = \varepsilon/5$,

$$\sum_{i=0}^{\infty} 4^{i+3} \lambda_P^{2\log(dD_{i+1}/\varepsilon)} \, \mathbf{Pr}\left[\tilde{S} \leq \frac{\sqrt{4kT}}{\sqrt{D_i}}\right] \leq \sum_{i=0}^{\infty} \frac{2^{2i+6+2\log\lambda_P[\log(dD_0/\varepsilon)+2(i+1)]}}{2^{k(i+1)}} \leq \varepsilon/5.$$

Hence, for $k = \Omega\left((\log^2 \lambda_P \cdot \log(d/\varepsilon)\right)$, with probability at least $1 - \varepsilon/5$, we have:

$$\forall p \in P: \ \|p - q\|_1 \geq D_0 + \varepsilon \implies \|h'(p) - f(q)\|_1 \geq 4.$$

Now, we are able to use Theorem 6 for points which are at distance at most $D_0 + \varepsilon$ from q, and the near neighbor. By Lemma 12, with constant probability, the number of grid points at distance $\leq D_0 + \varepsilon$, is at most $32 \cdot \lambda_P^{4\log(dD_0/\varepsilon)}$. Hence, by Theorem 6, for $\gamma = \varepsilon/10$ and $\delta = \varepsilon/(160\lambda_P^{4\log(dD_0/\varepsilon)})$, with probability at least $1 - \varepsilon/5$, it holds:

$$\forall p \in P: \ \|p - q\|_1 \in (1 + 9\varepsilon, D_0 + \varepsilon) \implies \|h'(p) - f(q)\|_1 > 1 + 3\varepsilon.$$

Moreover, with probability at least $\varepsilon/2$, we obtain:

$$\|h'(p^*) - f(q)\|_1 \leq 1 + 3\varepsilon.$$

As in Theorem 10, the target dimension needs to satisfy the following:

$$k \geq \frac{\left(\ln\left(160\lambda_P^{4\log(dD_0/\varepsilon)}/\varepsilon\right)\right)^{2/\varepsilon}}{\zeta(\varepsilon)}.$$

Hence, for $k = (\log \lambda_P \cdot \log(d/\varepsilon))^{\Theta(1/\varepsilon)} / \zeta(\varepsilon)$ we achieve total probability of success $\Omega(\varepsilon)$. ◀

6 Conclusion

We have filled in a gap in the spectrum of randomized embeddings with bounded distortion only for distances between the query and a pointset: such embeddings existed for ℓ_2 and ℓ_1 and for doubling subsets of ℓ_2. Here we settle the case of doubling subsets of ℓ_1 with a *near* neighbor-preserving embedding. In the meantime, we obtain concentration bounds on sums of independent Cauchy variables. Our algorithms are quite simple, therefore they should also be of practical interest.

We rely on approximate r-nets or randomly shifted grids. For the former, Theorem 10 provides with a trade-off between the preprocessing time required and the target dimension. On the other hand, Theorem 13 has the advantage of fast preprocessing: any point is embedded in $O(dk)$ time, and the embedding is oblivious to the pointset. In regards to the near-linear preprocessing time, the two results are comparable, since the dimension in Theorem 13 can be substituted by the target dimension of Theorem 6.

Notice that any potential improvements to Theorem 6 should lead to improvements to Theorems 10 and 13. The target dimension in these theorems follows from a direct application of Theorem 6 to the representative data points which lie inside a bounding ball centered at the query.

References

1 D. Achlioptas. Database-friendly Random Projections: Johnson-Lindenstrauss with Binary Coins. *J. Comput. Syst. Sci.*, 66(4):671–687, 2003.

2 N. Ailon and B. Chazelle. The Fast Johnson-Lindenstrauss Transform and Approximate Nearest Neighbors. *SIAM J. Comput.*, 39(1):302–322, May 2009.

3 E. Anagnostopoulos, I. Z. Emiris, and I. Psarros. Randomized Embeddings with Slack and High-Dimensional Approximate Nearest Neighbor. *ACM Trans. Algorithms*, 14(2):18:1–18:21, 2018. doi:10.1145/3178540.

4 A. Andoni and P. Indyk. Near-optimal Hashing Algorithms for Approximate Nearest Neighbor in High Dimensions. *Commun. ACM*, 51(1):117–122, 2008.

5 A. Andoni, T. Laarhoven, I. P. Razenshteyn, and E. Waingarten. Optimal Hashing-based Time-Space Trade-offs for Approximate Near Neighbors. In *Proc. ACM-SIAM Symposium on Discrete Algorithms, SODA, Barcelona, Spain*, pages 47–66, 2017.

6 A. Andoni, H. L. Nguyen, A. Nikolov, I. P. Razenshteyn, and E. Waingarten. Approximate near neighbors for general symmetric norms. In *Proc. ACM Symposium on Theory of Computing, STOC, Montreal, Canada*, pages 902–913, 2017.

7 Y. Bartal and L. A. Gottlieb. Dimension Reduction Techniques for $\ell_p, (1 < p < 2)$, with Applications. In *32nd International Symposium on Computational Geometry, SoCG 2016, June 14-18, 2016, Boston, MA, USA*, pages 16:1–16:15, 2016. doi:10.4230/LIPIcs.SoCG.2016.16.

8 Y. Bartal and L. A. Gottlieb. Approximate Nearest Neighbor Search for ℓ_p-Spaces $(2 < p < \infty)$ via Embeddings. In *Proc. LATIN: Theoretical Informatics - 13th Latin American Symp., Buenos Aires, Argentina*, pages 120–133, 2018. doi:10.1007/978-3-319-77404-6_10.

9 R. Cole and L. A. Gottlieb. Searching Dynamic Point Sets in Spaces with Bounded Doubling Dimension. In *Proc. ACM Symp. Theory of Computing*, pages 574–583, New York, USA, 2006. ACM.

10 D. Eppstein, S. Har-Peled, and A. Sidiropoulos. Approximate Greedy Clustering and Distance Selection for Graph Metrics. *CoRR*, abs/1507.01555, 2015. arXiv:1507.01555.

11 S. Har-Peled, P. Indyk, and R. Motwani. Approximate Nearest Neighbor: Towards Removing the Curse of Dimensionality. *Theory of Computing*, 8(1):321–350, 2012. doi:10.4086/toc.2012.v008a014.

12 S. Har-Peled and M. Mendel. Fast Construction of Nets in Low Dimensional Metrics, and Their Applications. In *Proc. Symp. Computational Geometry*, pages 150–158, 2005.

13 P. Indyk. On Approximate Nearest Neighbors under l_infinity Norm. *J. Comput. Syst. Sci.*, 63(4):627–638, 2001.

14 P. Indyk. Stable distributions, pseudorandom generators, embeddings, and data stream computation. *J. ACM*, 53(3):307–323, 2006. doi:10.1145/1147954.1147955.

15 P. Indyk and R. Motwani. Approximate Nearest Neighbors: Towards Removing the Curse of Dimensionality. In *Proc. ACM Symp. Theory of Computing*, pages 604–613, 1998.

16 P. Indyk and A. Naor. Nearest-neighbor-preserving Embeddings. *ACM Trans. Algorithms*, 3(3), 2007.

17 R. Krauthgamer and J. R. Lee. Navigating Nets: Simple Algorithms for Proximity Search. In *Proc. 15th Annual ACM-SIAM Symp. Discrete Algorithms*, SODA'04, pages 798–807, 2004.

18 J.R. Lee, M. Mendel, and A. Naor. Metric structures in L_1: dimension, snowflakes, and average distortion. *Eur. J. Comb.*, 26(8):1180–1190, 2005. doi:10.1016/j.ejc.2004.07.002.