Dimension Reduction by Random Hyperplane Tessellations

Yaniv Plan · Roman Vershynin

Abstract Given a subset K of the unit Euclidean sphere, we estimate the minimal number $m = m(K)$ of hyperplanes that generate a uniform tessellation of K, in the sense that the fraction of the hyperplanes separating any pair $x, y \in K$ is nearly proportional to the Euclidean distance between x and y. Random hyperplanes prove to be almost ideal for this problem; they achieve the almost optimal bound $m = O(w(K)^2)$ where $w(K)$ is the Gaussian mean width of K. Using the map that sends $x \in K$ to the sign vector with respect to the hyperplanes, we conclude that every bounded subset K of $\mathbb{R}^n$ embeds into the Hamming cube $\{-1, 1\}^m$ with a small distortion in the Gromov–Haussdorff metric. Since for many sets K one has $m = m(K) \ll n$, this yields a new discrete mechanism of dimension reduction for sets in Euclidean spaces.

Keywords Embedding · Dimension reduction · Hyperplane tessellations · Mean width · Near isometry

1 Introduction

Consider a bounded subset K of $\mathbb{R}^n$. We would like to find an arrangement of m affine hyperplanes in $\mathbb{R}^n$ that cut through K as evenly as possible; see Fig. 1 for an illustration. The intuitive notion of an "even cut" can be expressed more formally in the following way: The fraction of the hyperplanes separating any pair $x, y \in K$ should be proportional (up to a small additive error) to the Euclidean distance between x and y. What is the smallest possible number $m = m(K)$ of hyperplanes with this property? Besides having a natural theoretical appeal, this question is directly motivated by a certain problem of information theory, which we will describe later.

Y. Plan · R. Vershynin
Department of Mathematics, University of Michigan, 530 Church St., Ann Arbor, MI 48109, USA
e-mail: yplan@umich.edu

R. Vershynin
e-mail: romanv@umich.edu

Fig. 1 A hyperplane
tessellation of a set in the plane
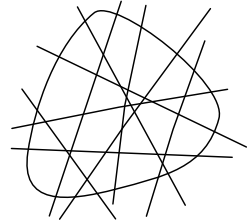
In the beginning, it will be most convenient to work with subsets K of the unit Euclidean sphere S^{n-1}, but we will lift this restriction later. Let $d(x, y)$ denote the normalized geodesic distance on S^{n-1}, so the distance between the opposite points on the sphere equals 1. A (linear) hyperplane in $\mathbb{R}^n$ can be expressed as $a^\perp$ for some $a \in \mathbb{R}^n$. We say that points $x, y \in \mathbb{R}^n$ are separated by the hyperplane[1] if $\text{sign}\langle a, x \rangle \neq \text{sign}\langle a, y \rangle$.

Definition 1.1 (Uniform tessellation) Consider a subset $K \subseteq S^{n-1}$ and an arrangement of m hyperplanes in $\mathbb{R}^n$. Let $d_A(x, y)$ denote the fraction of the hyperplanes that separate points x and y in $\mathbb{R}^n$. Given $\delta > 0$, we say that the hyperplanes provide a δ-*uniform tessellation* of K if

$$\left| d_A(x, y) - d(x, y) \right| \leq \delta, \quad x, y \in K. \tag{1.1}$$

The main result of this paper is a bound on the minimal number $m = m(K, \delta)$ of hyperplanes that provide a uniform tessellation of a set K. It turns out that for a fixed accuracy δ, an almost optimal estimate on m depends only on one global parameter of K, namely the mean width. Recall that the *Gaussian mean width* of K is defined as

$$w(K) = \mathbb{E} \sup_{x \in K} \left| \langle g, x \rangle \right|, \tag{1.2}$$

where $g \sim \mathcal{N}(0, I_n)$ is a standard Gaussian random vector in $\mathbb{R}^n$.

Theorem 1.2 (Random uniform tessellations) *Consider a subset $K \subseteq S^{n-1}$ and let $\delta > 0$. Let*

$$m \geq C\delta^{-6}w(K)^2$$

and consider an arrangement of m independent random hyperplanes in $\mathbb{R}^n$ uniformly distributed according to the Haar measure. Then with probability at least $1 - 2\exp(-c\delta^2 m)$, these hyperplanes provide a δ-uniform tessellation of K. Here and later, C, c denote positive absolute constants.

Remark 1.3 (Tessellations in stochastic geometry) By the rotation invariance of the Haar measure, it easily follows that $\mathbb{E}d_A(x, y) = d(x, y)$ for each pair $x, y \in \mathbb{R}^n$.

---

[1] For convenience of presentation, we prefer the sign function to take values $\{-1, 1\}$, so we define it as $\text{sign}(t) = 1$ for $t \geq 0$ and $\text{sign}(t) = -1$ for $t < 0$.

Theorem 1.2 states that with high probability, $d_A(x, y)$ almost matches its expected value *uniformly* over all $x, y \in K$. This observation highlights the principal difference between the problems studied in this paper and the classical problems on random hyperplane tessellations studied in stochastic geometry. The classical problems concern the shape of a *specific* cell (usually the one containing the origin) or certain *statistics* of cells (e.g., "how many cells have volume greater than a fixed number"?); see [7]. In contrast to this, the concept of uniform tessellation we propose his paper concerns *all* cells simultaneously; see Sect. 1.5 for a vivid illustration.

1.1 Embeddings into the Hamming Cube

Theorem 1.2 has an equivalent formulation in the context of metric embeddings. It yields that *every subset $K \subseteq S^{n-1}$ can be almost isometrically embedded into the Hamming cube $\{-1, 1\}^m$ with $m = O(w(K)^2)$.*

To explain this statement, let us recall a few standard notions. An ε-isometry (or almost isometry) between metric spaces (X, d_X) and (Y, d_Y) is a map $f : X \to Y$, which satisfies

$$\left| d_Y\big(f(x), f(x')\big) - d_X\big(x, x'\big) \right| \le \varepsilon, \quad x, x' \in X,$$

and such that for every $y \in Y$ one can find $x \in X$ satisfying $d_Y(y, f(x)) \le \varepsilon$. A map $f : X \to Y$ is an ε-isometric embedding of X into Y if the map $f : X \to f(X)$ is an ε-isometry between (X, d_X) and the subspace $(f(X), d_Y)$. It is not hard to show that X can be 2ε-isometrically embedded into Y (by means of a suitable map f) if X has the Gromov–Haussdorff distance at most ε from some subset of Y. Conversely, if there is an ε-isometry between X and $f(X)$ then the Gromov–Haussdorff distance between X and $f(X)$ is bounded by ε.

Finally, recall that the Hamming cube is the set $\{-1, 1\}^m$ with the (normalized) Hamming distance $d_H(u, v) = \frac{1}{m} \sum_{i=1}^m \mathbb{1}_{\{u_i \ne v_i\}} =$ the fraction of the coordinates where u and v are different.

An arrangement of m hyperplanes in $\mathbb{R}^n$ defines a *sign map* $f : \mathbb{R}^n \to \{-1, 1\}^m$, which sends $x \in \mathbb{R}^n$ to the sign vector of the orientations of x with respect to the hyperplanes. The sign map is uniquely defined up to the isometries of the Hamming cube. Let $a_1, \dots, a_m \in \mathbb{R}^n$ be normals of the hyperplanes, and consider the $m \times n$ matrix A with rows a_i. The sign map can be expressed as

$$f(x) = \operatorname{sign} Ax, \quad f : \mathbb{R}^n \to \{-1, 1\}^m,$$

where $\operatorname{sign} Ax$ denotes the vector of signs of the coordinates $\langle a_i, x \rangle$ of Ax. The fraction $d_A(x, y)$ of the hyperplanes that separate points x and y thus equals

$$d_A(x, y) = d_H(\operatorname{sign} Ax, \operatorname{sign} Ay), \quad x, y \in \mathbb{R}^n.$$

Then looking back at the definition of uniform tessellations, we observe the following fact:

Fact 1.4 (Embeddings by uniform tessellations) *Consider a δ-uniform tessellation of a set $K \subseteq S^{n-1}$ by m hyperplanes. Then the set K (with the induced geodesic*

Fig. 2 The graph of a tessellation of a set in the plane. The *dashed lines* represent the edges
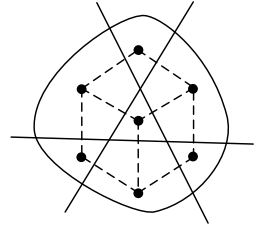
distance*) can be* δ*-isometrically embedded into the Hamming cube* $\{-1, 1\}^m$*. The sign map provides such an embedding.*

This allows us to state Theorem 1.2 as follows.

Theorem 1.5 (Embeddings into the Hamming cube) *Consider a subset* $K \subseteq S^{n-1}$ *and let* δ > 0*. Let*

$$m \geq C\delta^{-6}w(K)^2.$$

Then K *can be* δ*-isometrically embedded into the Hamming cube* $\{-1, 1\}^m$*.*
    Moreover, let A *be an* $m \times n$ *random matrix with independent* $\mathcal{N}(0, 1)$ *entries. Then with probability at least* $1 - 2\exp(-c\delta^2 m)$*, the sign map*

$$f(x) = \text{sign}\, Ax, \quad f : K \to \{-1, 1\}^m \tag{1.3}$$

is an δ*-isometric embedding.*

1.2 Almost Isometry of K and the Tessellation Graph

The image of the sign map f in (1.3) has a special meaning. When the Hamming cube $\{-1, 1\}^m$ is viewed as a graph (in which two points u, v are connected if they differ in exactly one coordinate), the image of f defines a subgraph of $\{-1, 1\}^m$, which is called the *tessellation graph* of K. The tessellation graph has a vertex for each cell and an edge for each pair of adjacent cells; see Fig. 2. Notice that the graph distance in the tessellation graph equals the number of hyperplanes that separate the two cells. Therefore, the definition of a uniform tessellation yields the following.

Fact 1.6 (Graphs of uniform tessellations) *Consider a* δ*-uniform tessellation of a set* $K \subseteq S^{n-1}$*. Then* K *is* δ*-isometric to the tessellation graph of* K*.*

Hence, we can read the conclusion of Theorem 1.2 as follows: K *is* δ*-isometric to the graph of its tessellation by* m *random hyperplanes, where* $m \sim \delta^{-6}w(K)^2$.

1.3 Computing Mean Width

Powerful methods to estimate the mean width $w(K)$ have been developed in connection with stochastic processes. These methods include Sudakov's and Dudley's

inequalities, which relate $w(K)$ to the covering numbers of K in the Euclidean metric, and the sharp technique of majorizing measures (see [15, 22]).

Mean width has a simple (and known) geometric interpretation. By the rotational invariance of the Gaussian random vector g in (1.2), one can replace g with a random vector θ that is uniformly distributed on S^{n-1}, as follows:

$$w(K) = c_n \sqrt{n} \cdot \bar{w}(K), \quad \text{where} \quad \bar{w}(K) = \mathbb{E} \sup_{x \in K} |\langle \theta, x \rangle|.$$

Here, c_n are numbers that depend only on n and such that $c_n \le 1$ and $\lim_{n \to \infty} c_n = 1$. We may refer to $\bar{w}(K)$ as the *spherical mean width* of K. Let us assume for simplicity that K is symmetric with respect to the origin. Then $2 \sup_{x \in K} |\langle \theta, x \rangle|$ is the width of K in the direction θ, which is the distance between the two supporting hyperplanes of K whose normals are θ. The spherical mean width $\bar{w}(K)$ is then twice the average width of K over all directions.

1.4 Dimension Reduction

Our results are already non-trivial in the particular case $K = S^{n-1}$. Since $w(S^{n-1}) \le \sqrt{n}$, Theorems 1.2 and 1.5 hold with $m \sim n$. But more importantly, many interesting sets $K \subset S^{n-1}$ satisfy $w(K) \ll \sqrt{n}$ and, therefore, make our results hold with $m \sim w(K)^2 \ll n$. In such cases, one can view the sign map $f(x) = \text{sign} \, Ax$ in Theorem 1.5 as a dimension reduction mechanism that transforms an n-dimensional set K into a subset of $\{-1, 1\}^m$.

A heuristic reason why dimension reduction is possible is that the quantity $w(K)^2$ measures the *effective dimension* of a set $K \subseteq S^{n-1}$. The effective dimension $w(K)^2$ of a set $K \subseteq S^{n-1}$ is always bounded by the algebraic dimension, but it may be much smaller and it is robust with respect to perturbations of K. In this regard, the notion of effective dimension is parallel to the notion of effective rank of a matrix from numerical linear algebra (see e.g. [20]). With these observations in mind, it is not surprising that the "true," effective dimension of K would be revealed (and would be the only obstruction according to Theorem 1.5) when K is being squeezed into a space of smaller dimension.

Let us illustrate dimension reduction on the example of finite sets $K \subset S^{n-1}$. Since $w(K) \le C \sqrt{\log |K|}$ (see, e.g., [15, (3.13)]), Theorem 1.5 holds with $m \sim \log |K|$, and we can state it as follows.

Corollary 1.7 (Dimension reduction for finite sets) *Let $K \subset S^{n-1}$ be a finite set. Let $\delta > 0$ and $m \ge C \delta^{-6} \log |K|$. Then K can be δ-isometrically embedded into the Hamming cube $\{-1, 1\}^m$.*

This fact should be compared to the *Johnson–Lindenstrauss lemma* for finite subsets $K \subset \mathbb{R}^n$ ([12], see [17, Sect. 15.2]), which states that if $m \ge C \delta^{-2} \log |K|$ then K can be Lipschitz embedded into $\mathbb{R}^m$ as follows:

$$\left| \left\| \bar{A}x - \bar{A}x' \right\|_2 - \left\| x - x' \right\|_2 \right| \le \delta \left\| x - x' \right\|_2, \quad x, x' \in K.$$

Here, $\bar{A} = m^{-1/2}A$ is the rescaled random Gaussian matrix A from Theorem 1.5. Note that while the Johnson–Lindenstrauss lemma involves a Lipschitz embedding from $\mathbb{R}^n$ to $\mathbb{R}^m$, it is generally impossible to provide a Lipschitz embedding from subsets of $\mathbb{R}^n$ to the Hamming cube (if there are points $x, x' \in K$ that are very close to each other); this is why we consider δ-isometric embeddings.

Like the Johnson–Lindenstrauss lemma, Corollary 1.7 can be proved directly by combining concentration inequalities for $d_A(x, y)$ with a union bound over $|K|^2$ pairs $(x, y) \in K \times K$. In fact, this method of proof allows for the weaker requirement $m \geq C\delta^{-2} \log |K|$. However, as we discuss later, this argument cannot be generalized in a straightforward way to prove Theorem 1.5 for general sets K. The Hamming distance $d_A(x, y)$ is highly discontinuous, which makes it difficult to extend estimates from points x, y in an ε-net of K to nearby points.

1.5 Cells of Uniform Tessellations

We mentioned two nice features of uniform tessellations in Facts 1.4 and 1.6. Let us observe one more property: all cells of a uniform tessellation have small diameter. Indeed, $d_A(x, y) = 0$ iff points x, y are in the same cell, so by (1.1) we have the following.

Fact 1.8 (Cells are small) *Every cell of a δ-uniform tessellation has diameter at most δ.*

With this, Theorem 1.2 immediately implies the following.

Corollary 1.9 (Cells of random uniform tessellations) *Consider a tessellation of a subset $K \subseteq S^{n-1}$ by $m \geq C\delta^{-6}w(K)^2$ random hyperplanes. Then, with probability at least $1 - \exp(-c\delta^2 m)$, all cells of the tessellation have diameter at most δ.*

This result has also a direct proof, which moreover gives a slightly better bound $m \sim \delta^{-4}w(K)^2$. We present this "curvature argument" in Sect. 3.

1.6 Uniform Tessellations in $\mathbb{R}^n$

So far, we only worked with subsets $K \subseteq S^{n-1}$. It is not difficult to extend our results to bounded sets $K \subset \mathbb{R}^n$. This can be done by embedding such a set K into S^n (the sphere in one more dimension) with small bi-Lipschitz distortion. This elementary argument is presented in Sect. 6, and it yields the following version of Theorem 1.2.

Theorem 1.10 (Random uniform tessellations in $\mathbb{R}^n$) *Consider a bounded subset $K \subset \mathbb{R}^n$ with $\mathrm{diam}(K) = 1$. Let*

$$m \geq C\delta^{-12}w(K - K)^2. \tag{1.4}$$

Then there exists an arrangement of m affine hyperplanes in $\mathbb{R}^n$ and a scaling factor $\lambda > 0$ such that

$$\left| \lambda \cdot d_A(x, y) - \|x - y\|_2 \right| \leq \delta, \quad x, y \in K.$$

Here, $d_A(x, y)$ denotes the fraction of the affine hyperplanes that separate x and y.

Remark 1.11 (Mean width in $\mathbb{R}^n$) While the quantity $w(K - K)$ appearing in (1.4) is clearly bounded by $2w(K)$, it is worth noting that the quantity $w(K - K)$ captures more accurately than $w(K)$ the geometric nature of the "mean width" of K. Indeed, $w(K - K) = \mathbb{E}h(g)$ where $h(g) = \sup_{x \in K} \langle g, x \rangle - \inf_{x \in K} \langle g, x \rangle$ is the distance between the two parallel supporting hyperplanes of K orthogonal to the random direction g, scaled by $\|g\|_2$.

1.7 Optimality

The main object of our study is $m(K) = m(K, \delta)$, the smallest number of hyperplanes that provide a δ-uniform tessellation of a set $K \subseteq S^{n-1}$. One has

$$\log_2 N(K, \delta) \leq m(K, \delta) \leq C\delta^{-6}w(K)^2, \tag{1.5}$$

where $N(K, \delta)$ denotes the covering number of K, i.e. the smallest number of balls of radius δ that cover K. The upper bound in (1.5) is the conclusion of Theorem 1.2. The lower bound holds because a δ-uniform tessellation provides a decomposition of K into at most 2^m cells each of which lies in a ball of radius δ by Fact 1.8.

To compare the upper and lower bounds in (1.5), recall Sudakov's inequality [15, Theorem 3.18] that yields

$$\log N(K, \delta) \leq C\delta^{-2}w(K)^2.$$

While Sudakov's inequality cannot be reversed in general, there are many situations where it is sharp. Moreover, according to Dudley's inequality (see [15, Theorem 11.17] and [18, Lemma 2.33]), Sudakov's inequality can always be reversed for some scale $\delta > 0$ and up to a logarithmic factor in n. (See also [16] for a discussion of sharpness of Sudakov's inequality.) So, the two sides of (1.5) are often close to each other, but there is in general some gap. We conjecture that the optimal estimate is

$$cw(K)^2 \leq \sup_{\delta > 0} \delta^2 m(K, \delta) \leq Cw(K)^2,$$

so the mean width of K seems to be completely responsible for the uniform tessellations of K.

Note that the lower bound in (1.5) holds in greater generality. Namely, it is not possible to have $m < \log_2 N(K, \delta)$ for *any* decomposition of K into 2^m pieces of diameter at most δ. However, from the upper bound, we see that with a slightly larger value $m \sim w(K)^2$, *an almost best decomposition of K is achieved by a random hyperplane tessellation.*

In this paper, we have not tried to optimize the dependence of $m(K, \delta)$ on δ. This interesting problem is related to the open question on the optimal dependence on distortion in Dvoretzky's theorem. We comment on this in Sect. 3.2.

1.8 Related Work: Embeddings of K into Normed Spaces

Embeddings of subsets $K \subseteq S^{n-1}$ into normed spaces were studied in geometric functional analysis [13, 21]. In particular, Klartag and Mendelson [13] were concerned with embeddings into ℓ_2^m. They showed that for $m \geq C\delta^{-2}w(K)^2$ there exists a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\left| m^{-1/2} \|Ax\|_2 - 1 \right| \leq \delta, \quad x \in K.$$

One can choose A to be an $m \times n$ random matrix with Gaussian entries as in Theorem 1.5, or with sub-Gaussian entries. Schechtman [21] gave a simpler argument for a Gaussian matrix, which also works for embeddings into general normed spaces X. In the specific case of $X = \ell_1^m$, Schechtman's result states that for $m \geq C\delta^{-2}w(K)^2$ one has

$$\left| m^{-1} \|Ax\|_1 - 1 \right| \leq \delta, \quad x \in K.$$

This result also follows from Lemma 2.1 below.

1.9 Related Work: One-Bit Compressed Sensing

Our present work was motivated by the development of *one-bit compressed sensing* in [6, 11, 19] where Theorem 1.5 is used in the following context. The vector x represents a signal; the matrix A represents a measurement map $\mathbb{R}^n \to \mathbb{R}^m$ that produces $m \ll n$ linear measurements of x; taking the sign of Ax represents quantization of the measurements (an extremely coarse, one-bit quantization). The problem of one-bit compressed sensing is to recover the signal x from the quantized measurements $f(x) = \text{sign } Ax$.

The problem of one-bit compressed sensing was introduced by Boufounos and Baraniuk [6]. Jacques, Laska, Boufounos, and Baraniuk [11] realized a connection of this problem to uniform tessellations of the set of sparse signals $K = \{x \in S^{n-1} : |\text{supp}(x)| \leq s\}$, and to almost isometric embedding of K into the Hamming cube $\{-1, 1\}^m$. For this set K, they proved Corollary 1.9 with $m \sim \delta^{-1} s \log(n/\delta)$ and a version of Theorem 1.5 for $m \sim \delta^{-2} s \log(n/\delta)$. The authors of the present paper analyzed in [19] a bigger set of "compressible" signals $K' = \{x \in S^{n-1} : \|x\|_1 \leq \sqrt{s}\}$ and proved for K' a version of Corollary 1.9 with $m \sim \delta^{-4} s \log(n/s)$. Since the mean widths of both sets K and K' are of the order $\sqrt{s \log(n/s)}$, Theorem 1.5 holds for these sets with $m \sim \delta^{-6} s \log(n/s)$. In other words, apart from the dependence of δ (which is an interesting problem), the prior results follow as partial cases from Theorem 1.5.

It is important to note that Theorem 1.5 addresses only the theoretical aspect of one-bit compressed sensing problem, which guarantees that the quantized measurement map $f(x) = \text{sign } Ax$ well preserves the geometry of signals. But one also faces an algorithmic challenge – how to efficiently recover x from $f(x)$, and specifically in polynomial time. We will not touch on this algorithmic aspect here but rather refer the reader to [19] and to our forthcoming work, which is based on the results of this paper.

1.10 Related Work: Locality-Sensitive Hashing

Locality-sensitive hashing is a method of dimension reduction. One takes a set of high-dimensional vectors in $\mathbb{R}^n$ and the goal is to hash nearby vectors to the same bin with high probability. More generally, one may desire that the distance between bins be nearly proportional to the distance between the original items. There have been a number of papers, which suggest to create such mappings onto the Hamming cube [1, 4, 8, 10, 14], some of which use a random hyperplane tessellation as defined in this paper. The new challenge considered herein is to create a locality-sensitive hashing for an infinite set.

1.11 Overview of the Argument

Let us briefly describe our proof of the results stated above. Since the distance in the Hamming cube $\{-1, 1\}^m$ can be expressed as $(2m)^{-1} \|x - y\|_1$, the Hamming cube is isometrically embedded in ℓ_1^m. Before trying to embed $K \subseteq S^{n-1}$ into the Hamming cube as claimed in Theorem 1.5, we shall make a simpler step and embed K almost isometrically into the bigger space ℓ_1^m with $m \sim \delta^{-2} w(K)^2$. A result of this type was given by Schechtman [21]. In Sect. 2, we prove a similar result by a simple and direct argument in probability in Banach spaces.

Our next and nontrivial step is to reembed the set from ℓ_1^m into its subset, the Hamming cube $\{-1, 1\}^m$. In Sect. 3, we give a simple "curvature argument" that allows us to deduce Corollary 1.9 on the diameter of cells, and even with a better dependence on δ, namely $m \sim \delta^{-4} w(K)^2$. However, a genuine limitation of the curvature argument makes it too weak to deduce Theorem 1.2 this way.

We instead attempt to prove Theorem 1.2 by an ε-net argument, which typically proceeds as follows: (a) show that $d_A(x, y) \approx d(x, y)$ holds for a fixed pair $x, y \in K$ with high probability; (b) take the union bound over all pairs x, y in an finite ε-net N_ε of K; (c) extend the estimate from N_ε to K by approximation. Unfortunately, as we indicate in Sect. 4 the approximation step (c) must fail due to the discontinuity of the Hamming distance $d_A(x, y)$.

A solution proposed in [5, 11] was to choose ε so small that none of the random hyperplanes pass near points $x, y \in N_\varepsilon$ with high probability. This strategy was effective for the set $K = \{x \in S^{n-1} : |\operatorname{supp}(x)| \leq s\}$ because the covering number of this specific set K has a mild (logarithmic) dependence on ε, namely $\log N(K, \varepsilon) \leq s \log(Cn/\varepsilon s)$. However, adapting this strategy to general sets K would cause our estimate on m to increase by a factor of n.

The solution we propose in the present paper is to "soften" the Hamming distance; see Sect. 4 for the precise notion. The *soft Hamming distance* enjoys some continuity properties as described in Lemmas 4.3 and 5.5. In Sect. 5.5, we develop the ε-net argument for the soft Hamming distance. Interestingly, the approximation step (c) for the soft Hamming distance will be based on the embedding of K into ℓ_1^m, which incidentally was our point of departure.

1.12 Notation

Throughout the paper, C, c, C_1, etc. denote positive absolute constants whose values may change from line to line. For integer n, we denote $[n] = \{1, \dots, n\}$. The ℓ_p norms

of a vector $x \in \mathbb{R}^n$ for $p \in \{0, 1, 2, \infty\}$ are defined as[2]

$$\|x\|_0 = |\operatorname{supp}(x)| = |\{i \in [n] : x(i) \neq 0\}|,$$

$$\|x\|_1 = \sum_{i=1}^{n} |x_i|, \qquad \|x\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}, \qquad \|x\|_\infty = \max_{i \in [n]} |x_i|.$$

We shall work with normed spaces $\ell_p^n = (\mathbb{R}^n, \|\cdot\|_p)$ for $p \in \{1, 2, \infty\}$. The unit Euclidean ball in $\mathbb{R}^n$ is denoted $B_2^n = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ and the unit Euclidean sphere is denoted $S^{n-1} = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$.

As usual, $\mathcal{N}(0, 1)$ stands for the univariate normal distribution with zero mean and unit variance, and $\mathcal{N}(0, I_n)$ stands for the multivariate normal distribution in $\mathbb{R}^n$ with zero mean and whose covariance matrix is identity I_n.

2 Embedding into ℓ_1

Lemma 2.1 (Concentration) *Consider a bounded subset $K \subset \mathbb{R}^n$ and independent random vectors $a_1, \ldots, a_m \sim \mathcal{N}(0, I_n)$ in $\mathbb{R}^n$. Let*

$$Z = \sup_{x \in K} \left| \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, x \rangle| - \sqrt{\frac{2}{\pi}} \|x\|_2 \right|.$$

(a) *One has*

$$\mathbb{E}Z \leq \frac{4w(K)}{\sqrt{m}}. \tag{2.1}$$

(b) *The following deviation inequality holds*:

$$\Pr Z > \frac{4w(K)}{\sqrt{m}} + u \leq 2\exp\left(-\frac{mu^2}{2d(K)^2}\right), \quad u > 0, \tag{2.2}$$

where $d(K) = \max_{x \in K} \|x\|_2$.

Proof (a) Note that $\mathbb{E}|\langle a_i, x \rangle| = \sqrt{\frac{2}{\pi}} \|x\|_2$ for all i. Let $\varepsilon_1, \ldots, \varepsilon_m$ be a sequence of iid Rademacher random variables. A standard symmetrization argument (see [15, Lemma 6.3]) followed by the contraction principle (see [15, Theorem 4.12]) yields that

$$\mathbb{E}Z \leq 2\mathbb{E} \sup_{x \in K} \left| \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i |\langle a_i, x \rangle| \right| \leq 4\mathbb{E} \sup_{x \in K} \left| \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i \langle a_i, x \rangle \right|$$

$$= 4\mathbb{E} \sup_{x \in K} \left| \left\langle \frac{1}{m} \sum_{i=1}^{m} \varepsilon_i a_i, x \right\rangle \right|.$$

---

[2] Note that, strictly speaking, $\|\cdot\|_0$ is not a norm on $\mathbb{R}^n$.

By the rotational invariance of the Gaussian distribution, $\frac{1}{m}\sum_{i=1}^{m}\varepsilon_i a_i$ is distributed identically with $g/\sqrt{m}$ where $g \sim \mathcal{N}(0, I_n)$. Therefore,

$$\mathbb{E}Z \leq \frac{4}{\sqrt{m}}\mathbb{E}\sup_{x\in K}\left|\langle g, x\rangle\right| = \frac{4w(K)}{\sqrt{m}}.$$

This proves the upper bound in (2.1).

(b) We combine the result of (a) with the Gaussian concentration inequality. To this end, we must first show that the map $A \mapsto Z = Z(A)$ is Lipschitz where $A = (a_1, \ldots, a_m)$ is considered as a matrix in the space $\mathbb{R}^{nm}$ equipped with Frobenius norm $\|\cdot\|_F$ (which coincides with the Euclidean norm on $\mathbb{R}^{nm}$). It follows from two applications of the triangle inequality followed by two applications of the Cauchy–Schwarz inequality that for $A = (a_1, \ldots, a_m), B = (b_1, \ldots, b_m) \in \mathbb{R}^{nm}$ we have

$$\left|Z(A) - Z(B)\right| \leq \sup_{x\in K}\frac{1}{m}\sum_{i=1}^{m}\left|\langle a_i - b_i, x\rangle\right| \leq \frac{d(K)}{m}\sum_{i=1}^{m}\|a_i - b_i\|_2$$

$$\leq \frac{d(K)}{\sqrt{m}}\|A - B\|_F.$$

Thus, Z has Lipschitz constant bounded by $d(K)/\sqrt{m}$. We may now bound the deviation probability for Z using the Gaussian concentration inequality (see [15, Eq. (1.6)]) as follows:

$$\Pr|Z - \mathbb{E}Z| \geq u \leq 2\exp\left(-mu^2/2d(K)^2\right).$$

The deviation inequality (2.2) now follows from the bound on $\mathbb{E}Z$ from (a). $\qquad\square$

Remark 2.2 (Random matrix formulation) One can state Lemma 2.1 in terms of random matrices. Indeed, let A be an $m \times n$ random matrix with independent $\mathcal{N}(0, 1)$ entries. Then its rows a_i satisfy the assumption of Lemma 2.1, and we can express Z as

$$Z = \sup_{x\in K}\left|\frac{1}{m}\|Ax\|_1 - \sqrt{\frac{2}{\pi}}\|x\|_2\right|. \tag{2.3}$$

Using this remark for the set $K - K$, we obtain a linear embedding of K into ℓ_1:

Corollary 2.3 (Embedding into ℓ_1) *Consider a subset $K \subset \ell_2^n$ and let $\delta > 0$. Let*

$$m \geq C\delta^{-2}w(K)^2.$$

Then, with probability at least $1 - 2\exp(-m\delta^2/32)$, the linear map $f : K \to \ell_1^m$ defined as $f(x) = \frac{1}{m}\sqrt{\frac{\pi}{2}}Ax$ is a δ-isometry. Thus K can be linearly embedded into ℓ_1^m with Gromov-Haussdorff distortion at most δ.

Proof Let A be the random matrix as in Remark 2.2. Using Lemma 2.1 for $K - K$ and noting the form of Z in (2.3), we conclude that the following event holds with

probability at least $1 - 2\exp(-m\delta^2/32)$:

$$\left|\frac{1}{m}\|Ax - Ay\|_1 - \sqrt{\frac{2}{\pi}}\|x - y\|_2\right| \leq \frac{8w(K - K)}{\sqrt{m}} \leq \frac{16w(K)}{\sqrt{m}} \leq \delta, \quad x, y \in K. \quad \square$$

Remark 2.4 The above argument shows in fact that Corollary 2.3 holds for

$$m \geq C\delta^{-2}w(K - K)^2.$$

As we noticed in Remark 1.11, the quantity $w(K - K)$ more accurately reflects the geometric meaning of the mean width than $w(K)$.

Remark 2.5 (Low M* estimate) Note that for the subspace $E = \ker A$ we have from (2.3) that $Z \geq \sup_{x \in K \cap E} \sqrt{\frac{2}{\pi}}\|x\|_2 = \sqrt{\frac{2}{\pi}}d(K \cap E)$. Then Lemma 2.1 implies that

$$\mathbb{E}d(K \cap E) \leq \frac{6w(K)}{\sqrt{m}}. \tag{2.4}$$

By rotation invariance of Gaussian distribution, inequality (2.4) holds for a random subspace E in $\mathbb{R}^n$ of given codimension $m \leq n$, uniformly distributed according to the Haar measure. This result recovers (up to the absolute constant 6 which can be improved) the so-called *low M* estimate* from geometric functional analysis; see [15, Sect. 15.1].

Remark 2.6 (Dimension reduction) As we emphasized in the Introduction, for many sets $K \subset \mathbb{R}^n$ one has $w(K) \ll n$. In such cases, Corollary 2.3 works for $m \ll n$. The embedding of K into ℓ_1^m yields dimension reduction for K (from n to $m \ll n$ dimensions).

    For example, if K is a finite set then $w(K) \leq C\sqrt{\log|K|}$ (see, e.g., [15, (3.13)]), and so Corollary 2.3 applies with $m \sim \log|K|$. This gives the following variant of the *Johnson–Lindenstrauss lemma*: every finite subset of a Euclidean space can be linearly embedded in ℓ_1^m with $m \sim \log|K|$ and with small distortion in the Gromov–Haussdorff metric. Stronger variants of Johnson–Lindenstrauss lemma are known for *Lipschitz* rather than Gromov–Haussdorff embeddings into ℓ_2^m and ℓ_1^m [2, 21]. However, for general sets K (in particular for any set with nonempty interior), a Lipschitz embedding into lower dimensions is clearly impossible; still a Gromov–Haussdorff embedding exists due to Corollary 2.3.

3 Proof of Corollary 1.9 by a Curvature Argument

In this section, we give a short argument that leads to a version of Corollary 1.9 with a slightly better dependence of m on δ.

Theorem 3.1 (Cells of random uniform tessellations) *Consider a subset $K \subseteq S^{n-1}$ and let $\delta > 0$. Let*

$$m \geq C\delta^{-4}w(K)^2$$

and consider an arrangement of m independent random hyperplanes in $\mathbb{R}^n$ that are uniformly distributed according to the Haar measure. Then, with probability at least $1 - 2\exp(-c\delta^4 m)$, all cells of the tessellation have diameter at most δ.

The argument is based on Lemma 2.1. If points $x, y \in K$ belong to the same cell, then the midpoint $z = \frac{1}{2}(x + y)$ also belongs to the same cell (after normalization). Using Lemma 2.1, one can then show that $\|z\|_2 \approx \frac{1}{2}(\|x\|_2 + \|y\|_2) = 1$. Due to the curvature of the sphere, this forces the length of the interval $\|x - y\|_2$ to be small, which means that the diameter of the cell is small. The formal argument is below.

Proof We represent the random hyperplanes as $\{a_i\}^\perp$, where $a_1, \ldots, a_m \sim \mathcal{N}(0, I_n)$ are independent random vectors in $\mathbb{R}^n$. Let δ, m be as in the assumptions of the theorem. We shall apply Lemma 2.1 for the sets K and $\frac{1}{2}(K + K)$ and for $u = \varepsilon/2$, where we set $\varepsilon = \delta^2/16$. Since the diameters of both these sets are bounded by 1, we obtain that with probability at least $1 - 2\exp(-c\delta^4 m)$ the following event holds:

$$\left| \sqrt{\frac{\pi}{2}} \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, v \rangle| - \|v\|_2 \right| < \varepsilon, \quad v \in K \cup \frac{1}{2}(K + K). \tag{3.1}$$

Assume that the event (3.1) holds. Consider a pair of points $x, y \in K$ that belong to the same cell of the tessellation, which means that

$$\text{sign}\langle a_i, x \rangle = \text{sign}\langle a_i, y \rangle, \quad i \in [m].$$

To complete the proof, it suffices to show that $\|x - y\|_2 \le \delta$. This will give desired diameter δ in the Euclidean metric. Furthermore, since for small δ the Euclidean and the geodesic distances are equivalent, the conclusion will hold for the geodesic distance as well.

We shall use (3.1) for $x, y \in K$ and for the midpoint $z := \frac{1}{2}(x + y) \in \frac{1}{2}(K + K)$. Clearly, $\text{sign}\langle a_i, z \rangle = \text{sign}\langle a_i, x \rangle = \text{sign}\langle a_i, y \rangle$, hence

$$|\langle a_i, z \rangle| = |\langle a_i, x \rangle| + |\langle a_i, y \rangle|, \quad i \in [m].$$

Therefore, we obtain from (3.1) that

$$\|z\|_2 \ge \sqrt{\frac{\pi}{2}} \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, z \rangle| - \varepsilon$$

$$= \frac{1}{2}\left[\sqrt{\frac{\pi}{2}} \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, x \rangle| + \sqrt{\frac{\pi}{2}} \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, y \rangle| \right] - \varepsilon$$

$$\ge \frac{1}{2}\left(\|x\|_2 - \varepsilon + \|y\|_2 - \varepsilon \right) - \varepsilon = 1 - 2\varepsilon. \tag{3.2}$$

By the parallelogram law, we conclude that

$$\|x - y\|_2^2 = 4 - \|x + y\|_2^2 = 4\left(1 - \|z\|_2^2\right) \le 16\varepsilon = \delta^2.$$

This completes the proof. □

3.1 Limitations of the Curvature Argument

Unfortunately, the curvature argument does not lend itself to proving the more general result, Theorem 1.2 on uniform tessellations. To see why, suppose $x, y \in K$ do not belong to the same cell but instead $d_A(x, y) = d$ for some small $d \in (0, 1)$. Consider the set of mismatched signs

$$T := \big\{ i \in [m] : \text{sign}\langle a_i, x \rangle \neq \text{sign}\langle a_i, y \rangle \big\}; \qquad \frac{|T|}{m} = d.$$

These signs create an additional error term on the right-hand side of (3.2), which is

$$\sqrt{\frac{\pi}{2}} \frac{1}{m} \sum_{i \in T} \big| \langle a_i, v_i \rangle \big| \quad \text{where } v_i \in \{x, y\}. \tag{3.3}$$

By analogy with Lemma 2.1, we can expect that this term should be approximately equal $|T|/m = d$. If this is true, then (3.2) becomes in our situation $\|z\|_2 \geq 1 - 2\varepsilon - d$, which leads as before to $\|x - y\|_2^2 \lesssim \varepsilon + d$. Ignoring ε, we see that the best estimate the curvature argument can give is $d(x, y) \lesssim \sqrt{d_A(x, y)}$ rather than $d(x, y) \lesssim d_A(x, y)$ that is required in Theorem 1.2.

The weak point of this argument is that it takes into account the size of T but ignores the nature of T. For every $i \in T$, the hyperplane $\{a_i\}^\perp$ passes through the arc connecting x and y. If the length of the arc $d(x, y)$ is small, this creates a strong constraint on a_i. Conditioning the distribution of a_i on the constraint that $i \in T$ creates a bias toward smaller values of $|\langle a_i, x \rangle|$ and $|\langle a_i, y \rangle|$. As a result, the conditional expected value of the error term (3.3) should be smaller than d. Computing this conditional expectation is not a problem for a given pair x, y, but it seems to be difficult to carry out a uniform argument over $x, y \in K$ where the (conditional) distribution of a_i depends on x, y.

We instead propose a different and somewhat more conceptual way to deduce Theorem 1.2 from Lemma 2.1. This argument will be developed in the rest of this paper.

3.2 Dvoretzky Theorem and Dependence on δ

The unusual dependence δ^{-4} in Theorem 3.1 is related to the open problem of the optimal dependence on distortion in the Dvoretzky theorem.

Indeed, consider the special case of the tessellation problem where $K = S^{n-1}$ and $w(K) \sim \sqrt{n}$. Then Lemma 2.1 in its geometric formulation (see Eq. (2.3) and Corollary 2.3) states that ℓ_2^n embeds into ℓ_1^m whenever $m \geq C\varepsilon^{-2}n$, meaning that

$$(1 - \varepsilon)\|x\|_2 \leq \|\Phi x\|_1 \leq (1 + \varepsilon)\|x\|_2, \quad x \in \mathbb{R}^n,$$

where $\Phi = \sqrt{\frac{\pi}{2}} \frac{1}{m} A$. Equivalently, there exists an n-dimensional subspace of ℓ_1^m that is $(1 + \varepsilon)$-Euclidean, where $n \sim \varepsilon^2 m$. This result recovers the well-known Dvoretzky theorem in V. Milman's formulation (see [9, Theorem 4.2.1]) for the space ℓ_1^m, and with the best known dependence on ε. However, it is not known whether ε^2 is
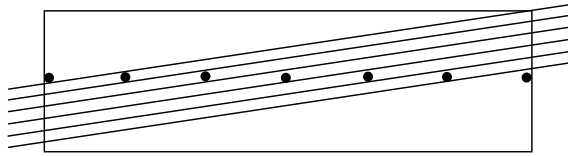
Fig. 3 This hyperplane tessellation of the set $K = [-\frac{1}{2}, \frac{1}{2}] \times [-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]$ is very non-uniform, as all cells have diameter at least 1. The tessellation is nevertheless very uniform for the ε-net $N_\varepsilon = \varepsilon\mathbb{Z} \cap K$, as $d_A(x, y) = \|x - y\|_2$ for all $x, y \in N_\varepsilon$

the optimal dependence for ℓ_1^m; see [21] for a discussion of the general problem of dependence on ε in the Dvoretzky theorem.

These observations suggest that we can reverse our logic. Suppose one can prove the Dvoretzky theorem for ℓ_1^m with a better dependence on ε, thereby constructing a $(1 + \varepsilon)$-Euclidean subspace of dimension $n \sim f(\varepsilon)m$ with $f(\varepsilon) \gg \varepsilon^2$. Then such construction can replace Lemma 2.1 in the curvature argument. This will lead to Theorem 3.1 for $K = S^{n-1}$ with an improved dependence on δ, namely with $m \sim f(\delta^2)n$. Concerning lower bounds, the best possible dependence of m on δ should be δ^{-1}, which follows by considering the case $n = 2$. This dependence will be achieved if the Dvoretzky theorem for ℓ_1^m is valid with $n \sim \varepsilon^{1/2}m$. This is unknown.

4 Toward Theorem 1.2: a Soft Hamming Distance

Our proof of Theorem 1.2 will be based on a covering argument. A standard covering argument of geometric functional analysis would proceed in our situation as follows:

(a) Show that $d_A(x, y) \approx d(x, y)$ with high probability for a fixed pair x, y. This can be done using standard concentration inequalities.
(b) Prove that $d_A(x, y) \approx d(x, y)$ uniformly for all x, y in a finite ε-net N_ε of K. Sudakov's inequality can be used to estimate the cardinality of N_ε via the mean width $w(K)$. The conclusion will follow from step 1 by the union bound over $(x, y) \in N_\varepsilon \times N_\varepsilon$.
(c) Extend the estimate $d_A(x, y) \approx d(x, y)$ from $x, y \in N_\varepsilon$ to $x, y \in K$ by approximation.

While the first two steps are relatively standard, step (c) poses a challenge in our situation. The Hamming distance $d_A(x, y)$ is a discontinuous function of x, y, so it is not clear whether the estimate $d_A(x, y) \approx d(x, y)$ can be extended from a pair points $x, y \in N_\varepsilon$ to a pair of nearby points. In fact, for some tessellations this task is impossible. Figure 3 shows that there exist very nonuniform tessellations that are nevertheless very uniform for an ε-net, namely one has $d_A(x, y) = d(x, y)$ for all $x, y \in N_\varepsilon$. The set K in that example is a subset of the plane $\mathbb{R}^2$, and one can clearly embed such a set with into the sphere S^2 as well.

To overcome the discontinuity problem, we propose to work with a soft version of the Hamming distance. Recall that m hyperplanes are determined by their normals $a_1, \ldots, a_m \in \mathbb{R}^n$, which we organize in an $m \times n$ matrix A with rows a_i. Then the

usual ("hard") Hamming distance $d_A(x, y)$ on $\mathbb{R}^n$ with respect to A with can be expressed as

$$d_A(x, y) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\mathcal{E}_i}, \quad \text{where } \mathcal{E}_i = \big\{\mathrm{sign}\langle a_i, x\rangle \neq \mathrm{sign}\langle a_i, y\rangle\big\}. \qquad (4.1)$$

Definition 4.1 (Soft Hamming distance) Consider an $m \times n$ matrix A with rows $a_1, \ldots, a_m$, and let $t \in \mathbb{R}$. The *soft Hamming distance* $d_A^t(x, y)$ on $\mathbb{R}^n$ is defined as

$$d_A^t(x, y) = \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\mathcal{F}_i}, \quad \text{where}$$

$$\mathcal{F}_i = \big\{\langle a_i, x\rangle > t, \langle a_i, y\rangle < -t\big\} \cup \big\{-\langle a_i, x\rangle > t, -\langle a_i, y\rangle < -t\big\}. \qquad (4.2)$$

Both positive and negative t may be considered. For positive t, the soft Hamming distance counts the hyperplanes that separate x, y well enough; for negative t, it counts the hyperplanes that separate or nearly separate x, y.

Remark 4.2 (Comparison of soft and hard Hamming distances) Clearly, $d_A^t(x, y)$ is a nonincreasing function of t. Moreover,

$$d_A^t(x, y) = d_A(x, y) \quad \text{for } t = 0;$$
$$d_A^t(x, y) \leq d_A(x, y) \quad \text{for } t \geq 0;$$
$$d_A^t(x, y) \geq d_A(x, y) \quad \text{for } t \leq 0.$$

The soft Hamming distance for a fixed t is as discontinuous as the usual (hard) Hamming distance. However, some version of continuity emerges when we allow t to vary slightly.

Lemma 4.3 (Continuity) *Let $x, y, x', y' \in \mathbb{R}^n$, and assume that $\|Ax'\|_\infty \leq \varepsilon$, $\|Ay'\|_\infty \leq \varepsilon$ for some $\varepsilon > 0$. Then for every $t \in \mathbb{R}$ one has*

$$d_A^{t+\varepsilon}(x, y) \leq d_A^t\big(x + x', y + y'\big) \leq d_A^{t-\varepsilon}(x, y).$$

Proof Consider the events $\mathcal{F}_i = \mathcal{F}_i(x, y, t)$ from the definition of the soft Hamming distance (4.2). By the assumptions, we have $|\langle a_i, x'\rangle| \leq \varepsilon$, $|\langle a_i, y'\rangle| \leq \varepsilon$ for all $i \in [m]$. This implies by the triangle inequality that

$$\mathcal{F}_i(x, y, t+\varepsilon) \subseteq \mathcal{F}_i\big(x + x', y + y', t\big) \subseteq \mathcal{F}_i(x, y, t-\varepsilon).$$

The conclusion of the lemma follows. $\qquad\qquad\square$

We are ready to state a stronger version of Theorem 1.2 for the soft Hamming distance.

Theorem 4.4 (Random uniform tessellations: soft version) *Consider a subset $K \subseteq S^{n-1}$ and let $\delta > 0$. Let*

$$m \geq C\delta^{-6}w(K)^2$$

and pick $t \in \mathbb{R}$. Consider an $m \times n$ random (Gaussian) matrix A with independent rows $a_1, \ldots, a_m \sim \mathcal{N}(0, I_n)$. Then with probability at least $1 - \exp(-c\delta^2 m)$, one has

$$\left| d_A^t(x, y) - d(x, y) \right| \leq \delta + 2|t|, \quad x, y \in K.$$

Note that if we take $t = 0$ in the above theorem, we recover Theorem 1.2. However, we find it easier to prove the result for general t, since in our argument we will work with different values of the t for the soft Hamming distance.

Theorem 4.4 is proven in the next section.

5 Proof of Theorem 4.4 on the Soft Hamming Distance

We will follow the covering argument outlined in the beginning of Sect. 4, but instead of $d_A(x, y)$ we shall work with the soft Hamming distance $d_A^t(x, y)$.

5.1 Concentration of Distance for a Given Pair

At the first step, we will check that $d_A^t(x, y) \approx d(x, y)$ with high probability for a fixed pair x, y. Let us first verify that this estimate holds in expectation, i.e., that $\mathbb{E}d_A^t(x, y) \approx d(x, y)$. One can easily check that

$$\mathbb{E}d_A(x, y) = d(x, y), \tag{5.1}$$

so we may just compare $\mathbb{E}d_A^t(x, y)$ to $\mathbb{E}d_A(x, y)$. Here is a slightly stronger result.

Lemma 5.1 (Comparing soft and hard Hamming distances in expectation) *Let A be a random Gaussian matrix be as in Theorem 4.4. Then, for every $t \in \mathbb{R}$ and every x, $y \in \mathbb{R}^n$, one has*

$$\left| \mathbb{E}d_A^t(x, y) - d(x, y) \right| \leq \mathbb{E}\left| d_A^t(x, y) - d_A(x, y) \right| \leq 2|t|.$$

Proof The first inequality follows from (5.1) and Jensen's inequality. To prove the second inequality, we use the events $\mathcal{E}_i$ and $\mathcal{F}_i$ from Eqs. (4.1), (4.2) defining the hard and soft Hamming distances, respectively. It follows that

$$\mathbb{E}\left| d_A^t(x, y) - d_A(x, y) \right|$$

$$= \mathbb{E}\left| \frac{1}{m} \sum_{i=1}^m (\mathbf{1}_{\mathcal{E}_i} - \mathbf{1}_{\mathcal{F}_i}) \right|$$

$$\leq \mathbb{E}|\mathbf{1}_{\mathcal{E}_1} - \mathbf{1}_{\mathcal{F}_1}| \quad \text{(by triangle inequality and identical distribution)}$$

$$= \mathbb{P}\{\mathcal{E}_1 \triangle \mathcal{F}_1\}$$

$$\leq \mathbb{P}\{|\langle a_1, x\rangle| \leq |t|\} + \mathbb{P}\{|\langle a_1, y\rangle| \leq |t|\}$$

$$\leq 2\mathbb{P}\{|g| \leq |t|\} \quad (\text{where } g \sim \mathcal{N}(0,1))$$

$$\leq 2|t| \quad (\text{by the density of the normal distribution}). \qquad \square$$

Now we upgrade Lemma 5.1 to an concentration inequality.

Lemma 5.2 (Concentration of distance) *Let A be a random Gaussian matrix as in Theorem 4.4. Then, for every $t \in \mathbb{R}$ and every $x, y \in \mathbb{R}^n$, the following deviation inequality holds*:

$$\mathbb{P}\{|d_A^t(x,y) - d(x,y)| > 2|t| + \delta\} \leq 2\exp(-2\delta^2 m), \quad \delta > 0.$$

Proof By definition, $m \cdot d_A^t(x,y)$ has the binomial distribution $\text{Bin}(m,p)$. The parameter $p = \mathbb{E}d_A^t(x,y)$ satisfies by Lemma 5.1 that

$$|p - d(x,y)| \leq 2|t|.$$

A standard Chernoff bound for binomial random variables states that

$$\mathbb{P}\{|d_A^t(x,y) - p| > \delta\} \leq 2\exp(-2\delta^2 m), \quad \delta > 0,$$

see, e.g., [3, Corollary A.1.7]. The triangle inequality completes the proof. $\qquad \square$

5.2 Concentration of Distance over an ε-Net

Let us fix a small $\varepsilon > 0$ whose value will be determined later. Let N_ε be an ε-net of K in the Euclidean metric. By Sudakov's inequality (see [15, Theorem 3.18]), we can arrange the cardinality of N_ε to satisfy

$$\log|N_\varepsilon| \leq C\varepsilon^{-2}w(K)^2. \tag{5.2}$$

We can decompose every vector $x \in K$ into a *center* x_0 and a *tail* x' so that

$$x = x_0 + x', \quad \text{where } x_0 \in N_\varepsilon, \ x' \in (K - K) \cap \varepsilon B_2^n. \tag{5.3}$$

We first control the centers by taking a union bound in Lemma 5.2 over the net N_ε.

Lemma 5.3 (Concentration of distance over a net) *Let A a random Gaussian matrix be as in Theorem 4.4. Let N_ε be a subset of S^{n-1} whose cardinality satisfies (5.2). Let $\delta > 0$, and assume that*

$$m \geq C\varepsilon^{-2}\delta^{-2}w(K)^2. \tag{5.4}$$

Let $t \in \mathbb{R}$. Then the following holds with probability at least $1 - 2\exp(-\delta^2 m)$:

$$|d_A^t(x_0, y_0) - d(x_0, y_0)| \leq 2|t| + \delta, \quad x_0, y_0 \in N_\varepsilon.$$

Proof By Lemma 5.3 and a union bound over the set of pairs $(x_0, y_0) \in N_\varepsilon \times N_\varepsilon$, we obtain

$$\mathbb{P}\left\{ \sup_{x,y \in N_\varepsilon} \left| d_A^t(x, y) - d(x, y) \right| > 2|t| + \delta \right\} \leq |N_\varepsilon|^2 \cdot 2\exp\left(-2\delta^2 m\right) \leq 2\exp\left(-\delta^2 m\right)$$

where the last inequality follows by (5.2) and (5.4). The proof is complete. ☐

5.3 Control of the Tails

Now we control the tails $x' \in (K - K) \cap \varepsilon B_2^n$ in decomposition (5.3).

Lemma 5.4 (Control of the tails) *Consider a subset $K \subseteq S^{n-1}$ and let $\varepsilon > 0$. Let*

$$m \geq C\varepsilon^{-2} w(K)^2.$$

Consider independent random vectors $a_1, \ldots, a_m \sim \mathcal{N}(0, I_n)$. Then with probability at least $1 - 2\exp(-cm)$, one has

$$\frac{1}{m} \sum_{i=1}^{m} |\langle a_i, x' \rangle| \leq \varepsilon \quad \text{for all } x' \in (K - K) \cap \varepsilon B_2^n.$$

Proof Let us apply Lemma 2.1 for the set $T = (K - K) \cap \varepsilon B_2^n$ instead of K, and for $u = \varepsilon/8$. Since $d(K) = \max_{x' \in T} \|x'\|_2 \leq \varepsilon$, we obtain that the following holds with probability at least $1 - 2\exp(-cm)$:

$$\sup_{x' \in T} \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, x' \rangle| \leq \sup_{x' \in T} \left| \frac{1}{m} \sum_{i=1}^{m} |\langle a_i, x' \rangle| - \sqrt{\frac{2}{\pi}} \|x'\|_2 \right| + \sqrt{\frac{2}{\pi}} \varepsilon$$

$$\leq \frac{4w(T)}{\sqrt{m}} + \frac{\varepsilon}{8} + \sqrt{\frac{2}{\pi}} \varepsilon. \tag{5.5}$$

Note that $w(T) \leq w(K - K) \leq 2w(K)$. So, using the assumption on m, we conclude that the quantity in (5.5) is bounded by ε, as claimed. ☐

5.4 Approximation

Now we establish a way to transfer the distance estimates from an ε-net N_ε to the full set K. This is possible by a continuity property of the soft Hamming distance, which we outlined in Lemma 4.3. This result requires the perturbation to be bounded in L_∞ norm. However, in our situation, the perturbations are going to be bounded only in L_1 norm due to Lemma 5.4. So, we shall prove the following relaxed version of continuity.

Lemma 5.5 (Continuity with respect to L_1 perturbations) *Let $x, y, x', y' \in \mathbb{R}^n$, and assume that $\|Ax'\|_1 \leq \varepsilon m$, $\|Ay'\|_1 \leq \varepsilon m$ for some $\varepsilon > 0$. Then for every $t \in \mathbb{R}$ and*

$M \geq 1$ *one has*

$$d_A^{t+M\varepsilon}(x, y) - \frac{2}{M} \leq d_A^t(x + x', y + y') \leq d_A^{t-M\varepsilon}(x, y) + \frac{2}{M}. \qquad (5.6)$$

Proof Consider the events $\mathcal{F}_i = \mathcal{F}_i(x, y, t)$ from the definition of the soft Hamming distance (4.2). By the assumptions, we have

$$\sum_{i=1}^{m} |\langle a_i, x' \rangle| \leq \varepsilon m, \qquad \sum_{i=1}^{m} |\langle a_i, y' \rangle| \leq \varepsilon m.$$

Therefore, the set

$$T := \left\{ i \in [m] : |\langle a_i, x' \rangle| \leq M\varepsilon, |\langle a_i, y' \rangle| \leq M\varepsilon \right\} \quad \text{satisfies} \quad |T^c| \leq 2m/M.$$

By the triangle inequality, we have

$$\mathcal{F}_i(x, y, t + M\varepsilon) \subseteq \mathcal{F}_i(x + x', y + y', t) \subseteq \mathcal{F}_i(x, y, t - M\varepsilon), \quad i \in T.$$

Therefore,

$$
\begin{aligned}
d_A^{t+M\varepsilon}(x, y) &= \frac{1}{m} \sum_{i=1}^{m} \mathbf{1}_{\mathcal{F}_i(x,y,t+M\varepsilon)} \leq \frac{|T^c|}{m} + \frac{1}{m} \sum_{i \in T} \mathbf{1}_{\mathcal{F}_i(x,y,t+M\varepsilon)} \\
&\leq \frac{2}{M} + \frac{1}{m} \sum_{i \in T} \mathbf{1}_{\mathcal{F}_i(x+x',y+y',t)} \leq \frac{2}{M} + d_A^t(x + x', y + y').
\end{aligned}
$$

This proves the first inequality in (5.6). The proof of the second inequality is similar. □

5.5 Proof of Theorem 4.4

Now we are ready to combine all the pieces and prove Theorem 4.4. To this end, consider the set K, numbers δ, m, t, and the random matrix A as in the theorem. Choose $\varepsilon = \delta^2/100$ and $M = 10/\delta$.

Consider an ε-net N_ε of K as we described in the beginning of Sect. 5.2. Let us apply Lemma 5.3 that controls the distances on N_ε along with Lemma 5.4 that controls the tails. By the assumption on m in the theorem and by our choice of ε, both requirements on m in these lemmas hold. By a union bound, with probability at least $1 - 4\exp(-c\delta^2 m)$ the following event holds: for every $x_0, y_0 \in N_\varepsilon$ and x', $y' \in (K - K) \cap \varepsilon B_2^n$, one has

$$
\begin{aligned}
\left| d_A^{t-M\varepsilon}(x_0, y_0) - d(x_0, y_0) \right| &\leq 2|t - M\varepsilon| + \delta/2, \\
\left| d_A^{t+M\varepsilon}(x_0, y_0) - d(x_0, y_0) \right| &\leq 2|t + M\varepsilon| + \delta/2,
\end{aligned}
\qquad (5.7)
$$

$$\|Ax'\|_1 \leq \varepsilon m, \qquad \|Ay'\|_1 \leq \varepsilon m. \qquad (5.8)$$

Let $x, y \in K$. As we described in (5.3), we can decompose the vectors as

$$x = x_0 + x', \quad y = y_0 + y', \quad \text{where } x_0, y_0 \in N_\varepsilon, \ x', y' \in (K - K) \cap \varepsilon B_2^n. \quad (5.9)$$

The bounds in (5.8) guarantee that the continuity property (5.6) in Lemma 5.5 holds. This gives

$$d_A^t(x, y) \le d_A^{t - M\varepsilon}(x_0, y_0) + \frac{2}{M}$$

$$\le d(x_0, y_0) + 2|t| + 2M\varepsilon + \frac{\delta}{2} + \frac{2}{M}$$

by (5.7) and the triangle inequality. Furthermore, using (5.9) we have

$$\left| d(x_0, y_0) - d(x, y) \right| \le d(x_0, x) + d(y_0, y) \le \|x_0 - x\|_2 + \|y_0 - y\|_2 \le 2\varepsilon.$$

It follows that

$$d_A^t(x, y) \le d(x, y) + 2|t| + 2M\varepsilon + \frac{\delta}{2} + \frac{2}{M} + 2\varepsilon.$$

Finally, by the choice of ε and M, we obtain

$$d_A^t(x, y) \le d(x, y) + 2|t| + \delta.$$

A similar argument shows that

$$d_A^t(x, y) \ge d(x, y) - 2|t| - \delta.$$

We conclude that

$$\left| d_A^t(x, y) - d(x, y) \right| \le \delta + 2|t|.$$

This completes the proof of Theorem 4.4.

6 Proof of Theorem 1.10 on Tessellations in $\mathbb{R}^n$

In this section, we deduce Theorem 1.10 from Theorem 1.2 by an elementary lifting argument into $\mathbb{R}^{n+1}$. We shall use the following notation: Given a vector $x \in \mathbb{R}^n$ and a number $t \in \mathbb{R}$, the vector $x \oplus t \in \mathbb{R}^n \oplus \mathbb{R} = \mathbb{R}^{n+1}$ is the concatenation of $x \in \mathbb{R}^n$ and t. Furthermore, $K \oplus t$ denotes the set of all vectors $x \oplus t$ where $x \in K$.

Assume $K \subset \mathbb{R}^n$ has $\text{diam}(K) = 1$. Translating K if necessary we may assume that $0 \in K$; then

$$\frac{1}{2} \le \sup_{x \in K} \|x\|_2 \le 1. \quad (6.1)$$

Also note that by assumption we have

$$m \ge C\delta^{-12} w(K - K) \ge C\delta^{-12} w(K). \quad (6.2)$$

Fix a large number $t \geq 2$ whose value will be chosen later and consider the set

$$K' = Q(K \oplus t) \subseteq S^n,$$

where $Q : \mathbb{R}^{n+1} \to S^n$ denotes the spherical projection map $Q(u) = u/\|u\|_2$. We have

$$
\begin{aligned}
w(K') &\leq t^{-1} w(K \oplus t) \quad \text{(as } \|u\|_2 \geq t \text{ for all } u \in K \oplus t\text{)} \\
&\leq t^{-1}\big(w(K) + t\mathbb{E}|\gamma|\big) \quad \big(\text{where } \gamma \sim \mathcal{N}(0, 1)\big) \\
&= t^{-1} w(K) + \sqrt{2/\pi} \leq 3w(K),
\end{aligned}
$$

where the last inequality holds because $w(K) \geq \sqrt{2/\pi} \sup_{x \in K} \|x\|_2 \geq 1/\sqrt{2\pi}$ by (6.1).

Then Theorem 1.2 implies that if $m \geq C\delta_0^{-6} w(K)^2$ for some $\delta_0 > 0$, then there exists an arrangement of m hyperplanes in $\mathbb{R}^{n+1}$ such that

$$\big|d_A(x', y') - d(x', y')\big| \leq \delta_0, \quad x', y' \in K'. \tag{6.3}$$

Consider arbitrary vectors x and y in K and the corresponding vectors $x' = Q(x \oplus t)$ and $y' = Q(x \oplus t)$ in K'. Let us relate the distances between x' and y' appearing in (6.3) to corresponding distances between x and y.

Let $a_i \oplus a \in \mathbb{R}^{n+1}$ denote normals of the hyperplanes. Clearly, x' and y' are separated by the ith hyperplane if and only if $x \oplus t$ and $y \oplus t$ are. This in turn happens if and only if x and y are separated by the affine hyperplane that consists of all $x \in \mathbb{R}^n$ satisfying $\langle a_i \oplus a, x \oplus t \rangle = \langle a_i, x \rangle + at = 0$. In other words, the hyperplane tessellation of K' induces an *affine* hyperplane tessellation of K, and the fraction $d_A(x', y')$ of the hyperplanes separating x' and y' equals the fraction of the affine hyperplanes separating x and y. With a slight abuse of notation, we express this observation as

$$d_A(x', y') = d_A(x, y). \tag{6.4}$$

Next, we analyze the normalized geodesic distance $d(x', y')$, which satisfies

$$\big|\pi \cdot d(x', y') - \|x' - y'\|_2\big| \leq C_0 \|x' - y'\|_2^2. \tag{6.5}$$

Denoting $t_x = \|x \oplus t\|_2$ and $t_y = \|y \oplus t\|_2$ and using the triangle inequality, we obtain

$$
\begin{aligned}
\varepsilon &:= \big|\,\|x' - y'\|_2 - t^{-1}\|x - y\|_2\,\big| \\
&= \big|\,\|t_x^{-1}(x \oplus t) - t_y^{-1}(y \oplus t)\|_2 - \|t^{-1}x - t^{-1}y\|_2\,\big| \\
&\leq \|x\|\,\big|t_x^{-1} - t^{-1}\big| + \|y\|\,\big|t_y^{-1} - t^{-1}\big| + t\,\big|t_x^{-1} - t_y^{-1}\big|. \tag{6.6}
\end{aligned}
$$

Note that (6.1) yields that $t \leq t_x, t_y \leq \sqrt{t^2 + 1}$. It follows that $|t_x^{-1} - t^{-1}| \leq 0.5 t^{-3}$ and the same bound holds for the other two similar terms in (6.6). Using this and (6.1),

we conclude that $\varepsilon \leq t^{-2}$. Putting this into (6.5) and using the triangle inequality twice, we obtain

$$\left| \pi \cdot d\big(x', y'\big) - t^{-1} \|x - y\|_2 \right| \leq C_0 \big(t^{-1} \|x' - y'\|_2 + \varepsilon\big)^2 + \varepsilon$$
$$\leq C_0 \big(2t^{-1} + t^{-2}\big)^2 + t^{-2} \leq C_1 t^{-2}.$$

Finally, we use this bound and (6.4) in (6.3), which gets us

$$\left| \pi t \cdot d_A(x, y) - \|x - y\|_2 \right| \leq \pi t \delta_0 + C_1 t^{-1}. \tag{6.7}$$

Now we can assign the values $t := 2C_1/\delta$ and $\delta_0 = \delta^2/(4\pi C_1)$ so the right-hand side of (6.7) is bounded by δ, as required. Note that the condition $m \geq C\delta_0^{-6} w(K)^2$ that we used above in order to apply Theorem 1.2 is satisfied by (6.2). This completes the proof of Theorem 1.10.

References

1. Ahmed, A., Ravi, S., Narayanamurthy, S., Smola, A.: Fastex: Hash clustering with exponential families. In: Advances in Neural Information Processing Systems, vol. 25 (2012)
2. Ailon, N., Chazelle, B.: The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. SIAM J. Comput. **39**, 302–322 (2009)
3. Alon, N., Spencer, J.: The Probabilistic Method, 2nd edn. Wiley, New York (2000)
4. Andoni, A., Indyk, P.: Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. In: 47th Annual IEEE Symposium on Foundations of Comp. Sci. (FOCS) (2006)
5. Boufounos, P.: Universal rate-efficient scalar quantization (2010). arXiv:1009.3145
6. Boufounos, P.T., Baraniuk, R.G.: 1-Bit compressive sensing. In: 42nd Annual Conference on Information Sciences and Systems (CISS) (2008)
7. Calka, P.: Tessellations. New Perspectives in Stochastic Geometry, pp. 145–169. Oxford University Press, Oxford (2010)
8. Charikar, M.: Similarity estimation techniques from rounding algorithms. In: Proceedings of the 34th Annual ACM Symposium on Theory of Computing (2002)
9. Giannopoulos, A., Milman, V.: Euclidean structure in finite dimensional normed spaces. In: Handbook of the Geometry of Banach Spaces, vol. I, pp. 707–779. North-Holland, Amsterdam (2001)
10. Goemans, M., Williamson, D.: Improved approximation algorithms for the maximum cut and satisfiability problems using semidefinite programming. J. ACM **42**, 1115–1145 (1995)
11. Jacques, L., Laska, J.N., Boufounos, P.T., Baraniuk, R.G.: Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. doi:10.1109/TIT.2012.2234823
12. Johnson, W., Lindenstrauss, J.: Extensions of Lipschitz mappings into a Hilbert space. Contemp. Math. **26**, 189–206 (1984)
13. Klartag, B., Mendelson, S.: Empirical processes and random projections. J. Funct. Anal. **225**, 229–245 (2005)
14. Kushilevitz, E., Ostrovsky, R., Rabani, Y.: Efficient search for approximate nearest neighbor in high dimensional spaces. SIAM J. Comput. **30**, 457–474 (2000)
15. Ledoux, M., Talagrand, M.: Probability in Banach Spaces. Isoperimetry and Processes. Springer, Berlin (1991)
16. Litvak, A.E., Milman, V.D., Pajor, A., Tomczak-Jaegermann, N.: On the Euclidean metric entropy of convex bodies. In: Geometric Aspects of Functional Analysis. Lecture Notes in Math., vol. 1910, pp. 221–235. Springer, Berlin (2007)
17. Matousek, J.: Lectures on Discrete Geometry. Springer, New York (2002)

18. Mendelson, S.: A few notes on statistical learning theory. In: Mendelson, S., Smola, A.J. (eds.) Advanced Lectures in Machine Learning. Lecture Notes in Computer Science, vol. 2600, pp. 1–40. Springer, Berlin (2003)
19. Plan, Y., Vershynin, R.: One-bit compressed sensing by linear programming (2011, submitted). arXiv: 1109.4299v4
20. Rudelson, M., Vershynin, R.: Sampling from large matrices: an approach through geometric functional analysis. J. ACM **21** (2007). 19 pp.
21. Schechtman, G.: Two observations regarding embedding subsets of Euclidean spaces in normed spaces. Adv. Math. **200**, 125–135 (2006)
22. Talagrand, M.: The Generic Chaining. Upper and Lower Bounds of Stochastic Processes. Springer, Berlin (2005)