Gaussian random projections for Euclidean membership problems

Vu Khac Ky, Pierre-Louis Poirion, Leo Liberti

CNRS LIX, École Polytechnique, F-91128 Palaiseau, France
Email: vu, poirion, liberti@lix.polytechnique.fr

November 19, 2015

Abstract

We discuss the application of random projections to the fundamental problem of deciding whether a given point in a Euclidean space belongs to a given set. We show that, under a number of different assumptions, the feasibility and infeasibility of this problem are preserved with high probability when the problem data is projected to a lower dimensional space. Our results are applicable to any algorithmic setting which needs to solve Euclidean membership problems in a high-dimensional space.

1 Introduction

Random projections are very useful dimension reduction techniques which are widely used in computer science [7, 13]. We assume we have an algorithm $A$ acting on a data set $X$ consisting of $n$ vectors in $\mathbb{R}^m$, where $m$ is large, and assume that the complexity of $A$ depends on $m$ and $n$ in a way that makes it impossible to run $A$ sufficiently fast. A random projection exploits the statistical properties of some random distribution to construct a mapping which embeds $X$ into a lower dimensional space $\mathbb{R}^k$ (for some appropriately chosen $k$) while preserving distances, angles, or other quantities used by $A$.

One striking example of random projections is the famous Johnson-Lindenstrauss lemma [9]:

1.1 Theorem (Johnson-Lindenstrauss Lemma)

Let $X$ be a set of $m$ points in $\mathbb{R}^m$ and $\varepsilon > 0$. Then there is a map $F: \mathbb{R}^m \to \mathbb{R}^k$ where $k$ is $O(\frac{\log m}{\varepsilon^2})$, such that for any $x, y \in X$, we have

$$(1 - \varepsilon)\|x - y\|_2^2 \leq \|F(x) - F(y)\|_2^2 \leq (1 + \varepsilon)\|x - y\|_2^2.$$  

(1)

Intuitively, this lemma claims that $X$ can be projected in a much lower dimensional space whilst keeping Euclidean distances approximately the same. The main idea to prove Thm. 1.1 is to construct a random linear mapping $T$ (called $JL$ random mapping onwards), sampled from certain distribution families, so that for each $x \in \mathbb{R}^m$, the event that

$$(1 - \varepsilon)\|x\|_2^2 \leq \|T(x)\|_2^2 \leq (1 + \varepsilon)\|x\|_2^2$$  

(2)

Supported by a Microsoft Research Ph.D. fellowship.
occurs with high probability. By Eq. (2) and the union bound, it is possible to show the existence of a map $F$ with the stated properties (see [2, 4]).

In this paper we employ random projections to study the following general problem:

**Euclidean Set Membership Problem (ESMP).** Given $p \in \mathbb{R}^m$ and $X \subseteq \mathbb{R}^m$, decide whether $p \in X$.

This is a fundamental class consisting of many problems, both in $\mathbf{P}$ (e.g. the Linear Feasibility Problem (LFP)) and $\mathbf{NP}$-hard (e.g. the Integer Feasibility Problem (IFP), which can naturally model sat, and also see [15]).

In this paper, we use a random linear projection operator $T$ to embed both $p$ and $X$ to a lower dimensional space, and study the relationship between the original membership problem and its projected version:

**Projected ESMP (PESMP).** Given $p, X, T$ as above, decide whether $T(p) \in T(X)$.

Note that, when $p \in X$, the fact that $T(p) \in T(X)$ follows by linearity of $T$. We are therefore only interested in the case when $p \notin X$, i.e. we want to estimate $\text{Prob}(T(p) \notin T(X))$, given that $p \notin X$.

1.1 Previous results

Random projections applying to some special cases of membership problems have been studied in [11], where we exploited some polyhedral structures of the problem to derive several results for polytopes and polyhedral cones. In the case $X$ is a polytope, we obtained the following result.

1.2 Proposition ([11])

Given $a_1, \ldots, a_n \in \mathbb{R}^m$, let $C = \text{conv}\{a_1, \ldots, a_n\}$, $b \in \mathbb{R}^m$ such that $b \notin C$, $d = \min_{x \in C} \|b - x\|$ and $D = \max_{1 \leq i \leq n} \|b - a_i\|$. Let $T : \mathbb{R}^m \rightarrow \mathbb{R}^k$ be a JL random mapping. Then

$$\text{Prob}(T(b) \notin T(C)) \geq 1 - 2n^2e^{-C(\varepsilon^2 - \varepsilon^3)k}$$

for some constant $C$ (independent of $m, n, k, d, D$) and $\varepsilon < \frac{d^2}{2n}$. If $X$ is a polyhedral cone, we obtained the following result.

1.3 Proposition ([11])

Given $b, a_1, \ldots, a_n \in \mathbb{R}^m$ of norms 1 such that $b \notin C = \text{cone}\{a_1, \ldots, a_n\}$, let $d = \min_{x \in C} \|b - x\|$ and $T : \mathbb{R}^m \rightarrow \mathbb{R}^k$ be a JL random mapping. Then:

$$\text{Prob}(T(b) \notin T(C)) \geq 1 - 2n(n + 1)e^{-C(\varepsilon^2 - \varepsilon^3)k}$$

for some constant $C$ (independent of $m, n, k, d$), where $\varepsilon = \frac{d^2}{\mu_A^2 + 2\sqrt{1 - d^2} \mu_A + 1}$,

$$\mu_A = \max\{\|x\|_A \mid x \in \text{cone}(a_1, \ldots, a_n) \land \|x\| \leq 1\},$$

and $\|x\|_A = \min\{\sum_i \theta_i \mid \theta \geq 0 \land x = \sum_i \theta_i a_i\}$ is the norm induced by $A = (a_1, \ldots, a_n)$.
We also recall the following Lemma, useful for the integer case.

1.4 Lemma ([11])
Let \( T : \mathbb{R}^m \to \mathbb{R}^k \) be a JL random mapping, let \( b, a_1, \ldots, a_n \in \mathbb{R}^m \) and let \( X \subseteq \mathbb{R}^m \) be a finite set. Then if \( b \neq \sum_{i=1}^{m} y_i a_i \) for all \( y \in X \), we have

\[
\text{Prob} \left( \forall y \in X \mid T(b) \neq \sum_{i=1}^{m} y_i T(a_i) \right) \geq 1 - 2|X|e^{-Ck};
\]

for some constant \( C > 0 \) (independent of \( m, k \)).

1.2 New results

In this paper, we consider the general case where the data set \( X \) has no specific structure, and use Gaussian random projections in our arguments to obtain some results about the relationship between ESMP and PESMP.

In the case when \( X \) is at most countable (i.e. finite or countable), using a straightforward argument, we prove that these two problems are equivalent almost surely. However, this result is only of theoretical interest due to round-off errors in floating point operations, which make its practical application difficult. We address this issue by introducing a threshold \( \delta > 0 \) with a corresponding Threshold ESMP (TESMP): if \( \Delta \) is the distance between \( T(p) \) and the closest point of \( T(X) \), decide whether \( \Delta \geq \delta \).

In the case when \( X \) may also be uncountable, we employ the doubling constant of \( X \), i.e. the smallest number \( \lambda_X \) such that any closed ball in \( X \) can be covered by at most \( \lambda_X \) closed balls of half the radius. Its logarithm \( \log_2 \lambda_X \) is called doubling dimension of \( X \). Recently, the doubling dimension has become a powerful tool for several classes of problems such as nearest neighbor [10, 8], low-distortion embeddings [3], clustering [12].

We show that we can project \( X \) into \( \mathbb{R}^k \), where \( k = O(\log_2 \lambda_X) \), whilst still ensure the equivalence between ESMP and PESMP with high probability. We also extend this result to the threshold case, and obtain a more useful bound for \( k \).

2 Finite and countable sets

In this section, we assume that \( X \) is either finite or countable. Let \( T \) be a JL random mapping from a Gaussian distribution, i.e. each entry of \( T \) is independently sampled from \( \mathcal{N}(0,1) \). It is well known that, for an arbitrary unit vector \( a \in S^{m-1} \), the random variable \( \| Ta \|^2 \) has a Chi-squared distribution \( \chi_k^2 \) with \( k \) degrees of freedom ([14]). Its corresponding density function is \( \frac{2^{-k/2}}{\Gamma(k/2)} x^{k/2-1} e^{k/2} \), where \( \Gamma(\cdot) \) is the gamma function. By [4], for any \( 0 < \delta < 1 \), taking \( z = \frac{\delta}{k} \) yields a cumulative distribution function

\[
F_{\chi_k^2}(\delta) \leq (ze^{1-z})^{k/2} < (ze)^{k/2} = \left( \frac{e\delta}{k} \right)^{k/2}.
\]

Thus, we have

\[
\text{Prob}(\| Ta \| \leq \delta) = F_{\chi_k^2}(\delta^2) < (3\delta^2)^{k/2}
\]
or, more simply, \( \text{Prob}(\|Ta\| < \delta) < \delta^k \) when \( k \geq 3 \).

Using this estimation, we immediately obtain the following result.

**2.1 Proposition**

Given \( p \in \mathbb{R}^m \) and \( X \subseteq \mathbb{R}^m \), at most countable, such that \( p \notin X \). Then, for a Gaussian random projection \( T : \mathbb{R}^m \to \mathbb{R}^k \) with any \( k \geq 1 \), we have \( T(p) \notin T(X) \) almost surely, i.e. \( \text{Prob}(T(p) \notin T(X)) = 1 \).

**Proof.** First, note that for any \( u \neq 0, Tu \neq 0 \) holds almost certainly. Indeed, without loss of generality we can assume that \( \|u\| = 1 \). Then for any \( 0 < \delta < 1 \):

\[
\text{Prob}(T(z) = 0) \leq \text{Prob}(\|Tz\| \leq \delta) = (3\delta^2)^{k/2} \to 0 \text{ as } \delta \to 0.
\]

Since the event \( T(p) \notin T(X) \) can be written as the intersection of at most countably many almost sure events \( T(p) \neq T(x) \) (for \( x \in X \)), it follows that \( \text{Prob}(T(p) \notin T(X)) = 1 \), as claimed. \( \square \)

Proposition 2.1 is simple, but it looks interesting because it suggests that we only need to project the data points to a line (i.e. \( k = 1 \)) and study an equivalent membership problem on a line. Furthermore, it turns out that this result remains true for a large class of random projections.

**2.2 Proposition**

Let \( \nu \) be a probability distribution on \( \mathbb{R}^m \) with bounded Lebesgue density \( f \). Let \( Y \subseteq \mathbb{R}^m \) be an at most countable set such that \( 0 \notin Y \). Then, for a random projection \( T : \mathbb{R}^m \to \mathbb{R}^1 \) sampled from \( \nu \), we have \( 0 \notin T(Y) \) almost surely, i.e. \( \text{Prob}(0 \notin T(Y)) = 1 \).

**Proof.** For any \( 0 \neq y \in Y \), consider the set \( \mathcal{E}_y = \{T : \mathbb{R}^m \to \mathbb{R}^1 \mid T(y) = 0\} \). If we regard each \( T : \mathbb{R}^m \to \mathbb{R}^1 \) as a vector \( t \in \mathbb{R}^m \), then \( \mathcal{E}_y \) is a hyperplane \( \{t \in \mathbb{R}^m \mid y \cdot t = 0\} \) and we have

\[
\text{Prob}(T(y) = 0) = \nu(\mathcal{E}_y) = \int_{\mathcal{E}_y} f d\mu \leq \|f\|_\infty \int_{\mathcal{E}_y} d\mu = 0
\]

where \( \mu \) denotes the Lebesgue measure on \( \mathbb{R}^m \). The proof then follows by the countability of \( Y \), similarly to Proposition 2.1 \( \square \)

Proposition 2.2 is based on the observation that the degree \( [\mathbb{R} : \mathbb{Q}] \) of the field extension \( \mathbb{R}/\mathbb{Q} \) is \( 2^{80} \), whereas \( Y \) is countable; so the probability that any row vector \( T_i \) of the random projection matrix \( T \) will yield a linear dependence relation \( \sum_{j \leq m} T_{ij} y_j = 0 \) for some \( 0 \neq y \in Y \) is zero. In practice, however, \( Y \) is part of the rational input of a decision problem, and the components of \( T \) are rational: hence any subsequence of them is trivially linearly dependent over \( \mathbb{Q} \). Moreover, floating point numbers have a bounded binary representation: hence, even if \( Y \) is finite, there is a nonzero probability that any subsequence of components of \( T \) will be linearly dependent by means of a nonzero multiplier vector in \( Y \).

This idea, however, does not work in practice: we tested it by considering the ESMP given by the IPF defined on the set \( \{x \in \mathbb{Z}_+^n \cap [L,U] \mid Ax = b\} \). Numerical experiments indicate that the corresponding PESMP \( \{x \in \mathbb{Z}_+^n \cap [L,U] \mid T(A)x = T(b)\} \), with \( T \) consisting of a one-row Gaussian projection matrix, is always feasible despite the infeasibility of the original IPF. Since Prop. 2.1
assumes that the components of $T$ are real numbers, we think that the reason behind this failure is the round-off error associated to the floating point representation used in computers. Specifically, when $T(A)x$ is too close to $T(b)$, floating point operations will consider them as a single point. In order to address this issue, we force the projected problems to obey stricter requirements. In particular, instead of only requiring that $T(p) \notin T(X)$, we ensure that

$$\text{dist}(T(p), T(X)) = \min_{x \in X} \|T(p) - T(x)\| > \tau,$$

where $\text{dist}$ denotes the Euclidean distance, and $\tau > 0$ is a (small) given constant. With this restriction, we obtain the following result.

### 2.3 Proposition

Given $\tau, \delta > 0$ and $p \notin X \subseteq \mathbb{R}^m$, where $X$ is a finite set, let

$$d = \min_{x \in X} \|p - x\| > 0.$$

Let $T : \mathbb{R}^m \rightarrow \mathbb{R}^k$ be a Gaussian random projection with $k \geq \frac{\log(|X|)}{\log\left(\frac{d}{\delta}\right)}$. Then:

$$\text{Prob}\left(\min_{x \in X} \|T(p) - T(x)\| > \tau\right) > 1 - \delta.$$

**Proof.** We assume that $k \geq 3$. For any $x \in X$ we have:

$$\text{Prob}(\|T(p - x)\| \leq \tau) = \text{Prob}\left(T\left(\frac{p - x}{\|p - x\|}\right) \leq \frac{\tau}{\|p - x\|}\right) \leq \text{Prob}\left(T\left(\frac{p - x}{\|p - x\|}\right) \leq \frac{\tau}{d}\right) < \frac{\tau^k}{d^k},$$

due to (3). Therefore, by the union bound,

$$\text{Prob}(\min_{x \in X} \|T(p) - T(x)\| > \tau) = 1 - \text{Prob}(\min_{x \in X} \|T(p) - T(x)\| \leq \tau) \geq 1 - \sum_{x \in X} \text{Prob}(\|T(p) - T(x)\| \leq \tau) > 1 - |X| \left(\frac{\tau}{d}\right)^k.$$

The RHS is greater than or equal to $1 - \delta$ if and only if \( \left(\frac{d}{\tau}\right)^k \geq \frac{|X|}{\delta} \), which is equivalent to \( k \geq \frac{\log(|X|)}{\log\left(\frac{d}{\delta}\right)} \), as claimed.

Note that $d$ is often unknown and can be arbitrarily small. However, if both $p, X$ are integral, then $d \geq 1$ and we can select $k > \frac{\log(|X|)}{\log\left(\frac{d}{\delta}\right)}$ in the above proposition.

In many cases, the set $X$ is infinite. We show that when this is the case, we can still overcome this difficulty under some assumptions. In particular, we prove that if $X = \{Ax \mid x \in \mathbb{Z}^n\}$ where $A$ is an $m \times n$ matrix with integer coefficients which are all positive in at least one row, then for any bounded vector $b \in \mathbb{Z}^m$ the problem $b \in X$ is equivalent, with high probability, to its projection to a $O(\log n)$-dimensional space. The idea is to separate one positive row and apply random projection to the others.
Formally, let us denote by $a^i$ the $i$-th row and by $a^j$ the $j$-th column of $A$. Assume that all entries in the row $a^i$ is positive and all entries of $b$ are bounded by a constant $B > 0$. Remove the row $i$ from $A$ and $b$ to obtain $\tilde{A} = (a_1', \ldots, a_n') \in \mathbb{Z}^{(m-1) \times n}$ and $\tilde{b} \in \mathbb{Z}^{m-1}$. Let $T : \mathbb{R}^{m-1} \to \mathbb{R}^k$ be a JL random mapping and denote by $Z = \{ x \in \mathbb{Z}_n^+ | a^i \cdot x = b_i \}$. Then we have:

**2.4 Proposition**
Assume that $b \notin X$, and let $0 < \delta < 1$. Using the terminology and given the assumptions above, if $k \geq \frac{1}{C} \ln\left(\frac{2}{\delta}\right) + \frac{B}{C} \log(n + B - 1)$ we have

$$\text{Prob}\left(T(b) \neq \sum_{j=1}^n x_j T(a^j) \text{ for all } x \in Z\right) \geq 1 - \delta$$

for some constant $C > 0$.

**Proof.** We first show that $|Z| \leq (n + B - 1)^B$. Since all the entries of $A$ are positive integers, we have

$$|Z| \leq \left| \{ x \in \mathbb{Z}_n^+ | \sum_{j=1}^n x_j = b_i \} \right| \leq \left| \{ x \in \mathbb{Z}_n^+ | \sum_{j=1}^n x_j = B \} \right|.$$

The number of elements in the RHS corresponds to the number of combinations with repetitions of $B$ items sampled from $n$, which is equal to $\binom{n+B-1}{n-1} = \binom{n+B-1}{B} \leq (n + B - 1)^B$.

Next, by Lemma 1.4, we have:

$$\text{Prob}\left(T(b) \neq \sum_{j=1}^n x_j T(a^j) \text{ for all } x \in Z\right) \geq 1 - 2(n + B - 1)^B e^{-Ck}, \quad (5)$$

which is greater than $1 - \delta$ when taking any $k$ such that $k \geq \frac{1}{C} \ln\left(\frac{2}{\delta}\right) + \frac{B}{C} \log(n + B - 1)$. The proposition is proved. \(\square\)

Note that in Prop. 2.4 we can choose the JL random mapping $T$ as a matrix with \{-1, +1\} entries (Rademacher variables). In this case, there is no need to worry about floating point errors.

## 3 Sets with low doubling dimension

In this section, we denote by $B(x,r)$ the closed ball centered at $x$ with radius $r > 0$, and $B_X(x,r) = B(x,r) \cap X$. We will also assume that $X$ is a doubling space, i.e. a set with bounded doubling dimension. One example of doubling spaces is a Euclidean space $\mathbb{R}^m$, we can show that the doubling dimension $\log_2(\lambda_X)$ of $X$ can be shown to be a constant factor of $m$ (16 6). However, many sets of low doubling dimensions are contained in high dimensional spaces (17). Note that computing the doubling dimension of a metric space is generally NP-hard (15). We shall make use of the following simple lemma.

**3.1 Lemma**
For any $p \in X$ and $\varepsilon, r > 0$, there is a set $S \subseteq X$ of size at most $\lambda_X^{\log_2(\frac{r}{\varepsilon})}$ such that

$$B_X(p,r) \subseteq \bigcup_{s \in S} B(s, \varepsilon).$$
Proof. By definition of the doubling dimension, \( B_X(p, r) \) is covered by at most \( \lambda_X \) closed balls of radius \( \frac{r}{2} \). Each of these balls in turn is covered by \( \lambda_X \) balls of radius \( \frac{r}{4} \), and so on: iteratively, for each \( k \geq 1 \), \( B_X(p, r) \) is covered by \( \lambda_X^k \) balls of radius \( \frac{r}{2^k} \). If we select \( k = \lceil \log_2(\frac{r}{\epsilon}) \rceil \) then \( k \geq \log_2(\frac{r}{\epsilon}) \), i.e. \( \frac{r}{2^k} \leq \epsilon \). This means \( B_X(p, r) \) is covered by \( \lambda_X^{\lceil \log_2(\frac{r}{\epsilon}) \rceil} \) balls of radius \( \epsilon \).

We will also use the following lemma, which is proved in [8] using a concentration estimation for sum of squared gaussian variables (Chi-squared distribution).

3.2 Lemma
Let \( X \subseteq B(0, 1) \) be a subset of the \( m \)-dimensional Euclidean unit ball. Then there exist universal constants \( c, C > 0 \) such that for \( k \geq C \log \lambda_X + 1 \) and \( \delta > 1 \), the following holds:

\[
\Pr(\exists x \in X \text{ s.t. } \|Tx\| > \delta) < e^{-c_1 k \delta^2}.
\]

In the proof of the next result (one of the main results in this section), we use the same idea as that in [8] for the nearest neighbor problem.

3.3 Theorem
Given \( 0 < \delta < 1 \) and \( p \notin X \subseteq \mathbb{R}^m \). Let \( T : \mathbb{R}^m \to \mathbb{R}^k \) be a Gaussian random projection. Then

\[
\Pr(T(p) \notin T(X)) = 1
\]

if \( k \geq C \log_2(\lambda_X) \), for some universal constant \( C \).

Proof. Let \( \epsilon > 0 \) and \( 0 = r_0 < r_1 < r_2 < \ldots \) be positive scalars (their values will be defined later). For each \( j = 1, 2, 3, \ldots \) we define a set

\[
X_j = X \cap B(p, r_j) \setminus B(p, r_{j-1}).
\]

Since \( X_j \subseteq B_X(p, r_j) \), by Lemma [3.1] we can find a point set \( S_j \subseteq X \) of size \( |S_j| \leq \lambda_X^{\lceil \log_2(\frac{r_j}{\epsilon}) \rceil} \) such that

\[
X_j \subseteq \bigcup_{s \in S_j} B(s, \epsilon).
\]

Hence, for any \( x \in X_j \), there is \( s \in S_j \) such that \( \|x - s\| < \epsilon \). Moreover, by the triangle inequality, any such \( s \) satisfies \( r_{j-1} - \epsilon < \|s - p\| < r_j + \epsilon \), so without loss of generality we can assume that

\[
S_j \subseteq B(p, r_j + \epsilon) \setminus B(p, r_{j-1} - \epsilon).
\]

We denote by \( E_j \) the event that:

\[
\exists s \in S_j, \exists x \in X_j \cap B(s, \epsilon) \text{ s.t. } \|Ts - Tx\| > \epsilon \sqrt{j}.
\]

By the union bound, we have

\[
\Pr(E_j) \leq \sum_{s \in S_j} \Pr(\exists x \in X_j \cap B(s, \epsilon) \text{ s.t. } \|Ts - Tx\| > \epsilon \sqrt{j})
\]

\[
\leq \sum_{s \in S_j} e^{-c_1 kj} \quad \text{(for some universal constant } c_1 \text{ by Lemma 3.2)}
\]

\[
\leq \lambda_X^{\lceil \log_2(\frac{r_j+\epsilon}{\epsilon}) \rceil} e^{-c_1 kj}.
\]
Again by the union bound, we have:

\[
\Pr \left( \exists x \in X \text{ s.t. } T(x) = T(p) \right) = \Pr \left( \exists x \in \bigcup_{j=1}^{\infty} X_j \text{ s.t. } T(x) = T(p) \right)
\leq \sum_{j=1}^{\infty} \Pr \left( \exists x \in X_j \text{ s.t. } T(x) = T(p) \right).
\]

Now we will estimate the individual probabilities:

\[
\Pr \left( \exists x \in X_j \text{ s.t. } T(x) = T(p) \right) \\
\leq \Pr \left( \exists x \in X_j \text{ s.t. } T(x) = T(p) \right) \cap \mathcal{E}_j + \Pr(\mathcal{E}_j)
\leq \Pr \left( \exists x \in X_j, s \in S_j \cap B(x, \varepsilon) \text{ s.t. } T(x) = T(p) \cap \|T(s) - T(x)\| \leq \varepsilon \sqrt{j} \right) + \Pr(\mathcal{E}_j)
\leq \Pr(\exists s \in S_j \text{ s.t. } \|T(s) - T(p)\| < \varepsilon \sqrt{j}) + \lambda_X^{\log \frac{c_1 k j}{\varepsilon}} e^{-c_1 k j}.
\]

Next, we choose \( \varepsilon = \frac{\delta}{N} \) for some large \( N \); and for each \( j \geq 1 \), we choose \( r_j = (2 + j) \varepsilon \). For \( j < N - 2 \), by definition it follows that \( X_j = \emptyset \). Therefore

\[
\Pr \left( \exists x \in X_j \text{ s.t. } T(s) = T(p) \right) = 0.
\]

On the other hand, for \( j \geq N - 2 \),

\[
\Pr \left( \exists s \in S_j \text{ s.t. } \|T(s) - T(p)\| \leq \varepsilon \sqrt{j} \right)
\leq \lambda_X^{\log \frac{r_j + k \varepsilon}{\varepsilon}} \Pr \left( \|T(z)\| \leq \frac{\varepsilon \sqrt{j}}{r_j - 1} + \varepsilon \right) \text{ for an arbitrary } z \in S^{n-1}
\leq \lambda_X^{\log (3+j)} \Pr \left( \|T(z)\| \leq \frac{1}{\sqrt{j}} \right) \text{ for an arbitrary } z \in S^{n-1}
\leq \lambda_X^{\log (3+j)} j^{-k/2} \text{ by the estimation (4)}.
\]

Note that \( \lambda_X^{\log (3+j)} \leq \lambda_X^{\log (6+2j)} = (6 + 2j) \lambda_X < j^2 \lambda_X \) for large enough \( N \). Therefore, we have

\[
\Pr \left( \exists x \in X_j \text{ s.t. } T(x) = T(p) \right) \leq \lambda_X^{\log (3+j)} \left( j^{-k/2} + e^{-c_1 k j} \right)
\leq j^{-c_2 k} + e^{-c_3 k j}
\]

for some universal constants \( c_2, c_3 \), provided that \( k \geq c_1 \log \lambda_X \) for some large enough constant \( c_1 \). Finally, by the union bound,

\[
\Pr(T(p) \notin T(X)) = 1 - \Pr(T(p) \in T(X)) \\
\geq 1 - \sum_{i=N-2}^{\infty} \left( i^{-c_2 k} + e^{-c_3 k j} \right)
\]

which tends to 1 when \( N \) tends to infinity.

Our final result in the section is an extension of Thm. 3.3 to the threshold case.
3.4 Theorem
Let \( p \notin X \subseteq \mathbb{R}^m \), \( T : \mathbb{R}^m \to \mathbb{R}^k \) be a Gaussian random projection, and \( d = \min_{x \in X} \| p - x \| \). Then for all \( 0 < \delta < 1 \) and all \( 0 < \tau < \kappa d \) for some constant \( \kappa < 1 \), we have
\[
\text{Prob}(\text{dist}(T(p), T(X)) > \tau) > 1 - \delta
\]
if \( k \) is \( O\left(\frac{\log \left(\frac{\lambda_X}{d}\right)}{\log \left(\frac{\epsilon}{\lambda_X}\right)}\right) \).

Proof. For \( j = 1, 2, \ldots \) we construct the sets \( X_j, S_j \) similarly as those in the proof of Thm. 3.3 (where the values of \( r_j \) and \( \epsilon \) will be defined later). Then we have
\[
\text{Prob}(\exists x \in X \text{ s.t } \|T(x) - T(p)\| < \tau) = \text{Prob}(\exists x \in \bigcup_{j=1}^{\infty} X_j \text{ s.t } \|T(x) - T(p)\| < \tau) \\
\leq \sum_{j=1}^{\infty} \text{Prob}(\exists x \in X_j \text{ s.t } \|T(x) - T(p)\| < \tau).
\]
For all \( j \geq 1 \), we have
\[
\text{Prob}(\exists x \in X_j \text{ s.t } \|T(x) - T(p)\| < \tau) \\
\leq \text{Prob}( (\exists x \in X_j \text{ s.t } \|T(x) - T(p)\| < \tau) \land x \in \mathcal{E}_j ) + \text{Prob}(\mathcal{E}_j) \\
\leq \text{Prob}(\exists x \in X_j, s \in S_j \cap B(x, \epsilon) \text{ s.t } \|T(x) - T(p)\| < \tau \land \|T(s) - T(x)\| \leq \epsilon \sqrt{j}) + \text{Prob}(\mathcal{E}_j) \\
\leq \text{Prob}(\exists s \in S_j \text{ s.t } \|T(s) - T(p)\| < \tau + \epsilon \sqrt{j}) + \lambda_X^{\left[\log_2 \left(\frac{r_j + \epsilon}{\epsilon}\right)\right]} e^{-c_1 k j}.
\]
Now we choose \( \epsilon = \frac{\tau}{N} \) for some \( N > 0 \) such that \( 1 + \frac{1}{N} < \frac{1}{\kappa} \) and for each \( j \geq 1 \), we choose \( r_j = \tau \sqrt{j} + 1 + (2 + j) \epsilon \). For \( j = 1 \), by the union bound we have
\[
\text{Prob}(\exists s \in S_1 \text{ s.t } \|T(s) - T(p)\| \leq \tau + \epsilon \sqrt{1}) \\
\leq \lambda_X^{\left[\log_2 \left(\frac{r_j + \epsilon}{\epsilon}\right)\right]} \text{Prob}(\|T(z)\| \leq \frac{\tau + \epsilon}{d}) \quad \text{for an arbitrary } z \in S_{m-1} \\
= \lambda_X^{\left[\log_2 (4 + N \sqrt{j})\right]} \text{Prob}(\|T(z)\| \leq (1 + \frac{1}{N}) \frac{\tau}{d}) \quad \text{for an arbitrary } z \in S_{m-1} \\
< \lambda_X^{\left[\log_2 (4 + N \sqrt{j})\right]} \left(1 + \frac{1}{N}\right)^{\frac{k}{2}} \quad \text{by estimation (1)} \\
< \left(1 + \frac{1}{N}\right)^{\frac{k}{d}} (c_2)^k
\]
for some universal constant \( c_2 > 0 \), as long as \( k > C \log(\lambda_X) \) for some \( C \) large enough.

For \( j \geq 2 \), we have
\[
\text{Prob}(\exists s \in S_j \text{ s.t } \|T(s) - T(p)\| \leq \tau + \epsilon \sqrt{j}) \\
\leq \lambda_X^{\left[\log_2 \left(\frac{r_j + \epsilon}{\epsilon}\right)\right]} \text{Prob}(\|T(z)\| \leq \frac{\tau + \epsilon \sqrt{j}}{r_{j-1} - \epsilon}) \quad \text{for an arbitrary } z \in S_{m-1} \\
= \lambda_X^{\left[\log_2 (3j + N \sqrt{j} + 1)\right]} \text{Prob}(\|T(z)\| \leq \frac{1}{\sqrt{j}}) \quad \text{for an arbitrary } z \in S_{m-1} \\
< \lambda_X^{\left[\log_2 (3j + N \sqrt{j} + 1)\right]} j^{-k/2} \quad \text{by estimation (1)} \\
< j^{-c_3 k}
\]
for some universal constant $c_3 > 0$, as long as $k > C \log(\lambda X)$ for some $C$ large enough.

Similarly, for all $1 \leq j$, we have

$$
\lambda_X^{[\log_2 \left( \frac{r_j \epsilon}{2} \right)]} e^{-c_1 kj} \leq e^{-c_4 kj},
$$

for some universal constant $c_4 > 0$, as long as $k > C \log(\lambda X)$ for some $C$ large enough.

From estimations (6), (7), (8) and by the union bound we have:

$$
\text{Prob}(\text{dist}(T(p), T(X)) \geq \tau) \geq 1 - \sum_{j=1}^{\infty} \text{Prob}(\text{dist}(T(p), T(X_j)) < \tau)
$$

$$
\geq 1 - \left(1 + \frac{1}{N} \right)^{\tau d} - \sum_{j=2}^{\infty} e^{-c_1 kj} - \sum_{j=1}^{\infty} e^{-c_4 kj}
$$

$$
\geq 1 - \delta \quad \text{for } k = O\left(\frac{\log(\lambda X)}{\log(\tau)}\right) \text{ large enough}.
$$

\[\square\]

References

[1] I. Abraham, Y. Bartal, and O. Neiman. Embedding metric spaces in their intrinsic dimension. In 19th Annual ACM-SIAM Symposium on Discrete Algorithms, page 363-372, 2008.

[2] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. Journal of Computer and System Sciences, 66:671–687, 2003.

[3] P. Agarwal, S. Har-Peled, and H. Yu. Embeddings on surfaces, curves, and moving points in Euclidean space. In Proceedings of the 23rd Symposium on Computational Geometry, pages 381–389. ACM, 2007.

[4] S. Dasgupta and A. Gupta. An elementary proof of a theorem by johnson and lindenstrauss. Random Structures and Algorithms, 22:60–65, 2002.

[5] L. Gottlieb and R. Krauthgamer. Proximity algorithms for nearly doubling spaces. SIAM Journal on Discrete Mathematics, 27(4):1759–1769, 2013.

[6] A. Gupta, R. Krauthgamer, and J.R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In IEEE Symposium on Foundations of Computer Science, page 534543. IEEE, 2003.

[7] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In 42nd Annual Symposium on Foundations of Computer Science, pages 10–33. IEEE, 2001.

[8] P. Indyk and A. Naor. Nearest neighbor preserving embeddings. ACM Transactions on Algorithms, 3(3):Art. 31, 2007.
REFERENCES

[9] W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In G. Hedlund, editor, *Conference in Modern Analysis and Probability*, volume 26 of *Contemporary Mathematics*, pages 189–206, Providence, 1984. AMS.

[10] R. Krauthgamer and J.R. Lee. Navigating nets: Simple algorithms for proximity search. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 791–801, 2004.

[11] Vu Khac Ky, P.-L. Poirion, and L. Liberti. Using the Johnson-Lindenstrauss lemma in linear and integer programming. Technical Report 07/4985, Optimization Online, 2015.

[12] A. Magen. Dimensionality reductions in $\ell_2$ that preserve volumes and distance to affine spaces. *Discrete and Computational Geometry*, 30(1):139–153, 2007.

[13] J. Matousek. *Lectures on Discrete Geometry*. Springer, New York, 2002.

[14] A. Mood, F. Graybill, and D. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, 1974.

[15] K. Murty and S. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129, 1987.

[16] J.-L. Verger-Gaugry. Covering a ball with smaller equal balls in $\mathbb{R}^n$. *Discrete Computational Geometry*, 33:143–155, 2005.