Gaussian random projections for Euclidean membership problems

Ky Vu[a], Pierre-Louis Poirion[b], Leo Liberti[c]

[a] ITCSC, Chinese University of Hong Kong, P.R. China
[b] Huawei Research Center, Paris, France
[c] CNRS LIX, Ecole Polytechnique, 91128 Palaiseau, France

Abstract

We discuss the application of Gaussian random projections to the fundamental problem of deciding whether a given point in a Euclidean space belongs to a given set. In particular, we consider the two cases, when the target set is either at most countable or of low doubling dimension. We show that, under a number of different assumptions, the feasibility (or infeasibility) of this problem is preserved almost surely when the problem data is projected to a lower dimensional space. We also consider the threshold version of this problem, in which we require that the projected point and the projected set are separated by a certain distance error. As a consequence of these results, we are able to improve the bound of Indyk-Naor on the Nearest Neigbour preserving embeddings. Our results are applicable to any algorithmic setting which needs to solve Euclidean membership problems in a high-dimensional space.

Keywords: Johnson-Lindenstrauss lemma, Machine Learning, Euclidean Distance Geometry, clustering.

1. Introduction

Random projections are useful dimension reduction techniques which are widely used in many areas such as Machine Learning [4, 8], Computer Science [12, 18], Numerical Linear Algebra [17, 23] and so on. In standard algorithmic settings, assume we have an algorithm $\mathcal{A}$ acting on a data set X consisting of n vectors in $\mathbb{R}^m$, where m is large, and assume that the complexity of $\mathcal{A}$ depends on m and n in a way that makes it impossible to run $\mathcal{A}$ 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 $\mathcal{A}$. Their simplicity notwithstanding, the performance of random projections is comparable to more complicated methods such as SVD and PCA [5]. One striking example of random projections is the famous Johnson-Lindenstrauss lemma [14]:

Email addresses: vukhacky@gmail.com (Ky Vu), pierre.louis.poirion@huawei.com (Pierre-Louis Poirion), liberti@lix.polytechnique.fr (Leo Liberti)

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

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

Intuitively, this lemma claims that X can be projected into a much lower dimensional space whilst keeping Euclidean distances approximately the same. The main idea to prove Theorem 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

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

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 [1, 7]. The JLL is a consequence of a general property of *sub-gaussian* random mappings $T = \frac{1}{\sqrt{d}}U$ [19]. Some of the most popular choices for U are:

Choices of random projection

1. $U = \sqrt{m}P$ where P is the orthogonal projections on a random k-dimensional linear subspace of $\mathbb{R}^m$ [14];
2. random $d \times m$ matrices with each entry independently drawn from the standard normal distribution $\mathcal{N}(0, 1)$ [13];
3. random $d \times m$ matrices with each entry independently taking values $+1$ and -1, each with probability $\frac{1}{2}$ [1];
4. random $d \times m$ matrices with entries independently taking values $+1$, 0, -1, respectively with probability $\frac{1}{6}$, $\frac{2}{3}$, $\frac{1}{6}$ [1].

In this paper we employ random projections to study the following meta-problem (parametrized on S):

> Euclidean Set Membership Problem (ESMP). Given $b \in \mathbb{R}^m$ and $S \subseteq \mathbb{R}^m$, decide whether $b \in S$.

This problem is important and extensively studied in optimization as feasibility problems. For example, any linear and integer programming problem can be transformed easily (by bisection arguments) into feasibility problems:

> Integer and Linear Feasibility Problems. Decide whether the set $\{x \in \mathbb{X} \mid Ax = b\}$ is empty or not,

where $\mathbb{X}$ is either $\mathbb{R}^n_+$ or $\mathbb{Z}^n_+$. This is exactly the membership problem where S is the restricted linear span $\{Ax \mid x \in \mathbb{X}\}$.

In this paper, we consider the general case where the data set S has no specific structure. We will use a random projection $T = \frac{1}{\sqrt{d}}U$ (we assume in the whole paper than $d \geq 3$), where U is a Gaussian random projection, i.e., a random $d \times m$ matrix with each entry independently drawn from the standard normal distribution $\mathcal{N}(0, 1)$, in our arguments to embed both b and S to a lower dimensional space, and solve its associated projected version:

> Projected ESMP (PESMP). Given $b \in \mathbb{R}^m$ and $S \subseteq \mathbb{R}^m$, and let $T : \mathbb{R}^m \to \mathbb{R}^d$ be a random projection. Decide whether $T(b) \in T(S)$.

Our objective is to investigate the relationships between ESMP and PESMP. It is easy to argue that $b \in S$ implies $T(b) \in T(S)$. In this paper, we are only interested in the problematic case when $b \notin S$. Since T is a random projection, we want to estimate $\mathsf{Prob}(T(b) \notin T(S))$, given that $b \notin S$. Notice that the probability does not change if we multiply T by some constant. Hence we could simply consider a Gaussian random matrix U instead of $T = \frac{1}{\sqrt{d}}U$, however for the sake of clarity in the proof, we will work with normalized Gaussian random matrix, i.e., T, in the whole paper.

1.1. Our contributions

We concentrate on the use of Gaussian random projections as a tool for dimension reduction in the general Euclidean membership problem. We consider two special cases: S is at most countable (i.e. finite or countable) and S is of low doubling dimension.

In the first case, using a straightforward argument, we prove that these two problems (ESMP and PESMP) are equivalent almost surely regardless of the projected dimension. However, this result is only of theoretical interest, possibly due to round-off errors in floating point operations which make its practical application difficult. We address this issue by introducing a threshold $\tau > 0$ and study the corresponding "threshold" problem:

> Threshold ESMP (TESMP): Given b, S, T as above. Let $\tau > 0$. Decide whether $\|T(b) - T(S)\| \geq \tau$.

In the case when S may also be uncountable, we prove that ESMP and PESMP are also equivalent if the projected dimension d is proportional to some intrinsic dimension of the set S. In particular, we employ the notion of doubling dimension (defined later) to prove that, if $b \notin S$, then $T(b) \notin T(S)$ almost surely as long as the projected dimension $d \geq \mathcal{C}\,\mathsf{ddim}(S)$, where $\mathsf{ddim}(S)$ is the doubling dimension of S and $\mathcal{C}$ is some universal constant. We extend this result to the threshold case, and obtain a more useful bound for d. It turns out that, as a consequence of that result, we are able to improve a bound of Indyk-Naor on the Nearest Neigbour preserving embeddings by a factor of $\frac{\log(1/\delta)}{\varepsilon}$.

1.2. Applicability

Although this paper only makes theoretical contributions, the applicability of random projections is beyond doubt. They are routinely used in large-scale problems arising in clustering, for example for images and text. We are ourselves pursuing random projections in the context of linear and integer programming [22].

2. Finite and countable sets

In this section, we assume that S is either finite or countable. Let $U \in \mathbb{R}^{d \times m}$ be a random matrix drawn from Gaussian distribution, i.e. each entry of U is independently sampled from $\mathcal{N}(0, 1)$. It is well known that, for an arbitrary unit vector $a \in \mathbb{S}^{m-1}$, $\|Ua\|^2$ is a random variable with a Chi-squared distribution χ^2_d with d degrees of freedom [20]. Its corresponding density function is

$\frac{2^{-d/2}}{\Gamma(d/2)}x^{d/2-1}e^{-x/2}$, where $\Gamma(\cdot)$ is the gamma function. By [7], for any $0 < \delta < 1$, taking $z = \frac{\delta}{d}$ yields a cumulative distribution function (CDF)

$$F_{\chi_d^2}(\delta) \leq (ze^{1-z})^{d/2} < (ze)^{d/2} = \left(\frac{e\delta}{d}\right)^{d/2}. \tag{3}$$

Thus, we have

$$\mathsf{Prob}(\|Ua\| \leq \delta) = F_{\chi_d^2}(\delta^2) < (\frac{3}{d}\delta^2)^{d/2} \tag{4}$$

or, more simply, $\mathsf{Prob}(\|Ua\| \leq \delta) < \delta^d$ when $d \geq 3$.
In the case of a normalized Gaussian random projection $T = \frac{1}{\sqrt{d}}U$, we have

$$\mathsf{Prob}(\|Ta\| \leq \delta) = \mathsf{Prob}(\|Ua\| \leq \sqrt{d}\delta) < (3\delta^2)^{\frac{d}{2}} \tag{5}$$

Using this estimation, we immediately obtain the following result.

2.1 Proposition
Given $b \in \mathbb{R}^m$ and $S \subseteq \mathbb{R}^m$, at most countable, such that $b \notin S$. Then, for any $d \geq 1$ and for a normalized Gaussian random projection $T : \mathbb{R}^m \to \mathbb{R}^d$ ($T = \frac{1}{\sqrt{d}}U$) , we have $T(b) \notin T(S)$ almost surely, i.e. $\mathsf{Prob}\big(T(b) \notin T(S)\big) = 1$.

Proof. First, note that for any $a \neq 0$, $Ta \neq 0$ holds almost certainly. Indeed, without loss of generality we can assume that $\|a\| = 1$. Then for any $0 < \delta < 1$:

$$\mathsf{Prob}\big(T(a) = 0\big) \leq \mathsf{Prob}\big(\|Ta\| \leq \delta\big) \overset{(5)}{<} (3\delta^2)^{d/2} \to 0 \text{ as } \delta \to 0. \tag{6}$$

It means that for any $y \neq b$, then $T(y) \neq T(b)$ almost surely. Since the event $T(b) \notin T(S)$ can be written as the intersection of at most countably many "almost sure" events $T(b) \neq T(y)$ (for $y \in S$), it follows that $\mathsf{Prob}\big(T(b) \notin T(S)\big) = 1$, as claimed. □

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. $d = 1$) and study an equivalent membership problem on a line. This idea, stated somewhat differently in terms of random aggregations of linear constraints, appears to be part of the folklore of linear and integer programming.

It turns out that this result remains true for a large class of random projections.

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

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

$$\mathsf{Prob}(T(s) = 0) = \nu(\mathcal{E}_s) = \int_{\mathcal{E}_s} f d\mu \leq \|f\|_\infty \int_{\mathcal{E}_s} d\mu = 0$$

where μ denotes the Lebesgue measure on $\mathbb{R}^m$. The proof then follows by the countability of S, similarly to Proposition 2.1. $\square$

Appealing as the idea may be, it does not seem to work in practice. We tested it by considering the ESMP given by the Integer Feasibility Problem (IPF) defined on the set $\{x \in \mathbb{Z}^n_+ \cap [L, U] \mid Ax = b\}$, where $L, U \in \mathbb{Q}^n$ and, for all $i = 1, \cdots, n,\ L_i \leq U_i$. 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, the reason behind this failure is possibly due to the round-off error associated to the floating point representation used in computers. Specifically, when $T(A)x$ is too close (but not equal) 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 $T(b) \notin T(S)$, we ensure that

$$\mathsf{dist}(T(b), T(S)) = \min_{x \in S} \|T(b) - T(x)\| > \tau, \tag{7}$$

where 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 $\tau > 0$, $0 < \delta < 1$ and $b \notin S \subseteq \mathbb{R}^m$, where S is a finite set. Let $R = \min_{x \in S} \|b - x\| > 0$ and $T : \mathbb{R}^m \to \mathbb{R}^d$ be a normalized Gaussian random projection with $d \geq \frac{\log(|S|/\delta)}{\log(R/(\sqrt{3}\tau))}$. Then:

$$\mathsf{Prob}\big(\min_{x \in S} \|T(b) - T(x)\| > \tau\big) > 1 - \delta.$$

Proof. For any $x \in S$, by the linearity of T, we have:

$$\begin{aligned}
\mathsf{Prob}\big(\|T(b - x)\| \leq \tau\big) &= \mathsf{Prob}\left(\left\|T\Big(\frac{b - x}{\|b - x\|}\Big)\right\| \leq \frac{\tau}{\|b - x\|}\right) \\
&\leq \mathsf{Prob}\left(\left\|T\Big(\frac{b - x}{\|b - x\|}\Big)\right\| \leq \frac{\tau}{R}\right) < \frac{(\sqrt{3}\tau)^d}{R^d},
\end{aligned}$$

due to (5). Therefore, by the union bound,

$$\begin{aligned}
\mathsf{Prob}\big(\min_{x \in S} \|T(b) - T(x)\| > \tau\big) &= 1 - \mathsf{Prob}\big(\min_{x \in S} \|T(b) - T(x)\| \leq \tau\big) \\
&\geq 1 - \sum_{x \in S} \mathsf{Prob}\big(\|T(b) - T(x)\| \leq \tau\big) \\
&> 1 - |S| \frac{(\sqrt{3}\tau)^d}{R^d}.
\end{aligned}$$

The RHS is greater than or equal to $1 - \delta$ if and only if $\frac{R^d}{(\sqrt{3}\tau)^d} \geq \frac{|S|}{\delta}$, which is equivalent to $d \geq \frac{\log(\frac{|S|}{\delta})}{\log(\frac{R}{\sqrt{3}\tau})}$, as claimed. $\square$

Note that R is often unknown and can be arbitrarily small. However, if both b, S are integral and $\tau < 1$, then $R \geq 1$ and we can select $d > \frac{\log \frac{|S|}{\delta}}{\log \frac{1}{\sqrt{3}\tau}}$ in the above proposition.

3. Sets with low doubling dimensions

In many real-world applications, the data set S is not finite or countable, but it lies in some intrinsically low-dimensional space. There are many examples of such sets, including human motion records, speed signals, image and text data and more [6, 3]. Random projections can provide a tool to extract the full information of the set, in spite of the (high) dimension of the ambient space that it is embedded in.

In this section, we will use *doubling dimension* as a measure for the intrinsic dimension of a set. Let denote by $B(x, r) = \{y \in S : \|y - x\| \leq r\}$, i.e. the closed ball centered at x with radius $r > 0$ (w.r.t S). We use the following definition:

3.1 Definition
The doubling constant *of a set* $S \subseteq \mathbb{R}^m$ *is the smallest number* λ_S *such that any closed ball in* S *can be covered by at most* λ_S *closed balls of half the radius. A set* $S \subseteq \mathbb{R}^m$ *is called a* doubling set *if it has a finite doubling constant. The number* $\log_2(\lambda_S)$ *is then called the* doubling dimension *of* S *and denoted by* $\mathsf{ddim}(S)$.

One popular example of doubling spaces is a Euclidean space $\mathbb{R}^m$. In this case, it is well-known that its doubling dimension is $O(m)$ [21, 11]. However, there are cases where the set $S \subseteq \mathbb{R}^m$ is of much lower doubling dimension. It is also easy to see that $\mathsf{ddim}(S)$ does not depend on the dimension m. The doubling dimension is therefore a powerful tool for reducing dimensions in several classes of problems such as nearest neighbor [15, 13], low distortion embeddings [2], and clustering [16].

In this section, we assume that $\mathsf{ddim}(S)$ is relatively small compared to m. Note that, computing the doubling dimension of S is generally **NP**-hard [9], although it can be approximated within a constant factor [10]. For simplicity, we assume that λ_S is a power of 2, i.e. the doubling dimension of S is a positive integer number.

We shall make use of the two following lemmas.

3.2 Lemma
For any $b \in S$ *and* $\varepsilon, r > 0$, *there is a set* $S_0 \subseteq S$ *of size at most* $\lambda_S^{\lceil \log_2(\frac{r}{\varepsilon}) \rceil}$ *such that*

$$B(b, r) \subseteq \bigcup_{s \in S_0} B(s, \varepsilon). \tag{8}$$

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

3.3 Lemma
Let $S \subseteq \{s \in \mathbb{R}^m | \, \|s\| \leq 1\}$ *be a subset of the* m*-dimensional Euclidean unit ball. Let* $T : \mathbb{R}^m \to \mathbb{R}^d$ *be a normalized Gaussian random projection. Then*

there exist universal constants $c, C > 0$ such that for $d \geq C \log \lambda_S$ and $\delta \geq 6$, the following holds:

$$\mathsf{Prob}(\exists s \in S \ s.t.\ \|Ts\| > \delta) < e^{-cd\delta^2}. \tag{9}$$

This lemma is proved in [13] using concentration estimations for sum of squared gaussian variables (Chi-squared distribution). In particular, we recall an important inequality, proved in [13] : for all $a \in \mathbb{R}^m$ of unit norm, all $\delta > 0$ and a mapping T as above,

$$\mathsf{Prob}\ (|\|Ta\| - 1| > \delta) \leq e^{-d\delta^2/8} \quad \text{and} \tag{10}$$

For the sake of completeness, we will present the original proof [13] of the above Lemma. Here, we use an additional requirement that $\delta \geq 6$ instead of $\delta > 1$, however the main argument is unchanged.

Proof. Choose $b = 0 \in S, r = 1$ and $\varepsilon_k = \frac{1}{2^k}$ in Lemma 3.2. By earlier convention that $B(x,r) = \{y \in S : \|y-x\| \leq r\}$, obviously we have $S \subseteq B(0,1)$. Then there is a set $S_k \subseteq S$ of size at most λ_S^k such that

$$S \subseteq \bigcup_{s \in S_k} B(s, \frac{1}{2^k}). \tag{11}$$

Therefore, for any $x \in S$, there is a sequence $(x_n)_{n\in\mathbb{N}}$ that converges to x, with $\forall n \in \mathbb{N}, x_n \in S_n$ and $\|x_n - x_{n+1}\| \leq \frac{1}{2^n}$ (this holds because each ball $B(x_n, \frac{1}{2^n})$ is itself covered by balls of radius $\frac{1}{2^{n+1}}$ with centers in S_{n+1}). We claim that, if $\|Tx\| \geq \delta$, then there must be some $k \in \mathbb{N}$ such that

$$\|T(x_k - x_{k+1})\| \geq \frac{\delta}{3}(\frac{3}{2})^{-k}$$

Indeed, if no such k exists, then

$$\delta \leq \|Tx\| \leq \sum_{k=0}^{\infty} \|T(x_k - x_{k+1})\| < \frac{\delta}{3}\sum_{k=0}^{\infty}(\frac{3}{2})^{-k} = \delta, \text{ a contradiction.}$$

Now, if we want to neglect x, we can treat x_k and x_{k+1} (found above) as two points u, v, in which $u \in S_k$ and $v \in B(u, \frac{1}{2^k}) \cap S_{k+1}$. Then we have

$$\|Tu - Tv\| \geq \frac{\delta}{3}(\frac{3}{2})^{-k} \geq \frac{\|u-v\|}{2^{-k}}\frac{\delta}{3}(\frac{3}{2})^{-k} = \frac{\delta}{3}(\frac{4}{3})^k\|u-v\|.$$

Therefore, we have

$$\begin{aligned}
& \mathsf{Prob}\big(\exists x \in S \text{ s.t } \|T(x)\| > \delta\big) \\
\leq\ & \mathsf{Prob}\bigg(\exists x \in S, \exists k \geq 0 \text{ s.t } \|T(x_k - x_{k+1})\| > \frac{\delta}{3}(\frac{3}{2})^{-k}\bigg) \\
\leq\ & \sum_{k=0}^{\infty}\mathsf{Prob}\bigg(\exists u \in S_k,\ \exists v \in B(u, \frac{1}{2^k}) \cap S_{k+1} \text{ s.t } \|Tu - Tv\| > \frac{\delta}{3}(\frac{4}{3})^k\|u-v\|\bigg) \\
\leq\ & \sum_{k=0}^{\infty}\mathsf{Prob}\bigg(\exists u \in S_k,\ \exists v \in B(u, \frac{1}{2^k}) \cap S_{k+1} \text{ s.t } \|T(\frac{u-v}{\|u-v\|})\| > \frac{\delta}{3}(\frac{4}{3})^k\bigg) \\
\leq\ & \sum_{k=0}^{\infty}\lambda_S^{2k+1}\mathsf{Prob}\bigg(\|Tz\| > \frac{\delta}{3}(\frac{4}{3})^k\bigg) \quad \text{for any unit } z \text{ (by the union bound).}
\end{aligned}$$

Since $\delta \geq 6$, for any $k \geq 0$ we have $\frac{\delta}{3}(\frac{4}{3})^k \geq 1 + \frac{\delta}{6}(\frac{4}{3})^k$. Therefore, from the inequality (10), the above expression is less than or equal to

$$\sum_{k=0}^{\infty} \lambda_S^{2k+1} \mathsf{Prob}\big(|\|Tz\| - 1| \geq \frac{\delta}{6}(\frac{4}{3})^k\big) \;\leq\; \sum_{k=0}^{\infty} \lambda_S^{2k+1} e^{-d\frac{\delta^2}{8*6^2}(\frac{4}{3})^{2k}} \;\leq\; e^{-cd\delta^2} \tag{12}$$

as long as $d \geq C \log(\lambda_S)$ for some universal constants c, C. □

We now state one of the main results in this paper. It says that we can maintain the equivalence between ESMP and its projected version by using Gaussian random projections with d proportional to the doubling dimension of S.

3.4 Theorem
Given $b \notin S \subseteq \mathbb{R}^m$ where S is a closed doubling set. Let $T : \mathbb{R}^m \to \mathbb{R}^d$ be a normalized Gaussian random projection. Then

$$\mathsf{Prob}(T(b) \notin T(S)) = 1 \tag{13}$$

if $d \geq \mathcal{C} \log_2(\lambda_S)$, for some universal constant $\mathcal{C}$.

Proof. Let $d \geq \mathcal{C} \log_2(\lambda_S)$ for some universal constant $\mathcal{C}$ (large). The proof is divided in three steps: first we build a sequence of annuli, $(X_k)_{k\in\mathbb{N}}$ that cover S; then on each annulus, X_k, we estimate the probability that there exists an element $x \in X_k$ such that $T(x) = T(b)$. Finally, we apply the union bound on the whole set S and prove that the probability that there exists $x \in S$ such that $T(x) = T(b)$ can be as small as needed. Assume that $R > 0$ is the distance between b and the set S. Let $\varepsilon_i, \Delta_i, i \in \mathbb{N}$ and $R = r_0 < r_1 < r_2 < \ldots$ be positive scalars (their values will be defined later). For each $k \in \mathbb{N}^*$ we define an annulus

$$X_k = S \cap B(b, r_k) \setminus B(b, r_{k-1}).$$

Since $X_k \subseteq S \cap B(b, r_k)$, by Lemma 3.2 we can find a point set $S_k \subseteq S$ of size $|S_k| \leq \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil}$ such that

$$X_k \subseteq \bigcup_{s \in S_k} B(s, \varepsilon_k).$$

Hence, for any $x \in X_k$, there is $s \in S_k$ such that $\|x - s\| < \varepsilon_k$. Moreover, by the triangle inequality, any such s satisfies $r_{k-1} - \varepsilon_k < \|s - b\| < r_k + \varepsilon_k$ (since x is inside the annulus X_k). So without loss of generality we can assume that

$$S_k \subseteq B(b, r_k + \varepsilon_k) \setminus B(b, r_{k-1} - \varepsilon_k).$$

Using the union bound, we have:

$$\begin{aligned} \mathsf{Prob}\big(\exists x \in S \text{ s.t. } T(x) = T(b)\big) &= \mathsf{Prob}\big(\exists x \in \bigcup_{k=1}^{\infty} X_k \text{ s.t. } T(x) = T(b)\big) \\ &\leq \sum_{k=1}^{\infty} \mathsf{Prob}\big(\exists x \in X_k \text{ s.t. } T(x) = T(b)\big). \end{aligned}$$

Now, we will try to estimate the individual probabilities inside this sum. More precisely we will consider two cases according to the following event: for each $k \geq 1$, we denote by $\mathcal{E}_k$ the event that:

$$\exists s \in S_k,\ \exists x \in X_k \cap B(s, \varepsilon_k) \text{ s.t. } \|Ts - Tx\| > \Delta_k \text{ (with } \Delta_k \geq 6\varepsilon_k).$$

Then we have

$$\begin{aligned} & \mathsf{Prob}\big(\exists x \in X_k \text{ s.t. } T(x) = T(b)\big) \\ \leq\ & \mathsf{Prob}\big((\exists x \in X_k \text{ s.t. } T(x) = T(b)) \wedge \bar{\mathcal{E}_k}\big) + \mathsf{Prob}(\mathcal{E}_k) \end{aligned} \tag{14}$$

For the second term in (14), by the union bound, we have

$$\begin{aligned} \mathsf{Prob}(\mathcal{E}_k) &\leq \sum_{s \in S_k} \mathsf{Prob}\big(\exists x \in X_k \cap B(s, \varepsilon_k) \text{ s.t. } \|Ts - Tx\| > \Delta_k\big) \\ &\leq \sum_{s \in S_k} e^{-c_1 d (\frac{\Delta_k}{\varepsilon_k})^2} \qquad \text{(for } c_1 \text{ a univ. constant as in Lemma 3.3)} \\ &\leq \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil}\ e^{-c_1 d (\frac{\Delta_k}{\varepsilon_k})^2}. \end{aligned}$$

(Note that here we must choose $\Delta_k \geq 6\varepsilon_k$ in order to apply Lemma 3.3).

For the first term in (14), we have

$$\begin{aligned} & \mathsf{Prob}\big((\exists x \in X_k \text{ s.t. } T(x) = T(b)) \wedge \bar{\mathcal{E}_k}\big) \\ \leq\ & \mathsf{Prob}\big(\exists x \in X_k, s \in S_k \cap B(x, \varepsilon_k) \text{ s.t. } T(x) = T(b) \wedge \|T(s) - T(x)\| \leq \Delta_k\big) \\ \leq\ & \mathsf{Prob}\big(\exists s \in S_k \text{ s.t. } \|T(s) - T(b)\| < \Delta_k\big) \\ \leq\ & \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil}\ \mathsf{Prob}\big(\|T(z)\| < \frac{\Delta_k}{r_{k-1} - \varepsilon_k}\big) \quad \text{for some unit vector } z \\ \leq\ & \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil}\ \big(\frac{\sqrt{3}\Delta_k}{r_{k-1} - \varepsilon_k}\big)^d \quad \text{(by inequality 5).} \end{aligned}$$

Putting all the estimations, we have obtained, together, we have:

$$\mathsf{Prob}\big(\exists x \in S \text{ s.t } T(x) = T(b)\big) \leq \sum_{k=1}^{\infty} \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil} \left(e^{-c_1 d (\frac{\Delta_k}{\varepsilon_k})^2} + \big(\frac{\sqrt{3}\Delta_k}{r_{k-1} - \varepsilon_k}\big)^d \right). \tag{15}$$

Next, we will show that there are choices of $\varepsilon_k, \Delta_k, r_k$ such that the RHS of (15) can be as small as needed.

Choices of $\varepsilon_k, \Delta_k, r_k$: For some $N > 1$ large, we will choose $\varepsilon_k, \Delta_k, r_k$ as follows:

1. $\varepsilon_k = \varepsilon$, for some $\varepsilon > 0$.
2. $\Delta_k = \frac{Nk\varepsilon}{\sqrt{3}}$
3. $r_k = (N^2(k+1)^2 + 1)\varepsilon$

Now the RHS of (15) can be rewritten as follows

$$\begin{aligned} & \sum_{k=1}^{\infty} \lambda_S^{\lceil \log_2(N^2(k+1)^2+1) \rceil} \left(e^{-c_1 d (Nk)^2} + \big(\frac{1}{Nk}\big)^d \right) \\ \leq\ & \sum_{k=1}^{\infty} \lambda_S^{3 \log_2(Nk)} \left(e^{-c_1 d (Nk)^2} + \big(\frac{1}{Nk}\big)^d \right) \\ \leq\ & \sum_{k=1}^{\infty} 2^{3\,\mathsf{ddim}(S) \log_2(Nk)} \left(e^{-c_1 d (Nk)^2} + \big(\frac{1}{Nk}\big)^d \right). \end{aligned} \tag{16}$$

Note that $2^{3\,\mathsf{ddim}(S)\log_2(Nk)}$ does not increase fast enough compared to the decreasing speeds of both $e^{-c_1 d(Nk)^2}$ and $(\frac{1}{Nk})^d$ when $d \geq \mathcal{C}\mathsf{ddim}(S)$ with $\mathcal{C}$ large enough (and also independent of N). Therefore, there are universal constants $c_2, c_3 > 0$ such that the value of (16) is less than or equal to

$$\sum_{k=1}^{\infty} e^{-c_2(Nk)^2} + \sum_{k=1}^{\infty} (\frac{1}{Nk})^{c_3 d} \tag{17}$$

Both the two infinite sums tend to 0 when N tends to ∞. This means that

$$\mathsf{Prob}\big(\exists x \in S \text{ s.t. } T(x) = T(b)\big) = 0,$$

which proves our theorem. □

Our next result in this section is an extension of Thm. 3.4 to the threshold case.

3.5 Theorem
Let $b \notin S$ where $S \subseteq \mathbb{R}^m$ is a closed doubling set, $T : \mathbb{R}^m \to \mathbb{R}^d$ be a normalized Gaussian random projection, and $r = \min_{x\in S} \|b - x\|$. Let $\kappa < 1$ be some fixed constant. Then for all $0 < \delta < 1$ and $0 < \tau < \kappa r$, we have

$$\mathsf{Prob}(\mathsf{dist}(T(b), T(S)) > \tau) > 1 - \delta$$

if the projected dimension is $d = \Omega(\frac{\log(\frac{\lambda_S}{\delta})}{1-\kappa})$.

Proof. We follow the same scheme, as in the proof of Theorem 3.4. Let $d \geq \mathcal{C}(\frac{\log(\frac{\lambda_S}{\delta})}{1-\kappa})$ for some universal constant $\mathcal{C}$ (large). As before, let $\varepsilon_i, \Delta_i, i \in \mathbb{N}$ and $r = r_0 < r_1 < r_2 < \ldots$ be positive scalars whose values will be decided later. The annuli X_k and point sets S_k are also defined similarly. Using the union bound, now we have:

$$\begin{aligned}\mathsf{Prob}\big(\exists x \in S \text{ s.t. } \|T(x) - T(b)\| < \tau\big) &= \mathsf{Prob}\big(\exists x \in \bigcup_{k=1}^{\infty} X_k \text{ s.t. } \|T(x) - T(b)\| < \tau\big)\\ &\leq \sum_{k=1}^{\infty} \mathsf{Prob}\big(\exists x \in X_k \text{ s.t. } \|T(x) - T(b)\| < \tau\big).\end{aligned}$$

Now, we will try to estimate the individual probabilities inside this sum.
For each $k \geq 1$, we denote by $\mathcal{E}_k$ the event that:

$$\exists s \in S_k,\ \exists x \in X_k \cap B(s, \varepsilon_k) \text{ s.t. } \|Ts - Tx\| > \Delta_k \text{ (with } \Delta_k \geq 6\varepsilon_k). \tag{18}$$

Then we have

$$\begin{aligned}&\mathsf{Prob}\big(\exists x \in X_k \text{ s.t. } \|T(x) - T(b)\| < \tau\big)\\ \leq\ &\mathsf{Prob}\big((\exists x \in X_k \text{ s.t. } \|T(x) - T(b)\| < \tau) \wedge \bar{\mathcal{E}_k}\big) + \mathsf{Prob}(\mathcal{E}_k)\end{aligned} \tag{19}$$

For the second term in (19), from the previous proof, we already had:

$$\mathsf{Prob}(\mathcal{E}_k) \leq \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k})\rceil}\ e^{-c_1 d(\frac{\Delta_k}{\varepsilon_k})^2}. \tag{20}$$

(Note that here we must choose $\Delta_k \geq 6\varepsilon_k$ in order to apply Lemma 3.3).

Now, for the first term in (19), we have

$$\begin{aligned}
& \mathsf{Prob}\big((\exists x \in X_k \text{ s.t. } \|T(x) - T(b)\| < \tau) \wedge \bar{\mathcal{E}_k}\big) \\
\leq\ & \mathsf{Prob}\big(\exists x \in X_k, s \in S_k \cap B(x, \varepsilon_k) \text{ s.t. } \|T(x) - T(b)\| < \tau \wedge \|T(s) - T(x)\| \leq \Delta_k\big) \\
\leq\ & \mathsf{Prob}\big(\exists s \in S_k \text{ s.t. } \|T(s) - T(b)\| < \Delta_k + \tau\big) \qquad \text{(by triangle inequality)} \\
\leq\ & \begin{cases} \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil}\ \mathsf{Prob}\big(\|T(z)\| < \frac{\Delta_k+\tau}{r_{k-1}-\varepsilon_k}\big) & \text{for some unit vector } z \quad \text{if } k \geq 2 \\ \lambda_S^{\lceil \log_2(\frac{r_1}{\varepsilon_1}) \rceil}\ \mathsf{Prob}\big(\|T(z)\| < \frac{\Delta_1+\tau}{r}\big) & \text{for some unit vector } z \quad \text{if } k = 1 \end{cases} \\
\leq\ & \begin{cases} \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil}\ (\frac{\Delta_k+\tau}{r_{k-1}-\varepsilon_k})^d & \text{if } k \geq 2 \\ \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil}\ (\frac{\Delta_1+\tau}{r})^d & \text{if } k = 1. \end{cases}
\end{aligned}$$

Putting all the estimations we have obtained together, we have:

$$\begin{aligned}
& \mathsf{Prob}\big(\exists x \in S \text{ s.t. } \|T(x) - T(b)\| < \tau\big) \\
\leq\ & \left(\sum_{k=1}^{\infty} \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil}\ e^{-c_1 d (\frac{\Delta_k}{\varepsilon_k})^2} + \sum_{k=2}^{\infty} \lambda_S^{\lceil \log_2(\frac{r_k}{\varepsilon_k}) \rceil} (\frac{\Delta_k + \tau}{r_{k-1} - \varepsilon_k})^d\right) + \lambda_S^{\lceil \log_2(\frac{r_1}{\varepsilon_1}) \rceil}\ (\frac{\Delta_1 + \tau}{r})^d.
\end{aligned} \tag{21}$$

Here we separate one term out, and we will prove that the remaining expression can be made as small as wanted by certain choices of parameters.

Choices of $\varepsilon_k, \Delta_k, r_k$: Let $N > 0$ be the number such that $(\frac{7}{N} + 1) = \frac{r}{\tau}$. From the assumptions, we have $N = 7\frac{\tau}{r-\tau}$ and $0 < \tau < \kappa r$. Hence

$N < \frac{7\kappa r}{r-\tau} < \frac{7\kappa}{1-\frac{\tau}{r}} < \frac{7\kappa}{1-\kappa} < \frac{7}{1-\kappa}$. We will choose $\varepsilon_k, \Delta_k, r_k$ as follows:

1. $\varepsilon_k = \varepsilon = \frac{\tau}{N}$,
2. $\Delta_k = 6\sqrt{k}\varepsilon$,
3. $r_k = (6k + 7)\varepsilon + \sqrt{k+1}\,\tau$.

(Our purpose is to choose the parameters so that $\frac{\Delta_k+\tau}{r_{k-1}-\varepsilon_k} = \frac{1}{\sqrt{k}}$ and $\Delta_k \geq 6\varepsilon_k$).

From this choice, it is obvious $r_0 = r$. Now the RHS of (21) can be rewritten as follows:

$$\begin{aligned}
& \left(\sum_{k=1}^{\infty} \lambda_S^{\lceil \log_2(6k+7+N\sqrt{k+1}) \rceil}\ e^{-c_1 d(36k)} + \sum_{k=2}^{\infty} \lambda_S^{\lceil \log_2(6k+7+N\sqrt{k+1}) \rceil} (\frac{1}{\sqrt{k}})^d\right) + \\
+\ & \lambda_S^{\lceil \log_2(13+N\sqrt{2}) \rceil}\ ((\frac{6}{N} + 1)\frac{\tau}{r})^d \leq \\
\leq\ & \left(\sum_{k=1}^{\infty} \lambda_S^{c_3 \log_2(N(k+1))}\ e^{-c_1 d(36k)} + \sum_{k=2}^{\infty} \lambda_S^{c_3 \log_2(N(k+1))} (\frac{1}{\sqrt{k}})^d\right) + \\
+\ & \lambda_S^{c_2}\ ((\frac{6}{N} + 1)\frac{\tau}{r})^d
\end{aligned} \tag{22}$$

for some universal constants c_1, c_2, c_3.

It is easy to show that the expression in the big bracket is bounded above by $e^{-c_4 d}$ as long as $d \geq \mathcal{C} \log_2(\lambda_S) \log(\frac{7}{1-\kappa}) > \mathcal{C} \log_2(\lambda_S) \log(N)$ (for some large constants $c_4, \mathcal{C}$). Moreover,

$$e^{-c_4 d} \leq \frac{\delta}{2} \quad \text{if and only if} \quad d \geq \frac{1}{c_4} \log(\frac{2}{\delta}) \tag{23}$$

and $\lambda_S^{c_2}\,((\frac{6}{N}+1)\frac{\tau}{r})^d \leq \frac{\delta}{2}$ if and only if

$$d \geq \frac{c_2 \log(\lambda_S) + \log(\frac{2}{\delta})}{\log((\frac{N}{6+N})\frac{r}{\tau})}. \tag{24}$$

However, $\log((\frac{N}{6+N})\frac{r}{\tau}) = \log(\frac{7r}{6r+\tau}) \geq \log(\frac{7}{6+\kappa}) \geq \log(1+\frac{1-\kappa}{7}) \geq \frac{1-\kappa}{7}$, from the Taylor series of the logarithm function. Therefore, (24) holds if we select

$$d \geq \mathcal{C}\frac{\log(\frac{\lambda_S}{\delta})}{1-\kappa} \tag{25}$$

for some universal constants $\mathcal{C}$. The proof follows immediately from (23) and (25) by an application of the union bound. □

One of the interesting consequence of Theorem 3.5 is the following application to the Approximate Nearest Neighbour problem.

3.6 Corollary
For $T : \mathbb{R}^m \to \mathbb{R}^d$ be a normalized Gaussian random projection, $X \subseteq \mathbb{R}^m$, $\varepsilon \in (0, 1/2)$ and $\delta \in (0, 1/2)$, we can choose

$$d = \max\left\{O(\frac{\log(\frac{1}{\delta})}{\varepsilon^2}),\ O(\frac{\log(\frac{\lambda_S}{\delta})}{\varepsilon})\right\}$$

such that for every $x_0 \in X$, with probability at least $1-\delta$,

1. $\mathsf{dist}\ (Tx_0, T(X \setminus \{x_0\})) \leq (1+\varepsilon)\mathsf{dist}\ (x_0, X \setminus \{x_0\})$,
2. *Every $x \in X$ with $\|x_0 - x\| \geq (1+2\varepsilon)\mathsf{dist}\ (x_0, X\setminus\{x_0\})$ satisfies*

$$\|Tx_0 - Tx\| > (1+\varepsilon)\mathsf{dist}\ (x_0, X \setminus \{x_0\}).$$

Note that, this result improves the bound provided by Indyk-Naor in [13]. In that paper, the authors give the bound for the projected dimension to be

$$d = O\bigg(\frac{\log(2/\varepsilon)}{\varepsilon^2}\log(\frac{1}{\delta})\log(\lambda_S)\bigg),$$

which is significantly larger than our bound. *Proof.* Similar as the proof of Theorem 4.1 in [13], we have: for $d \geq \mathcal{C}\frac{\log(\frac{1}{\delta})}{\varepsilon^2}$ with some large constant $\mathcal{C}$:

$$\text{Prob}\ [\mathsf{dist}\ (Tx_0, T(X \setminus \{x_0\})) \leq (1+\varepsilon)\mathsf{dist}\ (x_0, X \setminus \{x_0\})] < \frac{\delta}{2}. \tag{26}$$

Now, assume that $\mathsf{dist}\ (x_0, X \setminus \{x_0\}) = 1$. Set $b = x_0$ and $S = \{x \in X : \|x - x_0\| \geq 1+2\varepsilon\}$ and $\tau = 1+\varepsilon$ as in Theorem 3.5. We then have $r := \min_{x\in S}\|x-b\| \geq 1+2\varepsilon$, which implies $\frac{\tau}{r} \leq \frac{1+\varepsilon}{1+2\varepsilon} = 1 - \frac{\varepsilon}{1+2\varepsilon} < 1 - \frac{\varepsilon}{2}$. Therefore, we can choose $\kappa = 1-\varepsilon/2$. Applying Theorem 3.5, we have:

$$\mathsf{Prob}(\mathsf{dist}(T(b), T(S)) \leq \tau) \leq \frac{\delta}{2} \tag{27}$$

if the projected dimension $d = \Omega(\frac{\log(\frac{\lambda_S}{\delta})}{1-\kappa}) = \Omega(\frac{\log(\frac{\lambda_S}{\delta})}{\varepsilon})$.

From (26) and (27), we conclude that the two required conditions hold for some

$$d = \max\left\{O(\frac{\log(\frac{1}{\delta})}{\varepsilon^2}),\ O(\frac{\log(\frac{\lambda_S}{\delta})}{\varepsilon})\right\},$$

as claimed. □

4. Conclusion

In this paper we discussed the possibility of applying random projections of the type used in Johnson-Lindenstrauss lemma to a very general class of set membership problems in Euclidean spaces. Our results are directly applicable, as it suffices to pre-multiply the input vectors by a randomly generated matrix, and yield the correct answer with high probability.

Acknowledgements
The authors are grateful to the Siebel Energy Institute for a seed grant.

References

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

[2] 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.

[3] E. Bingham and H. Mannila. Random projection in dimensionality reduction: Applications to image and text data. In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01, pages 245–250, New York, NY, USA, 2001. ACM.

[4] A. Blum. Random projection, margins, kernels, and feature-selection. In *Proceedings of the 2005 International Conference on Subspace, Latent Structure and Feature Selection*, SLSFS'05, pages 52–68, Berlin, Heidelberg, 2006. Springer-Verlag.

[5] S. Dasgupta. Experiments with random projection. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, UAI '00, pages 143–151, San Francisco, CA, USA, 2000. Morgan Kaufmann Publishers Inc.

[6] S. Dasgupta and Y. Freund. Random projection trees and low dimensional manifolds. In *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 537–546, New York, NY, USA, 2008. ACM.

[7] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65, January 2003.

[8] D. Fradkin and D. Madigan. Experiments with random projections for machine learning. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 517–522, New York, NY, USA, 2003. ACM.

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

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

[11] A. Gupta, R. Krauthgamer, and J. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *Proceedings of the 44th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, pages 534–, Washington, DC, USA, 2003. IEEE Computer Society.

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

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

[14] 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.

[15] 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.

[16] A. Magen. Dimensionality reductions in ℓ_2 that preserve volumes and distance to affine spaces. *Discrete and Computational Geometry*, 30(1):139–153, 2007.

[17] M. Mahoney. Randomized algorithms for matrices and data. *Found. Trends Mach. Learn.*, 3(2):123–224, February 2011.

[18] Jiri Matousek. *Lectures on Discrete Geometry*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2002.

[19] J. Matoušek. On variants of the Johnson-Lindenstrauss lemma. *Random Structures and Algorithms*, 33:142–156, 2008.

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

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

[22] K. Vu, P.-L. Poirion, and L. Liberti. Random projections for linear programming. *Mathematics of Operations Research (to appear)*, 2017.

[23] D. Woodruff. Sketching as a tool for numerical linear algebra. *Found. Trends Theor. Comput. Sci.*, 10:1–157, October 2014.