# Database-friendly Random Projections

Dimitris Achlioptas
*Microsoft Research, One Microsoft Way, Redmond WA, 98052, U.S.A.*

E-mail: optas@microsoft.com

A classic result of Johnson and Lindenstrauss asserts that any set of $n$ points in $d$-dimensional Euclidean space can be embedded into $k$-dimensional Euclidean space — where $k$ is logarithmic in $n$ and independent of $d$ — so that all pairwise distances are maintained within an arbitrarily small factor. All known constructions of such embeddings involve projecting the $n$ points onto a spherically random $k$-dimensional hyperplane through the origin. Here we give two novel constructions of such embeddings, having the additional property that all elements of the projection matrix belong in $\{-1, 0, +1\}$. This makes our constructions particularly well suited for database environments, as the computation of the embedding reduces to evaluating a single aggregate over $k$ random partitions of the attributes.

## 1. INTRODUCTION

Consider projecting the points of your favorite sculpture first onto the plane and then onto a single line. The result amply demonstrates the power of dimensionality.

Conversely, given a high-dimensional pointset it is natural to ask whether it could be embedded into a lower dimensional space without suffering great distortion. In this paper, we will consider this question for finite sets of points in Euclidean space. It will be convenient to think of $n$ points in $\mathbb{R}^d$ as an $n \times d$ table (matrix) $A$ with each point represented as a row (vector) with $d$ attributes (coordinates).

Given such a matrix representation of the pointset, one of the most commonly used embeddings is the one suggested by the Singular Value Decomposition of $A$. That is, in order to embed the $n$ points into $\mathbb{R}^k$ we project them onto the $k$-dimensional space spanned by the singular vectors corresponding to the $k$ largest singular values of $A$. If one rewrites the result of this projection as a (rank $k$) $n \times d$ matrix $A_k$, we are guaranteed that for any other $k$-dimensional pointset represented as an $n \times d$ matrix $D$,

$$|A - A_k|_F \leq |A - D|_F \ ,$$

where, for any matrix $Q$, $|Q|_F^2 = \sum Q_{i,j}^2$. To interpret this result observe that $|\cdot|_F$ measures the embedding's distortion as follows: for each point (row) consider the difference-vector between the original and the new position; the distortion is then the sum of the squared lengths of all such vectors. Thus, if moving a point by $z$ takes energy proportional to $z^2$, then $A_k$ represents the $k$-dimensional configuration reachable from $A$ with least energy.

It turns out that $A_k$ is also "optimal" under many other matrix norms. Specifically, it is well-known that for any rank $k$ matrix $D$ and for *any* rotationally invariant norm

$$|A - A_k| \leq |A - D| \ .$$

For each such norm, just as we did above for $|\cdot|_F$, one can give a natural interpretation of how it measures the *global* distortion resulting from the embedding. At the same time, though, in all these cases there are no guarantees whatsoever regarding *local* properties of the resulting embedding. For example, it is not hard to devise examples where the new distance between a pair of points is arbitrarily smaller than the original distance.

The lack of any such local guarantees makes it very hard to exploit such embeddings *algorithmically*. In a seminal paper [9], Linial, London and Rabinovich were the first to consider embeddings that respect local properties and their algorithmic applications. By now, such embeddings have become an important tool in algorithmic design.

A real gem in this area has been the following result of Johnson and Lindenstrauss [7].

LEMMA 1.1 ([7]). *Given $\epsilon > 0$ and an integer $n$, let $k$ be a positive integer such that $k \geq k_0 = O(\epsilon^{-2} \log n)$. For every set $P$ of $n$ points in $\mathbb{R}^d$ there exists $f : \mathbb{R}^d \to \mathbb{R}^k$ such that for all $u, v \in P$*

$$(1 - \epsilon)||u - v||^2 \leq ||f(u) - f(v)||^2 \leq (1 + \epsilon)||u - v||^2 \ .$$

We will refer to embeddings providing a guarantee akin to that of Lemma 1.1 as JL-embeddings. In the last few years, such embeddings have been useful in solving a variety of problems. The rough idea is the following. By providing a low dimensional representation of the data, JL-embeddings speed up certain algorithms dramatically, in particular algorithms whose run-time depends exponentially in the dimension of the working space (for a number of practical problems the best known algorithms indeed have such behavior). At the same time, the provided guarantee regarding pairwise distances often allows one to establish that the solution found by working in the low dimensional space is a good approximation to the solution in the original space. We give a few examples below.

Papadimitriou, Raghavan, Tamaki and Vempala [10], proved that embedding the points of $A$ in a low-dimensional space can significantly speed up the computation of a low rank approximation to $A$, without significantly affecting its quality. In [6], Indyk and Motwani showed that JL-embeddings are useful in solving the $\varepsilon$-approximate nearest neighbor problem, where (after some preprocessing of the pointset $P$) one is to answer queries of the following type: "Given an arbitrary point $x$, find a point $y \in P$, such that for every point $z \in P$, $||x - z|| \geq (1 - \varepsilon)||x - y||$." In a different vein, Schulman [11] used JL-embeddings as part of an approximation algorithm for the version of clustering where we seek to minimize the sum of the squares of intracluster distances. Recently, Indyk [5] showed that JL-embeddings can also be used in the context of "data-stream" computation, where one has limited memory and is allowed only a single pass over the data (stream).

### 1.1. Our contribution

Over the years, the probabilistic method has allowed for the original proof of Johnson and Lindenstrauss to be greatly simplified and sharpened, while at the same time giving conceptually simple randomized algorithms for constructing the embedding [4, 6, 3]. Roughly speaking, all such algorithms project the input points onto a spherically random

hyperplane through the origin. While conceptually simple, in practice all such algorithms amount to multiplying $A$ with a dense matrix of real numbers. This can be a non-trivial task in many practical computational environments. Moreover, it is mathematically interesting to investigate the precise role of spherical symmetry in our choice of hyperplane.

Our main result, below, asserts that one can replace projections onto random hyperplanes with much simpler and faster operations. In particular, in a database environment these operations can be implemented readily using standard SQL primitives without any additional functionality. Somewhat surprisingly, we prove that this comes without *any* sacrifice in the quality of the embedding. In fact, we will see that for every fixed value of $d$ we can get slightly better bounds than all current methods.

We state our result below as Theorem 1.1. Following that, we discuss how to compute the embedding in terms of database operations. As in Lemma 1.1, the parameter $\epsilon$ controls the accuracy in distance preservation, while now $\beta$ controls the probability of success.

THEOREM 1.1. *Let $P$ be an arbitrary set of $n$ points in $\mathbb{R}^d$, represented as an $n \times d$ matrix $A$. Given $\epsilon, \beta > 0$ let*

$$k_0 = \frac{4+2\beta}{\epsilon^2/2 - \epsilon^3/3} \log n \ .$$

*For integer $k \geq k_0$, let $R$ be a $d \times k$ random matrix with $R(i,j) = r_{ij}$, where $\{r_{ij}\}$ are independent random variables from either one of the following two probability distributions:*

$$r_{ij} = \begin{cases} +1 & \textit{with probability} \quad 1/2 \\ -1 & \cdot\cdot \qquad\qquad\quad 1/2 \end{cases},$$

$$r_{ij} = \sqrt{3} \times \begin{cases} +1 & \textit{with probability} \quad 1/6 \\ 0 & \cdot\cdot \qquad\qquad\quad 2/3 \\ -1 & \cdot\cdot \qquad\qquad\quad 1/6 \end{cases}.$$

*Let*

$$E = \frac{1}{\sqrt{k}} A R$$

*and let $f : \mathbb{R}^d \to \mathbb{R}^k$ map the $i^{th}$ row of $A$ to the $i^{th}$ row of $E$.*

*With probability at least $1 - n^{-\beta}$, for all $u, v \in P$*

$$(1-\epsilon)||u-v||^2 \leq ||f(u) - f(v)||^2 \leq (1+\epsilon)||u-v||^2 \ .$$

We see that to construct a JL-embedding via Theorem 1.1, we need a very simple probability distribution to generate the projection matrix, while the computation of the projection itself reduces to aggregate evaluation. Moreover, when $r_{ij} \in \{-1, +1\}$, the construction is also conceptually extremely simple. On the other hand, when $r_{ij} \in \{-1, 0, +1\}$, we get a threefold speedup, as we only need to process a third of all attributes for each of the $k$ coordinates.

**Database-friendliness.** To apply the theorem in a database system using, say, the second distribution above one needs to generate $k$ new attributes, each one formed by performing the same random experiment: throw away $2/3$ of the original attributes at random; partition the remaining attributes randomly into two equal parts; for each partition, produce a new attribute equal to the sum of all attributes; take the difference of the two sum-attributes.

**Projecting onto random lines.** Looking a bit more closely into the computation of the embedding we see that each row (vector) of $A$ is projected onto $k$ random vectors whose coordinates $\{r_{ij}\}$ are independent random variables with mean 0 and variance 1. If the $\{r_{ij}\}$ were independent Normal random variables with mean 0 and variance 1, it is well-known that each resulting vector would point to uniformly random direction in space. Projections onto such vectors have been considered in a number of settings, including the work of Kleinberg on approximate nearest neighbors [8] and of Vempala on learning intersections of halfspaces [12]. More recently, such projections have also been used in learning mixture of Gaussians models, starting with the work of Dasgupta [2] and later with the work of Arora and Kannan [1].

Our proof implies that for any fixed vector $\alpha$, the behavior of its projection onto a random vector $c$ is mandated by the even moments of the random variable $||\alpha \cdot c||$. In fact, our result follows by showing that for every vector $\alpha$, under our distributions for $\{r_{ij}\}$, these moments are dominated by the corresponding moments for the case where $c$ is spherically symmetric. As a result, projecting onto vectors whose entries are distributed like the columns of matrix $R$ could replace projection onto spherically random vectors; it is computationally simpler and results in projections that are at least as nicely behaved.

**Randomization.** Perhaps a naive attempt at constructing JL-embeddings would be to pick $k$ of the original coordinates in $d$-dimensional space as the new coordinates. Naturally, as two points can be very far apart while differing only along a single original dimension, this approach is doomed. At the same time, though, if for each pair of points, all coordinates contributed "roughly equally" to the their distance, then a sampling scheme as above would make a lot of sense. Thus, it is very natural to first apply a random *rotation* to the original pointset in $\mathbb{R}^d$ and then, say, pick the first $k$ of the resulting coordinates as our new coordinates. Of course, this is exactly the same as projecting onto spherically random $k$-dimensional hyperplane! The random rotation can be viewed as a form of insurance, similar to the random permutation usually applied before applying Quicksort.

**Derandomization.** Finally, we note that Theorem 1.1 allows one to use significantly fewer random bits than all previous methods for constructing JL-embeddings. While the amount of randomness needed is still quite large, such attempts for randomness reduction are of independent interest and our result can be viewed as a first step in that direction.

## 2. PREVIOUS WORK

As we will see, in all methods for producing JL-embeddings, including ours, the heart of the matter is showing that for any vector, the squared length of its projection is sharply concentrated around its expected value. The original proof of Johnson and Lindenstrauss [7] uses quite heavy geometric approximation machinery to yield such a concentration bound. That proof was greatly simplified and sharpened by Frankl and Meahara [4] who explicitly considered a projection onto $k$ random orthonormal vectors (as opposed to viewing such vectors as the basis of a random hyperplane), yielding the following result.

THEOREM 2.1 ([4]). *For any $\epsilon \in (0, 1/2)$, any sufficiently large set $P \in \mathbb{R}^d$, and $k \geq k_0 = \lceil 9(\epsilon^2 - 2\epsilon^3/3)^{-1} \log |P| \rceil + 1$, there exists a map $f : P \to \mathbb{R}^k$ such that for all $u, v \in P$,*

$$(1-\epsilon)||u-v||^2 \leq ||f(u)-f(v)||^2 \leq (1+\epsilon)||u-v||^2 \ .$$

The next great simplification of the proof of Lemma 1.1 was given, independently, by Indyk and Motwani [6] and Dasgupta and Gupta [3], the latter also giving a slight sharpening of the bound for $k_0$. By combining the analysis of [3] with the viewpoint of [6] it is in fact not hard to show that Theorem 1.1 holds if for all $i, j$, $r_{ij} \stackrel{D}{=} N(0,1)$. Below we state our rendition of how each of these simplifications were achieved as it prepares the ground for our own work. Let us write $X \stackrel{D}{=} Y$ to denote that $X$ is distributed as $Y$ and recall that $N(0,1)$ denotes the standard Normal random variable having mean 0 and variance 1.

[6]: Assume that we try to implement the scheme of Frankl and Maehara [4] but we are lazy about enforcing either normality (unit length) or orthogonality among our $k$ vectors. Instead, we just pick our $k$ vectors independently, in a spherically symmetric manner, by taking as the coordinates of each vector independent $N(0,1)$ random variables and then merely scaling each vector by $1/\sqrt{d}$ so that its expected length is 1.

An immediate gain of this approach is that now, for any fixed vector $\alpha$, the length of its projection onto each of our vectors is also a Normal random variable. This is due to a powerful and deep fact, namely the 2-stability of the Gaussian distribution: for any real numbers $\alpha_1, \alpha_2, \ldots, \alpha_d$, if $\{Z_i\}_{i=1}^d$ is a family of independent Normal random variables and $X = \sum_{i=1}^d \alpha_i Z_i$, then $X \stackrel{D}{=} c\,N(0,1)$, where $c = (\alpha_1^2 + \cdots + \alpha_d^2)^{1/2}$. As a result, if we take these $k$ projection lengths to be the coordinates of the embedded vector in $\mathbb{R}^k$, then the squared length of the embedded vector follows the Chi-square distribution for which strong concentration bounds are readily available.

Remarkably, very little is lost due to our laziness. Although, we did not explicitly enforce either orthogonality, or normality, the resulting $k$ vectors, with high probability, will come very close to having both of these properties. In particular, the length of each of the $k$ vectors is sharply concentrated (around 1) as the sum of $d$ independent random variables. Moreover, since the $k$ vectors point in uniformly random directions in $\mathbb{R}^d$, they get rapidly closer to being nearly orthogonal as $d$ grows.

[3]: Here we will exploit spherical symmetry without appealing directly to the 2-stability of the Gaussian distribution. Instead observe that, by symmetry, the projection of any unit vector $\alpha$ on a random hyperplane through the origin is distributed exactly like the projection of a random point from the surface of the $d$-dimensional sphere onto a fixed subspace of dimension $k$. Such a projection can be studied readily, though, as now each coordinate is a scaled Normal random variable. With a somewhat tighter analysis than [6], this approach gave the strongest known bound, namely $k \geq k_0 = (4+2\beta)(\epsilon^2/2 - \epsilon^3/3)^{-1} \log n$, which is exactly the same as the bound in Theorem 1.1.

## 3. SOME INTUITION

Our contribution begins with the realization that spherical symmetry, while making life extremely comfortable, is not essential. What is essential is concentration. So, at least in principle, one is free to consider other candidate distributions for the $\{r_{ij}\}$, if perhaps at the expense of comfort.

As we saw earlier, each column of our matrix $R$ will give us a coordinate of the projection in $\mathbb{R}^k$ and thus the squared length of the projection is merely the sum of the squares of these coordinates. So, effectively, the projection is equivalent to the following: each column acts as an independent estimator of the original vector's length (its estimate being the inner product with it) and in the end we take the consensus estimate (sum) of our $k$ estimators.

Seen from this angle, requiring the $k$ vectors to be orthonormal has the pleasant statistical overtone of "maximizing mutual information" (since all estimators have equal weight and are orthogonal). Nonetheless, even if we only require that each column simply gives an unbiased, bounded variance estimator, the Central Limit Theorem implies that if we take sufficiently many columns, we can get an arbitrarily good estimate of the original length. Naturally, the number of columns needed, depends on the variance of the estimators.

From the above we see that the key issue is the concentration of the projection of an arbitrary fixed vector $\alpha$ onto a single random vector. The main technical difficulty, resulting from giving up spherical symmetry, is that this concentration can depend on $\alpha$. Our technical contribution lies in determining probability distributions for the $\{r_{ij}\}$ under which, for all vectors, this concentration is at least as good as in the spherically symmetric case. In fact, it will turn out that for every *fixed* value of $d$, we can get a (minuscule) improvement of concentration. Thus, for every fixed $d$, we can actually get a *strictly better* bound for $k$, albeit marginally, than by taking spherically random vectors.

The reader might be wondering "how can it be that perfect spherical symmetry does not buy us anything (and is in fact slightly worse for each fixed $d$)?". To at least show why we don't lose too much by giving up spherical symmetry, we have the following intuitive argument. When all vectors are not equal with respect to the variability of the length of their projection an adversary could try to pick a worst-case such vector $w$. So, we can rephrase the question as "How much are we empowering the adversary by committing to picking our column vectors among lattice points rather than arbitrary points in $\mathbb{R}^d$?".

As we will see, and this lies at the heart of our proof, the worst-case vector $w$ is $\frac{1}{\sqrt{d}}\ (1,\ldots,1)$ (along with all $2^d$ vectors resulting by sign-flipping $w$'s coordinates). So, the worst-case vector, at least in terms of the magnitudes of its coordinates, turns out to be a more or less "typical" vector, unlike say $(1,0,\ldots,0)$. Therefore, it is not hard to believe that the adversary would not fare much worse by replacing $w$ with a spherically random vector. In that case, though, the adversary does not benefit at all from our commitment!

To get a more satisfactory answer, it seems like one has to delve into the proof. In particular, both for the spherically random case and for our distributions, the bound on $k$ is mandated by the probability of *over*estimating the projected length. Thus, the "bad events" amount to the spanning vectors being too "well-aligned" with $\alpha$. Now, in the spherically symmetric setting it is possible to have alignment that is arbitrarily close to perfect, albeit with correspondingly smaller probability. In our case, if we don't have perfect alignment then we are guaranteed a certain, bounded amount of misalignment. It is precisely this tradeoff between the probability and the extent of alignment that drives the proof.

Consider, for example, the case when $d = 2$ with $r_{ij} \in \{-1, +1\}$. As we said above, the worst case vector is $w = (1/\sqrt{2})(1,1)$. So, with probability $1/2$ we have perfect alignment (when our random vector is $\pm w$) and with probability $1/2$ we have orthogonality. On the other hand, for the spherically symmetric case, we have to consider the integral over all points on the plane, weighted by their probability under the two-dimensional Gaussian distribution. It's a rather instructive exercise to explore this tradeoff directly and might also give the interested reader some intuition for the general case.

## 4. PRELIMINARIES AND THE SPHERICALLY SYMMETRIC CASE

### 4.1. Preliminaries

Let $x \cdot y$ denote the inner product of vectors $x, y$. To simplify notation in the calculations we will work with matrix $R$ scaled by $1/\sqrt{d}$ and, as a result, to get $E$ we need to scale

$A \times R$ by $\sqrt{d/k}$ rather than $1/\sqrt{k}$. So, $R$ is a random $d \times k$ matrix with $R(i,j) = r_{ij}/\sqrt{d}$, where the $\{r_{ij}\}$ are distributed as in Theorem 1.1. Therefore, if $c_j$ denotes the $j^{th}$ column of $R$, then $\{c_j\}_{j=1}^k$ is a family of $k$ i.i.d. random unit vectors in $\mathbb{R}^d$ and for all $\alpha \in \mathbb{R}^d$, $f(\alpha) = \sqrt{d/k}\ (\alpha \cdot c_1, \ldots, \alpha \cdot c_d)$.

In practice, of course, such scaling can be postponed until after the matrix multiplication (projection) has been performed, so that we maintain the advantage of only having $\{-1, 0, +1\}$ in the projection matrix.

Let us start by computing $\mathbf{E}\left(||f(\alpha)||^2\right)$ for an arbitrary vector $\alpha \in \mathbb{R}^d$. Let $\{Q_j\}_{j=1}^k$ be defined as

$$Q_j = \alpha \cdot c_j \ .$$

Then

$$\mathbf{E}\left(Q_j\right) = \mathbf{E}\left(\frac{1}{\sqrt{d}} \sum_{i=1}^{d} \alpha_i r_{ij}\right) = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} \alpha_i \mathbf{E}\left(r_{ij}\right) = 0 \ , \tag{1}$$

and

$$\begin{aligned}
\mathbf{E}\left(Q_j^2\right) &= \mathbf{E}\left(\left(\frac{1}{\sqrt{d}} \sum_{i=1}^{d} \alpha_i r_{ij}\right)^2\right) \\
&= \frac{1}{d}\mathbf{E}\left(\sum_{i=1}^{d} \left(\alpha_i r_{ij}\right)^2 + \sum_{l=1}^{d} \sum_{m=1}^{d} 2\alpha_l \alpha_m r_{lj} r_{mj}\right) \\
&= \frac{1}{d} \sum_{i=1}^{d} \alpha_i^2\, \mathbf{E}\left(r_{ij}^2\right) + \frac{1}{d} \sum_{l=1}^{d} \sum_{m=1}^{d} 2\alpha_l \alpha_m \mathbf{E}\left(r_{lj}\right) \mathbf{E}\left(r_{mj}\right) \\
&= \frac{1}{d} \times ||\alpha||^2 \ .
\end{aligned} \tag{2}$$

Note that to get (1) and (2) we only used that $\{r_{ij}\}$ are independent, $\mathbf{E}(r_{ij}) = 0$ and $\mathrm{Var}(r_{ij}) = 1$. Now from (2) we see that

$$\mathbf{E}\left(||f(\alpha)||^2\right) = \mathbf{E}\left(\left(||\sqrt{d/k}\ (\alpha \cdot c_1, \ldots, \alpha \cdot c_d)||\right)^2\right) = \frac{d}{k} \sum_{j=1}^{k} \mathbf{E}\left(Q_j^2\right) = ||\alpha||^2 \ .$$

That is for *any* independent family of $\{r_{ij}\}$ with $\mathbf{E}(r_{ij}) = 0$ and $\mathrm{Var}(r_{ij}) = 1$ we get an independent estimator, i.e., $\mathbf{E}\left(||f(\alpha)||^2\right) = ||\alpha||^2$. Note now that in order to have a JL-embedding we need that for each of the $\binom{n}{2}$ pairs $u, v \in P$, the squared norm of the vector $u - v$, is maintained within a factor of $1 \pm \epsilon$. Therefore, if for some such family of $\{r_{ij}\}$ we can further prove that for some $\beta > 0$ and any fixed vector $\alpha \in \mathbb{R}^d$,

$$\Pr[(1-\epsilon)||\alpha||^2 \leq ||f(\alpha)||^2 \leq (1+\epsilon)||\alpha||^2] \geq 1 - \frac{2}{n^{2+\beta}} \ , \tag{3}$$

then the probability of not getting a JL-embedding is bounded by $\binom{n}{2} \times 2/n^{2+\beta} < 1/n^{\beta}$. Thus, our entire task has been reduced to determining a zero mean, unit variance distribution for the $\{r_{ij}\}$ such that (3) holds for *any* fixed vector $\alpha$. In fact, since for any fixed projection matrix, $||f(\alpha)||^2$ is proportional to $||\alpha||^2$, it suffices to prove that (3) holds for arbitrary *unit*

vectors. Finally, observe that since $\mathbf{E}(||f(\alpha)||^2) = ||\alpha||^2$, inequality (3) merely asserts that the random variable $||f(\alpha)||^2$ is concentrated around its expectation.

### 4.2. The spherically symmetric case

As a warm up for proving concentration for our distributions for the $\{r_{ij}\}$, let us first wrap up the spherically random case. Getting a concentration inequality for $||f(\alpha)||^2$ when $r_{ij} \stackrel{D}{=} N(0,1)$ is straightforward. Due to the 2-stability of the Normal distribution, for *every* unit vector $\alpha$, we have $||f(\alpha)||^2 \stackrel{D}{=} \chi^2(k)/k$, where $\chi^2(k)$ denotes the Chi-square distribution with $k$ degrees of freedom. The fact that we get the same distribution for every vector $\alpha$ corresponds to the intuition that "all vectors are the same" with respect to projection onto a spherically random vector. Standard tail-bounds for the Chi-square distribution readily yield the following.

LEMMA 4.1. *Let $r_{ij} \stackrel{D}{=} N(0,1)$ for all $i, j$. Then, for any $\epsilon > 0$ and any unit-vector $\alpha \in \mathbb{R}^d$,*

$$\Pr\left[||f(\alpha)||^2 > 1+\epsilon\right] \;<\; \exp\left(-\frac{k}{2}(\epsilon^2/2 - \epsilon^3/3)\right) \;\;,$$
$$\Pr\left[||f(\alpha)||^2 < 1-\epsilon\right] \;<\; \exp\left(-\frac{k}{2}(\epsilon^2/2 - \epsilon^3/3)\right) \;\;.$$

Thus, to get a JL-embedding we need only require

$$2 \times \exp\left(-\frac{k}{2}(\epsilon^2/2 - \epsilon^3/3)\right) \leq \frac{2}{n^{2+\beta}} \;\;,$$

which holds for

$$k \geq \frac{4+2\beta}{\epsilon^2/2 - \epsilon^3/3}\;\log n \;\;.$$

Let us note that the bound on the upper tail of $||f(\alpha)||^2$ above is *tight* (up to lower order terms). As a result, as long as the union bound is used, one cannot hope for a better bound on $k$ while using spherically random vectors.

To prove our result we will use the exact same approach, arguing that for every unit vector $\alpha \in \mathbb{R}^d$, the random variable $||f(\alpha)||^2$ is sharply concentrated around its expectation. In the next section we state a lemma analogous to Lemma 4.1 above and show how it follows from bounds on certain moments of $Q_1^2$. We prove those bounds in Section 6.

## 5. TAIL BOUNDS

To simplify notation let us define for an arbitrary vector $\alpha$,

$$S = S(\alpha) = \sum_{j=1}^{k} (\alpha \cdot c_j)^2 = \sum_{j=1}^{k} Q_j^2(\alpha) \;\;,$$

where $c_j$ is the $j^{th}$ column of $R$, so that $||f(\alpha)||^2 = S \times d/k$.

LEMMA 5.1. *Let $r_{ij}$ have any one of the two distributions in Theorem 1.1. Then, for any $\epsilon > 0$ and any unit vector $\alpha \in \mathbb{R}^d$,*

$$\begin{aligned} \Pr\left[S(\alpha) > (1+\epsilon)k/d\right] &< \exp\left(-\frac{k}{2}(\epsilon^2/2 - \epsilon^3/3)\right) \ , \\ \Pr\left[S(\alpha) < (1-\epsilon)k/d\right] &< \exp\left(-\frac{k}{2}(\epsilon^2/2 - \epsilon^3/3)\right) \ . \end{aligned}$$

In proving Lemma 5.1 we will generally omit the dependence of probabilities on $\alpha$, making it explicit only when it affects our calculations. We will use the standard technique of applying Markov's inequality to the moment generating function of $S$, thus reducing the proof of the lemma to bounding certain moments of $Q_1$. In particular, we will need the following lemma which will be proved in Section 6.

LEMMA 5.2. *For all $h \in [0, d/2)$, all $d \geq 1$ and all unit vectors $\alpha$,*

$$\mathbf{E}\left(\exp\left(hQ_1(\alpha)^2\right)\right) \leq \frac{1}{\sqrt{1-2h/d}} \ , \tag{4}$$

$$\mathbf{E}\left(Q_1(\alpha)^4\right) \leq \frac{3}{d^2} \ . \tag{5}$$

**Proof of Lemma 5.1**. We start with the upper tail. For arbitrary $h > 0$ let us write

$$\begin{aligned} \Pr\left[S > (1+\epsilon)\frac{k}{d}\right] &= \Pr\left[\exp(hS) > \exp\left(h(1+\epsilon)\frac{k}{d}\right)\right] \\ &< \mathbf{E}\left(\exp\left(hS\right)\right)\exp\left(-h(1+\epsilon)\frac{k}{d}\right) \ . \end{aligned}$$

Since $\{Q_j\}_{j=1}^k$ are i.i.d. we have

$$\begin{aligned} \mathbf{E}\left(\exp\left(hS\right)\right) &= \mathbf{E}\left(\prod_{j=1}^{k}\exp\left(hQ_j^2\right)\right) &(6)\\ &= \prod_{j=1}^{k}\mathbf{E}\left(\exp\left(hQ_j^2\right)\right) &(7)\\ &= \left(\mathbf{E}\left(\exp\left(hQ_1^2\right)\right)\right)^k \ , &(8) \end{aligned}$$

where passing from (6) to (7) uses that the $\{Q_j\}_{j=1}^k$ are independent, while passing from (7) to (8) uses that they are identically distributed. Thus, for any $\epsilon > 0$

$$\Pr\left[S > (1+\epsilon)\frac{k}{d}\right] < \left(\mathbf{E}\left(\exp\left(hQ_1^2\right)\right)\right)^k \exp\left(-h(1+\epsilon)\frac{k}{d}\right) \ . \tag{9}$$

Substituting (4) in (9) we get (10). To optimize the bound we set the derivative in (10) with respect to $h$ to 0. This gives $h = \frac{d}{2}\frac{\epsilon}{1+\epsilon} < \frac{d}{2}$. Substituting this value of $h$ we get (11)

and series expansion yields (12).

$$\Pr\left[S > (1+\epsilon)\frac{k}{d}\right] < \left(\frac{1}{\sqrt{1-2h/d}}\right)^k \exp\left(-h(1+\epsilon)\frac{k}{d}\right) \quad (10)$$

$$= \left((1+\epsilon)\exp(-\epsilon)\right)^{k/2} \quad (11)$$

$$< \exp\left(-\frac{k}{2}(\epsilon^2/2 - \epsilon^3/3)\right) \ . \quad (12)$$

Similarly, but now considering $\exp(-hS)$ for arbitrary $h > 0$, we get that for any $\epsilon > 0$

$$\Pr\left[S < (1-\epsilon)\frac{k}{d}\right] < \left(\mathbf{E}\left(\exp\left(-hQ_1^2\right)\right)\right)^k \exp\left(h(1-\epsilon)\frac{k}{d}\right) \ . \quad (13)$$

Rather than bounding $\mathbf{E}\left(\exp\left(-hQ_1^2\right)\right)$ directly, let us expand $\exp(-hQ_1^2)$ to get

$$\Pr\left[S < (1-\epsilon)\frac{k}{d}\right] < \left(\mathbf{E}\left(1 - hQ_1^2 + \frac{\left(-hQ_1^2\right)^2}{2!}\right)\right)^k \exp\left(h(1-\epsilon)\frac{k}{d}\right)$$

$$= \left(1 - \frac{h}{d} + \frac{h^2}{2}\mathbf{E}\left(Q_1^4\right)\right)^k \exp\left(h(1-\epsilon)\frac{k}{d}\right) \ , \quad (14)$$

where $\mathbf{E}(Q_1^2)$ was given by (2).

Now, substituting (5) in (14) we get (15). This time taking $h = \frac{d}{2}\frac{\epsilon}{1+\epsilon}$ is not optimal but is still "good enough", giving (16). Again, series expansion yields (17).

$$\Pr\left[S < (1-\epsilon)\frac{k}{d}\right] \leq \left(1 - \frac{h}{d} + \frac{3}{2}\left(\frac{h}{d}\right)^2\right)^k \exp\left(h(1-\epsilon)\frac{k}{d}\right) \quad (15)$$

$$= \left(1 - \frac{\epsilon}{2(1+\epsilon)} + \frac{3\epsilon^2}{8(1+\epsilon)^2}\right)^k \exp\left(\frac{\epsilon(1-\epsilon)k}{2(1+\epsilon)}\right) \quad (16)$$

$$< \exp\left(-\frac{k}{2}(\epsilon^2/2 - \epsilon^3/3)\right) \ . \quad (17)$$

□

## 6. MOMENT BOUNDS

To simplify notation in this section we will drop the subscript and refer to $Q_1$ as $Q$. It should be clear that the distribution of $Q$ depends on $\alpha$, i.e., $Q = Q(\alpha)$. This is precisely what we give up by not projecting onto spherically symmetric vectors. Our strategy for giving bounds on the moments of $Q$ will be to determine a "worst case" unit vector $w$ and bound the moments of $Q(w)$. We claim the following.

LEMMA 6.1. *Let*

$$w = \frac{1}{\sqrt{d}}(1, \ldots, 1) \ .$$

*For every unit vector $\alpha \in \mathbb{R}^d$, and for all $k = 0, 1, \ldots$*

$$\mathbf{E}\left(Q(\alpha)^{2k}\right) \leq \mathbf{E}\left(Q(w)^{2k}\right) \quad . \tag{18}$$

We will also prove that the even moments of $Q(w)$ are dominated by the corresponding moments from the spherically symmetric case. That is,

LEMMA 6.2. *Let*

$$T \stackrel{D}{=} N(0, 1/d) \quad .$$

*For all $d \geq 1$ and all $k = 0, 1, \ldots$*

$$\mathbf{E}\left(Q(w)^{2k}\right) \leq \mathbf{E}\left(T^{2k}\right) \quad . \tag{19}$$

Using Lemmata 6.1 and 6.2 we can prove Lemma 5.2 as follows.

**Proof of Lemma 5.2.** To prove (5) we observe that for any unit vector $\alpha$, by (18) and (19),

$$\mathbf{E}\left(Q(\alpha)^4\right) \leq \mathbf{E}\left(Q(w)^4\right) \leq \mathbf{E}\left(T^4\right) \quad ,$$

while

$$\mathbf{E}\left(T^4\right) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-\lambda^2/2) \left(\frac{\lambda^4}{d^2}\right) d\lambda = \frac{3}{d^2} \quad .$$

To prove (4) we first observe that for any real-valued random variable $U$ and for all $h$ such that $\mathbf{E}\left(\exp\left(hU^2\right)\right)$ is bounded, the Monotone Convergence Theorem (MCT) allows us to swap the expectation with the sum and get

$$\mathbf{E}\left(\exp\left(hU^2\right)\right) = \mathbf{E}\left(\sum_{k=0}^{\infty} \frac{(hU^2)^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{h^k}{k!} \mathbf{E}\left(U^{2k}\right) \quad .$$

So, below, we proceed as follows. Taking $h \in [0, d/2)$ makes the integral in (20) converge, giving us (21). Thus, for such $h$, we can apply the MCT to get (22). Now, applying (18) and (19) to (22) gives (23). Applying the MCT once more gives (24).

$$\begin{aligned}
\mathbf{E}\left(\exp\left(hT^2\right)\right) &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} \exp(-\lambda^2/2) \exp\left(h\frac{\lambda^2}{d}\right) d\lambda && (20)\\
&= \frac{1}{\sqrt{1 - 2h/d}} && (21)\\
&= \sum_{k=0}^{\infty} \frac{h^k}{k!} \mathbf{E}\left(T^{2k}\right) && (22)\\
&\geq \sum_{k=0}^{\infty} \frac{h^k}{k!} \mathbf{E}\left(Q(\alpha)^{2k}\right) && (23)\\
&= \mathbf{E}\left(\exp\left(hQ(\alpha)^2\right)\right) \quad . && (24)
\end{aligned}$$

Thus, $\mathbf{E}\left(\exp\left(hQ^2\right)\right) \leq 1/\sqrt{1 - 2h/d}$ for $h \in [0, d/2)$, as desired. □

Before proving Lemma 6.1 we will need to prove the following lemma.

LEMMA 6.3. *Let $r_1, r_2$ be i.i.d. r.v. having one of the following two probability distributions: $r_i \in \{-1, +1\}$, each value having probability 1/2, or, $r_i \in \{-\sqrt{3}, 0, +\sqrt{3}\}$ with 0 having probability 2/3 and $\pm\sqrt{3}$ being equiprobable.*

*For any $a, b \in \mathbb{R}$ let $c = \sqrt{(a^2+b^2)/2}$. Then for any $M \in \mathbb{R}$ and all $k = 0, 1, \ldots$*

$$\mathbf{E}\left((M + ar_1 + br_2)^{2k}\right) \leq \mathbf{E}\left((M + cr_1 + cr_2)^{2k}\right) \ .$$

**Proof**. We first consider the case where $r_i \in \{-1, +1\}$, each value having probability $1/2$.

If $a^2 = b^2$ then $a = c$ and the lemma holds with equality. Otherwise, observe that

$$\mathbf{E}\left((M + cr_1 + cr_2)^{2k}\right) - \mathbf{E}\left((M + ar_1 + br_2)^{2k}\right) = \frac{S_k}{4}$$

where

$$\begin{aligned} S_k \ &= \ (M+2c)^{2k} + 2M^{2k} + (M-2c)^{2k} - (M+a+b)^{2k} \\ &\quad -(M+a-b)^{2k} - (M-a+b)^{2k} - (M-a-b)^{2k} \ . \end{aligned}$$

We will show that $S_k \geq 0$ for all $k \geq 0$.

Since $a^2 \neq b^2$ we can use the binomial theorem to expand every term other than $2M^{2k}$ in $S_k$ and get

$$S_k = 2M^{2k} + \sum_{i=0}^{2k} \binom{2k}{i} M^{2k-i} D_i \ ,$$

where

$$D_i = (2c)^i + (-2c)^i - (a+b)^i - (a-b)^i - (-a+b)^i - (-a-b)^i \ .$$

Observe now that for odd $i$, $D_i = 0$. Moreover, we claim that $D_{2j} \geq 0$ for all $j \geq 1$. To see this claim observe that $(2a^2 + 2b^2) = (a+b)^2 + (a-b)^2$ and that for all $j \geq 1$ and $x, y \geq 0$, $(x+y)^j \geq x^j + y^j$. Thus,

$$S_k \ = \ 2M^{2k} + \sum_{j=0}^{k} \binom{2k}{2j} M^{2(k-j)} D_{2j} \ = \ \sum_{j=1}^{k} \binom{2k}{2j} M^{2(k-j)} D_{2j} \ \geq \ 0 \ .$$

The proof for the case where $r_i \in \{-\sqrt{3}, 0, +\sqrt{3}\}$ is just a more cumbersome version of the proof above, so we omit it. That proof, though, brings forward an interesting point. If one tries to take $r_i = 0$ with probability greater than $2/3$, while maintaining a range of size 3 and variance 1, the lemma fails. In other words, 2/3 is tight in terms of how much probability mass we can put to $r_i = 0$ and still have the current lemma hold. □

**Proof of Lemma 6.1**. Recall that for any vector $\alpha$, $Q(\alpha) = Q_1(\alpha) = \alpha \cdot c_1$ where

$$c_1 = \frac{1}{\sqrt{d}} \left(r_{11}, \ldots, r_{d1}\right) \ .$$

If $\alpha = (\alpha_1, \ldots, \alpha_d)$ is such that $\alpha_i^2 = \alpha_j^2$ for all $i, j$, then by symmetry, $Q(\alpha)$ and $Q(w)$ are identically distributed and the lemma holds trivially. Otherwise, we can assume

without loss of generality, that $\alpha_1^2 \neq \alpha_2^2$ and consider the "more balanced" unit vector $\theta = (c, c, \alpha_3, \ldots, \alpha_d)$, where $c = \sqrt{(\alpha_1^2 + \alpha_2^2)/2}$. We will prove that

$$\mathbf{E}\left(Q(\alpha)^{2k}\right) \leq \mathbf{E}\left(Q(\theta)^{2k}\right) \quad . \tag{25}$$

Applying this argument repeatedly yields the lemma, as $\theta$ eventually becomes $w$.

To prove (25), below we first express $\mathbf{E}\left(Q(\alpha)^{2k}\right)$ as a sum of averages over $r_{11}, r_{21}$ and then apply Lemma 6.3 to get that each term (average) in the sum, is bounded by the corresponding average for vector $\theta$. More precisely,

$$\begin{aligned}
\mathbf{E}\left(Q(\alpha)^{2k}\right) &= \frac{1}{d^k} \sum_M \mathbf{E}\left((M + \alpha_1 r_{11} + \alpha_2 r_{21})^{2k}\right) \Pr\left[\sum_{i=3}^{d} \alpha_i r_{i1} = \frac{M}{\sqrt{d}}\right] \\
&\leq \frac{1}{d^k} \sum_M \mathbf{E}\left((M + c r_{11} + c r_{21})^{2k}\right) \Pr\left[\sum_{i=3}^{d} \alpha_i r_{i1} = \frac{M}{\sqrt{d}}\right] \\
&= \mathbf{E}\left(Q(\theta)^{2k}\right) \quad .
\end{aligned}$$

□

**Proof of Lemma 6.2**. Recall that $T \stackrel{D}{=} N(0, 1/d)$. We will first express $T$ as the scaled sum of $d$ independent standard Normal random variables. This will allow for a direct comparison of the terms in each of the two expectations.

Specifically, let $\{T_i\}_{i=1}^d$ be a family of i.i.d. standard Normal random variables. Then $\sum_{i=1}^d T_i$ is a Normal random variable with variance $d$. Therefore,

$$T \stackrel{D}{=} \frac{1}{d} \sum_{i=1}^{d} T_i \quad .$$

Recall also that $Q(w) = Q_1(w) = w \cdot c_1$ where

$$c_1 = \frac{1}{\sqrt{d}} \left(r_{11}, \ldots, r_{d1}\right) \quad .$$

To simplify notation let us write $r_{i1} = Y_i$ and let us also drop the dependence of $Q$ on $w$. Thus,

$$Q = \frac{1}{d} \sum_{i=1}^{d} Y_i \quad ,$$

where $\{Y_i\}_{i=1}^d$ are i.i.d. r.v. having one of the following two distributions: $Y_i \in \{-1, +1\}$, each value having probability $1/2$, or $Y_i \in \{-\sqrt{3}, 0, +\sqrt{3}\}$ with 0 having probability 2/3 and $\pm\sqrt{3}$ being equiprobable.

We are now ready to compare $\mathbf{E}\left(Q^{2k}\right)$ with $\mathbf{E}\left(T^{2k}\right)$. We first observe that for every $k = 0, 1, \ldots$

$$\begin{aligned}
\mathbf{E}\left(T^{2k}\right) &= \frac{1}{d^{2k}} \sum_{i_1=1}^{d} \cdots \sum_{i_{2k}=1}^{d} \mathbf{E}\left(T_{i_1} \cdots T_{i_{2k}}\right) \text{ , and} \\
\mathbf{E}\left(Q^{2k}\right) &= \frac{1}{d^{2k}} \sum_{i_1=1}^{d} \cdots \sum_{i_{2k}=1}^{d} \mathbf{E}\left(Y_{i_1} \cdots Y_{i_{2k}}\right) \quad .
\end{aligned}$$

To prove the lemma we will show that for every value assignment to the indices $i_1, \ldots, i_{2k}$,

$$\mathbf{E}\left(Y_{i_1} \cdots Y_{i_{2k}}\right) \leq \mathbf{E}\left(T_{i_1} \cdots T_{i_{2k}}\right) \quad . \tag{26}$$

Let $V = \langle v_1, v_2, \ldots, v_{2k} \rangle$ be the value assignment considered. For $i \in \{1, \ldots, d\}$, let $c_V(i)$ be the number of times that $i$ appears in $V$. Observe that if for some $i$, $c_V(i)$ is odd then both expectations appearing in (26) are 0, since both $\{Y_i\}_{i=1}^d$ and $\{T_i\}_{i=1}^d$ are independent families and $\mathbf{E}(Y_i) = \mathbf{E}(T_i) = 0$ for all $i$. Thus, we can assume that there exists a set $\{j_1, j_2, \ldots, j_p\}$ of indices and corresponding values $\ell_1, \ell_2, \ldots, \ell_p$ such that

$$\begin{aligned} \mathbf{E}\left(Y_{i_1} \cdots Y_{i_{2k}}\right) &= \mathbf{E}\left(Y_{j_1}^{2\ell_1} Y_{j_2}^{2\ell_2} \cdots Y_{j_p}^{2\ell_p}\right), \quad \text{and} \\ \mathbf{E}\left(T_{i_1} \cdots T_{i_{2k}}\right) &= \mathbf{E}\left(T_{j_1}^{2\ell_1} T_{j_2}^{2\ell_2} \cdots T_{j_p}^{2\ell_p}\right) \quad . \end{aligned}$$

Note now that since the indices $j_1, j_2, \ldots, j_p$ are distinct, $\{Y_{j_t}\}_{t=1}^p$ and $\{T_{j_t}\}_{t=1}^p$ are families of i.i.d. r.v. Therefore,

$$\begin{aligned} \mathbf{E}\left(Y_{i_1} \cdots Y_{i_{2k}}\right) &= \mathbf{E}\left(Y_{j_1}^{2\ell_1}\right) \times \cdots \times \mathbf{E}\left(Y_{j_p}^{2\ell_p}\right), \quad \text{and} \\ \mathbf{E}\left(T_{i_1} \cdots T_{i_{2k}}\right) &= \mathbf{E}\left(T_{j_1}^{2\ell_1}\right) \times \cdots \times \mathbf{E}\left(T_{j_p}^{2\ell_p}\right) \quad . \end{aligned}$$

So, without loss of generality, in order to prove (26) it suffices to prove that for every $\ell = 0, 1, \ldots$

$$\mathbf{E}\left(Y_1^{2\ell}\right) \leq \mathbf{E}\left(T_1^{2\ell}\right) \quad . \tag{27}$$

This, though, is completely trivial. First recall the well-known fact that the $(2\ell)$th moment of $N(0,1)$ is $(2\ell-1)!! = (2\ell)!/(\ell! 2^\ell) \geq 1$. Now:

– If $Y_1 \in \{-1, +1\}$ then $\mathbf{E}\left(Y_1^{2\ell}\right) = 1$, for all $\ell \geq 0$.

– If $Y_1 \in \{-\sqrt{3}, 0, +\sqrt{3}\}$ then $\mathbf{E}(Y_1^{2\ell}) = 3^{\ell-1} \leq (2\ell)!/(\ell! 2^\ell)$, where the last inequality follows by an easy induction.

It is worth pointing out that, along with Lemma 6.3, these are the only two points were we used any properties of the distributions for the $r_{ij}$ (here called $Y_i$) other than them having zero mean and unit variance. □

Finally, we note that

• Since $\mathbf{E}\left(Y_1^{2\ell}\right) < \mathbf{E}\left(T_1^{2\ell}\right)$ for certain $l$, we see that for each fixed $d$, both inequalities in Lemma 5.2 are actually strict, yielding slightly better tails bounds for $S$ and a correspondingly better bound for $k_0$.

• By using Jensen's inequality one can get a direct bound for $\mathbf{E}(Q^{2k})$ when $Y_i \in \{-1, +1\}$, i.e., without comparing it to $\mathbf{E}(T^{2k})$. That simplifies the proof for that case and shows that, in fact, taking $Y_i \in \{-1, +1\}$ is the minimizer of $\mathbf{E}\left(\exp\left(hQ^2\right)\right)$ for all $h$.

## ACKNOWLEDGMENT

I am grateful to Marek Biskup for his help with the proof of Lemma 6.2 and to Jeong Han Kim for suggesting the approach of equation (14). Many thanks also to Paul Bradley, Aris Gionis, Anna Karlin, Elias Koutsoupias and Piotr Indyk for comments on earlier versions of the paper and useful discussions.

## REFERENCES

1. Sanjeev Arora and Ravi Kannan, *Learning mixtures of arbitrary Gaussians*, 33rd Annual ACM Symposium on Theory of Computing (Creete, Greece), ACM, New York, 2001, pp. 247–257.
2. Sanjoy Dasgupta, *Learning mixtures of Gaussians*, 40th Annual Symposium on Foundations of Computer Science (New York, NY, 1999), IEEE Comput. Soc. Press, Los Alamitos, CA, 1999, pp. 634–644.
3. Sanjoy Dasgupta and Anupam Gupta, *An elementary proof of the Johnson-Lindenstrauss lemma*, Technical report 99-006, UC Berkeley, March 1999.
4. Peter Frankl and Hiroshi Maehara, *The Johnson-Lindenstrauss lemma and the sphericity of some graphs*, J. Combin. Theory Ser. B **44** (1988), no. 3, 355–362.
5. Piotr Indyk, *Stable distributions, pseudorandom generators, embeddings and data stream computation*, 41st Annual Symposium on Foundations of Computer Science (Redondo Beach, CA, 2000), IEEE Comput. Soc. Press, Los Alamitos, CA, 2000, pp. 189–197.
6. Piotr Indyk and Rajeev Motwani, *Approximate nearest neighbors: towards removing the curse of dimensionality*, 30th Annual ACM Symposium on Theory of Computing (Dallas, TX), ACM, New York, 1998, pp. 604–613.
7. William B. Johnson and Joram Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conference in modern analysis and probability (New Haven, Conn., 1982), Amer. Math. Soc., Providence, R.I., 1984, pp. 189–206.
8. Jon Kleinberg, *Two algorithms for nearest-neighbor search in high dimensions*, 29th Annual ACM Symposium on Theory of Computing (El Paso, TX, 1997), ACM, New York, 1997, pp. 599–608.
9. Nathan Linial, Eran London, and Yuri Rabinovich, *The geometry of graphs and some of its algorithmic applications*, Combinatorica **15** (1995), no. 2, 215–245.
10. Christos H. Papadimitriou, Prabhakar Raghavan, Hisao Tamaki, and Santosh Vempala, *Latent semantic indexing: A probabilistic analysis*, 17th Annual Symposium on Principles of Database Systems (Seattle, WA, 1998), 1998, pp. 159–168.
11. Leonard J. Schulman, *Clustering for edge-cost minimization*, 32nd Annual ACM Symposium on Theory of Computing (Portland, OR, 2000), ACM, New York, 2000, pp. 547–555.
12. Santosh Vempala, *A random sampling based algorithm for learning the intersection of half-spaces*, 38th Annual Symposium on Foundations of Computer Science (Miami, FL, 1997), IEEE Comput. Soc. Press, Los Alamitos, CA, 1997, pp. 508–513.