HAL
open science

Lower bounds for k-distance approximation

Quentin Mérigot

To cite this version:

Quentin Mérigot. Lower bounds for k-distance approximation. SoCG '13 - 29th ACM Symposium on Computational Geometry, Jun 2013, Rio de Janeiro, Brazil. pp.435-440, 10.1145/2462356.2462367 . hal-00803410

HAL Id: hal-00803410

https://hal.science/hal-00803410

Submitted on 21 Mar 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Lower Bounds for k-Distance Approximation

Quentin Mérigot

March 21, 2013

Abstract

Consider a set P of N random points on the unit sphere of dimension $d-1$, and the symmetrized set $S = P \cup (-P)$. The *halving polyhedron* of S is defined as the convex hull of the set of centroids of N distinct points in S. We prove that after appropriate rescaling this halving polyhedron is Hausdorff close to the unit ball with high probability, as soon as the number of points grows like $\Omega(d \log(d))$. From this result, we deduce probabilistic lower bounds on the complexity of approximations of the distance to the empirical measure on the point set by distance-like functions.

1 Introduction

The notion of *distance to a measure* was introduced in order to extend existing geometric and topological inference results from the usual Hausdorff sampling condition to a more probabilistic model of noise [CCSM11]. Consider a finite subset P of the Euclidean space $\mathbb{R}^d$ and a positive number k in the range $\{1, \dots, |P|\}$, where $|P|$ denotes the cardinality of P. The distance to the empirical measure on P is given by the following formula:

$$\mathrm{d}_{P,k}(x) := \left(\frac{1}{k} \left[\min_{p_1,\dots,p_k \in P} \sum_{i=1}^{k} \|x - p_i\|^2 \right] \right)^{1/2}, \tag{1.1}$$

where the minimum is taken over the sets consisting of k distinct points in P. We will call this function the *k-distance to the point set P*. Equation (1.1) allows to compute the value of the k-distance at a certain point x easily, using a nearest-neighbor data structure. On the other hand, this formula cannot be used to perform more global computations, such as estimating the Betti numbers of a sublevel set $\mathrm{d}_{P,k}^{-1}(0,r) = \{x \in \mathbb{R}^d;\ \mathrm{d}_{P,k}(x) \leqslant r\}$, not to mention reconstructing a simplicial complex homotopic to this set.

There is another representation of the k-distance that allows to perform such global operations. It is computational geometry folklore that $\mathrm{d}_{P,k}$ can be rewritten as the square root the minimum of a finite number of quadratic functions. More precisely $\mathrm{d}_{P,k}(x)^2 = \min_{\overline{p}} \|x - \overline{p}\|^2 + w_{\overline{p}}$, where the minimum

is taken over the set of centroids of k distinct points in P, and $w_{\overline{p}}$ is chosen adequately (see §2.2). This implies that sublevel sets of $\mathrm{d}_{P,k}$ are simply union of balls, and this allows one to compute their homotopy type using weighted alpha-complexes or similar constructions [Ede92]. However, since the number of centroids of k distinct points in P grows exponentially with the number of points, this formulation is not very practical either.

Fortunately, many geometric and topological inference result continue to hold if one replaces $\mathrm{d}_{P,k}$ in the computation by a "good approximation" φ [CCSM11]. This means that the error $\|\varphi - \mathrm{d}_{P,k}\|_\infty := \max_{\mathbb{R}^d} |\varphi - \mathrm{d}_{P,k}|$ has to be small enough, and that the approximating function φ should be *distance-like*. For the purpose of this work, φ is distance-like if there exists a finite set of *sites* $Q \subseteq \mathbb{R}^d$ and non-negative *weights* $(w_q)_{q\in Q}$ such that $\varphi = \varphi_Q^w$, where φ_Q^w is defined by

$$\varphi_Q^w(x) := \left(\min_{q\in Q} \|x-q\|^2 + w_q\right)^{1/2}, \quad w_q \geqslant 0. \tag{1.2}$$

To summarize, in order to estimate the topology of the sublevel set of the k-distance, it makes sense to try to replace it by a distance-like function that uses much fewer sites.

Note that one could try to find approximations of the k-distance in a class of functions $\mathcal{F}$ different from the class of distance-like functions. It is indeed possible that a well chosen class of functions would produce more compact approximations of the k-distance and of similar functions. However, changing the class of function would practically forbid to use these approximations for the purpose of geometric inference, because of the lack of (i) computational topology tools to compute with the sublevel sets of functions in $\mathcal{F}$ and (ii) a geometric inference theory adapted to this class of function.

Complexity of k-distance The natural formalization of our approximation problem is as follows. Given a finite point set P in $\mathbb{R}^d$, a number $k > 0$ and a target approximation error ε, what is the minimum cardinality of a weighted point set (Q, w) with non-negative weights such that the approximation error $\left\|\varphi_Q^w - \mathrm{d}_{P,k}\right\|_\infty$ is bounded by ε ? We call this cardinality the ε-complexity of the k-distance function $\mathrm{d}_{P,k}$. When ε is zero, the 0-complexity of the k-distance function $\mathrm{d}_{P,k}$ is equal to the number of order-k Voronoi cell of P that have non-empty interior. One can then translate lower-bounds on the number of order-k Voronoi cells into lower bounds for the 0-complexity of the k-distance. The purpose of this article is to provide lower bounds on ε-complexity of the k-distance function for a *non-zero approximation error* ε, when P is a random point cloud on the unit $(d-1)$-dimensional sphere and $k = |P|/2$.

1.1 Prior work

Approximation of the k-distance The question of approximating the distance to the measure by a distance-like function with few sites has been originally raised in [GMM12]. In this article, the authors proposed an approximation of the k-distance, called the *witnessed k-distance* and denoted by $\mathrm{d}^{\mathrm{w}}_{P,k}$, which involves only a *linear* number of sites. They also give a probabilistic upper bound on the approximation error $\|\mathrm{d}^{\mathrm{w}}_{P,k} - \mathrm{d}_{P,k}\|_\infty$ under the hypothesis that the point cloud P is obtained by sampling a ℓ-dimensional submanifold of the Euclidean space. The upper bound on the approximation error degrade as the intrinsic dimension ℓ of the underlying submanifold increases. This suggests that approximating the distance to the uniform measure on a point cloud drawn from a high-dimensional submanifold might be difficult.

It is possible to build data structures that allow to compute approximate pointwise values of the k-distance in time that is logarithmic in the number of points — but exponential in the ambient dimension [HPK12]. The same data structure can be used to compute generalizations of the k-distance, such as the sum of the pth power to the k-nearest neighbors for an exponent p larger than one.

Complexity of order-k Voronoi As mentioned earlier, there exists upper and lower bounds for the number of cells in an order-k Voronoi diagrams, and those bounds can be translated into bounds on the number of sites that one needs to use in order to get an exact representation of the k-distance function by a distance-like function. When k is half the cardinality of the point cloud P, we will speak of *halving Voronoi diagram.* A *halving hyperplane* for P is a hyperplane that separates P into two sets with equal cardinality. The number of halving hyperplanes yielding different partitions of P is a lower bound on the number of infinite halving Voronoi cells. The best lower bound on the number of halving hyperplanes, that holds for an arbitrary ambient dimension d, is given by $N^{d-1}\mathrm{e}^{\Omega(\sqrt{\log N})}$, where N is the cardinality in the point set [Tót01]. This bound improves on previous lower bounds by several authors, e.g. [ELSS73, EW85, Sei87].

In [BS94], the authors study the expectation of the number of k-sets of a point cloud P obtained by sampling independent random point on the sphere. Recall that a k-set is a subset of k points in P that can be separated from other points in P by a hyperplane; in particular, each k-set corresponds to an infinite order-k Voronoi cell. The authors prove that the expected number of k-sets in a random point cloud P on the sphere is upper bounded by $\mathrm{O}(|P|^{d-1})$.

In order to obtain our lower bounds on the number of sites needed to approximate $\mathrm{d}_{P,k}$, we will use the notion of k-set polyhedron, originally introduced in [EVW97]. This polyhedron is the convex hull of the set of

centroids of k distinct points in P. The relation between this polyhedron and the notion of k-set is that the number of extreme points of the k-set polyhedron is equal to the number of k-sets in the point set. When k is half the cardinality of P, we will call this polyhedron the halving polyhedron.

1.2 Contributions

The main result of this work concerns the geometry of halving polyhedra of finite point sets on the unit sphere. Our theorem shows that even with relatively few points, the halving polyhedron of a certain random point set S on the unit $(d-1)$-dimensional sphere is Hausdorff-close to a ball with high probability. The random point set S is obtained by picking N random, independent and uniformly distributed points $p_1, \ldots, p_N$ on the $(d-1)$-dimensional unit sphere, and by letting $S = \{\pm p_i\}_{1 \leqslant i \leqslant N}$. The halving polyhedron of S is a by definition a random convex polyhedron, which we denote by L_N^d. The statement and proof of this theorem are inspired by the main theorem of [AAFM06]. We use the quantity

$$m_d := \mathbb{E}(|X \cdot u|) \simeq \left(\frac{2}{\pi d}\right)^{1/2},$$

where X is a uniformly distributed random vector on the unit $(d-1)$-sphere, and u is an arbitrary unit vector. Note that this quantity turns out not to depend on u.

Theorem 3.1. *There exists an absolute constants $c > 0$ such that for every positive number η, the inequality*

$$\mathrm{d_H}\left(\frac{1}{m_d} L_N^d, \mathcal{B}(0,1)\right) \leqslant \eta$$

holds with probability at least $1 - 2\exp\left[c \cdot \left(d\log(1/\delta) - N\eta^2\right)\right]$, where $\delta = \min(\eta, 1/\sqrt{d})$.

In Section 4, we deduce from this theorem a probabilistic lower bound on the ε-complexity of the $\mathrm{d}_{S,N}$, where S is the point set defined in the previous paragraph. The exact statement of the lower bound can be found in Theorem 4.1.

2 Traces at infinity and k-set polyhedra

Background: support function The support function of a convex subset K of $\mathbb{R}^d$ is a function $x \mapsto \mathrm{h}(K, x)$ from the unit sphere to $\mathbb{R}$. It is defined by the following formula:

$$\mathrm{h}(K, u) := \max\{x \cdot u;\ x \in K\}. \tag{2.3}$$

The application that maps a convex set to its support function on the sphere satisfies the following isometry property:

$$\|\mathrm{h}(K,.) - \mathrm{h}(L,.)\|_{\infty,\mathcal{S}^{d-1}} = \mathrm{d_H}(K,L). \tag{2.4}$$

where $\|f\|_{\infty,\mathcal{S}^{d-1}} = \max_{\mathcal{S}^{d-1}} |f|$ is the infinity norm on the unit sphere, and where $\mathrm{d_H}$ denotes the *Hausdorff distance*. The proof of this equality is given in Theorem 1.8.11 in [Sch93]. All the elementary facts about support functions that we will need can be found in the first chapter of this book.

2.1 Traces at infinity of distance-like functions.

We call *distance-like* a non-negative function whose square can be written as the minimum of a family of unit paraboloids $\|x-q\|^2 + w_q$, with $w_q \geqslant 0$. Note that this definition is equivalent to the one given in [CCSM11], thanks to the remark following Proposition 3.1 in this article.

Given a finite subset Q of the Euclidean space and non-negative weights w, we let $\varphi := \varphi_Q^w$ be the distance-like function defined by (1.2). We call *trace at infinity* of φ and denote by $K(\varphi)$ the convex polyhedra obtained by taking the convex hull of the set of sites Q used to define φ.

The name trace at infinity is explained by the following asymptotic development of the values of φ, along a unit-speed ray starting at the origin, i.e. $\gamma_t := tu$ with $\|u\| = 1$:

$$\begin{aligned}
\varphi(\gamma_t) &= \min_{q\in Q} \left(\|tu-q\|^2 + w_q\right)^{1/2} \\
&= \min_{q\in Q} t\left(1 - \frac{2}{t}u\cdot q + \frac{w_q + \|q\|^2}{t^2}\right)^{1/2} \\
&= t - \max_{q\in Q} u\cdot q + \mathrm{O}(1/t) && (2.5)\\
&= t - \mathrm{h}(K(\varphi),u) + \mathrm{O}(1/t) && (2.6)
\end{aligned}$$

The last equality follows from the fact that the support function of the convex hull of a set Q is given by $\max_{q\in Q} u\cdot q$. Using these computations, we get the following lemma. This lemma implies in particular that if two functions $\varphi := \varphi_Q^w$ and $\psi := \varphi_P^w$ coincide on $\mathbb{R}^d$, then $K(\varphi) = K(\psi)$.

Lemma 2.1. *Given two distance-like functions $\varphi := \varphi_Q^w$ and $\psi := \varphi_P^v$ as in* (1.2),

$$\mathrm{d_H}(K(\varphi), K(\psi)) \leqslant \|\varphi - \psi\|_\infty . \tag{2.7}$$

Proof. The asymptotic developments for φ and ψ given in equation (2.6) imply the following inequality for every unit direction u: $|\mathrm{h}(K(\varphi),u) - \mathrm{h}(K(\psi),u)| \leqslant \|\varphi-\psi\|_\infty$. With Equation (2.4), this yields the desired bound on the Hausdorff distance between $K(\varphi)$ and $K(\psi)$. □

The *power diagram* of a weighted point set (Q, w) is a decomposition of the space into convex polyhedra, one per point in Q, defined by

$$\mathrm{Pow}_Q^w(q) = \{x \in \mathbb{R}^d;\ \forall p \in Q,\ \ \|x-q\|^2 + w_q \leqslant \|x-p\|^2 + w_p\}.$$

The following lemma shows the relation between the vertices of the trace at infinity of φ_Q^w and unbounded cells in the power diagram of (Q, w).

LEMMA 2.2. *Consider a finite weighted point set (Q, w):*

(i) if the power cell of a point q in Q is unbounded, then q lies on the boundary of the polyhedron $K(\varphi_Q^w)$;

(i) conversely, if q is an extreme point in $K(\varphi_Q^w)$, then the power cell of q is unbounded.

Note that there might exist point q that lie on the boundary of the trace at infinity and whose power cell is empty. For instance, consider $Q = \{q_{-1}, q_0, q_1\}$ in $\mathbb{R}^2$ with $q_i = (0, i)$, and weights $w_{-1} = w_1 = 0$ and $w_0 > 1$. Then q_0 lies on the boundary of the convex hull while its power cell is empty.

Proof. Being convex, the power cell of a point q is unbounded if and only it contains a ray $\gamma_t := r + tu$, where u is a unit vector. By definition of the power cell, we have $\|q - \gamma_t\|^2 + w_q \leqslant \|p - \gamma_t\|^2 + w_p$ for all point p in the set Q. Expanding both sides expressions and simplifying, we get:

$$\|q-r\|^2 - \|p-r\|^2 + 2tu \cdot (p-q) + w_q - w_p \leqslant 0$$

This inequality holds for $t \to +\infty$, thus implying $u \cdot p \leqslant u \cdot q$. Thus, q lies on the boundary of $K(\varphi)$, and u is an exterior normal vector to $K(\varphi)$ at q.

Conversely, if q is an extreme point of the convex polyhedron $K(\varphi)$, there must exist a unit vector u, such that $u \cdot (p - q) \leqslant -\varepsilon < 0$ for any point p in Q distinct from q. Tracing back the above inequalities, and using the strict bound, one can show that for any point r, the ray $\gamma_t := r + tu$ belongs to the power cell of q for t large enough. □

2.2 Trace at infinity of the k-distance.

Given a set of points P in $\mathbb{R}^d$ and an integer k between one and $|P|$, the *k-distance* to P is defined by equation (1.1). The fact that this function is distance-like can be seen in the following equivalent formulation, whose proof can be found for instance in Proposition 3.1 of [GMM12]:

$$\mathrm{d}_{P,k}(x) = \left(\min_{\overline{p}} \sum_{i=1}^{k} \|x - \overline{p}\|^2 + w_{\overline{p}} \right)^{1/2}, \tag{2.8}$$

where the minimum is taken over the centroids $\overline{p}$ of k distinct points in P, i.e. $\overline{p} = \frac{1}{k}\sum_{1 \leqslant i \leqslant k} p_i$ and where the weight is given by $w_{\overline{p}} := \frac{1}{k}\sum_{1 \leqslant i \leqslant k} \|\overline{p} - p_i\|^2$.

Definition 2.1. We denote $K_k^d(P)$ the convex hull of the set of centroids of k distinct points in P. The support function of this polyhedron is given by the following formula:

$$\mathrm{h}(K_k^d(P), u) := \max_{p_1,\dots,p_k \in P} \frac{1}{k} \sum_{i=1}^{k} u \cdot p_i \tag{2.9}$$

This polyhedron has been introduced first in [EVW97] under the name of *k-set polyhedron* of the point set P. Equations (2.8)–(2.9) above imply that this polyhedron is the trace at infinity of the k-distance to P. Moreover, the number of infinite order-k Voronoi of P is at least equal to the number of extreme points in $K_k^d(P)$.

Halving distance When the number of points in P is even and k is equal to half this number, we rename the k-distance the *halving distance.* Similarly, we will refer to the order-k Voronoi diagram as the *halving Voronoi diagram* and to the k-set polyhedron as the *halving polyhedron.*

3 Approximating the sphere by halving polyhedra

In this section, we consider a family of random polyhedra constructed as halving polyhedron of symmetric point sets on the unit sphere. More precisely, we define:

Definition 3.1. Given a set P of N points on the unit $(d-1)$-sphere, we define $L_N^d(P)$ as the halving polyhedron of the symmetrization of P, i.e.

$$L_N^d(P) := K_N^d(P \cup (-P))$$

We obtain a (random) convex polyhedron $L_N^d := L_N^d(P)$, by letting P be an independent random sampling of N points on the unit $(d-1)$-sphere.

Our main theorem consists in a lower bound on the probability of the halving polyhedron L_N^d to be Hausdorff-close to a ball centered at the origin.

THEOREM 3.1. *There exists an absolute constants $c > 0$ such that for every positive number η, the inequality*

$$\mathrm{d_H}\left(\frac{1}{m_d} L_N^d, \mathcal{B}(0,1)\right) \leqslant \eta$$

holds with probability at least $1 - 2\exp\left[c \cdot \left(d\log(1/\delta) - N\eta^2\right)\right]$, *where* $\delta = \min(\eta, 1/\sqrt{d})$.

The proof of this theorem is postponed to Section 5. As a first corollary, one can show that the random polyhedron L_N^d is approximately round with high probability as soon as the cardinality of the set of points sampled on the sphere grows faster than $d \log d$, where d is the ambient dimension.

Corollary 3.2. *For any $\kappa > 0$, there is a constant C_κ such that for $\eta > 0$, $d \geqslant 1/\eta^2$ and $N \geqslant \frac{d}{\eta^2}(\log(d) + C_\kappa)$ the following inequality*

$$\mathrm{d_H}\left(\frac{1}{m_d} L_N^d, \mathcal{B}(0,1)\right) \leqslant \eta$$

holds with probability at least $1 - \exp(-\kappa d)$.

Proof. From Theorem 3.1 the probability bound in the statement holds if $-\kappa d \geqslant c(d\log(1/\delta) - N\eta^2)$. Since δ is equal to $1/\sqrt{d}$, this is the case if $N\eta^2 \geqslant d\left(\log(d) + \frac{2\kappa}{c}\right)$. □

4 Application: approximation of distance-to-measures

The results of the previous section can be used to obtain a probabilistic statement on the complexity of the halving distance to a random point set on a high-dimensional sphere. We call *ε-complexity* of a distance-like function $\varphi : \mathbb{R}^d \to \mathbb{R}$ the minimum number of sites that one needs in order to be able to construct a distance-like function ψ such that the infinity norm $\|\varphi - \psi\|_\infty$ is at most ε, i.e.

$$\mathcal{N}(\varphi, \varepsilon) := \min\left\{|Q|\,;\ \left\|\varphi - \varphi_Q^w\right\|_\infty \leqslant \varepsilon,\ \varphi_Q^w \text{ as in (1.2)}\right\}.$$

The following theorem provides a probabilistic lower bound on the ε-complexity of a family of distance-like functions.

Theorem 4.1. *For any constant $\kappa > 0$ there exists a constant $C(\kappa)$ such that the following hold. Let $\eta > 0$, $d \geqslant 1/\eta^2$, and S be the symmetrization of an random point cloud of cardinality $N = \frac{d}{\eta^2}(\log(d) + C_\kappa)$ on the unit sphere. Then, the inequality*

$$\mathcal{N}\left(\mathrm{d}_{S,N}, m_d\eta\right) \geqslant 2\sqrt{d}\left(\frac{N}{64d(\log(d) + C_\kappa)}\right)^{\frac{d-1}{4}}$$

holds with probability at least $1 - \exp(-\kappa d)$.

Taking $\kappa = 1$ and $d = 1/\eta^2$, this implies:

Corollary 4.2. *There exists a sequence of point clouds S_d of cardinality $d^2(\log(d) + C_1)$ on the sphere $\mathcal{S}^{d-1}$ such that*

$$\mathcal{N}\left(\mathrm{d}_{S_d, \frac{1}{2}|S_d|}, \frac{\sqrt{2/\pi}}{d}\right) \geqslant 2\sqrt{d}\left(\frac{d}{64}\right)^{\frac{d-1}{4}}.$$

Proofof Theorem 4.1. The halving polyhedron $L_N^d(S)$ is by definition equal to the trace at infinity of the halving distance $K(\mathrm{d}_{S,N})$. Therefore, we can apply Corollary 3.2: there exist a constant C_κ such that for a random point set S distributed as in the statement of the theorem, the following inequality holds:

$$\mathrm{d_H}\left(\frac{1}{m_d}K(\mathrm{d}_{S,N}), \mathrm{B}(0,1)\right) \leqslant \eta. \tag{4.10}$$

with probability at least $1 - \exp(-\kappa d)$. We now consider a deterministic point set S on the sphere that satisfies the above inequality, and we consider a distance-like function $\psi : x \mapsto \min_{q\in Q} \|x-q\|^2 + w_q$ such that the approximation error $\|\psi - \mathrm{d}_{S,N}\|_\infty$ is bounded by $m_d\eta$. By definition of the complexity of $\mathrm{d}_{S,N}$, our goal is to prove a lower bound on the cardinality of the point set Q. Using the triangular inequality for the Hausdorff distance first, and then Lemma 2.1.(ii), we get the following inequalities:

$$\begin{aligned}
&\mathrm{d_H}\left(\frac{1}{m_d}K(\psi), \mathrm{B}(0,1)\right)\\
&\quad\leqslant \frac{1}{m_d}\mathrm{d_H}(K(\psi), K(\mathrm{d}_{S,N})) + \mathrm{d_H}\left(\frac{1}{m_d}K(\mathrm{d}_{S,N}), \mathrm{B}(0,1)\right)\\
&\quad\leqslant \frac{1}{m_d}\|\psi - \mathrm{d}_{S,N}\|_\infty + \eta \leqslant 2\eta
\end{aligned}$$

Now, recall that $K(\psi)$ is equal to the convex hull of the point set Q. If we define R as the rescaled set $\{m_d^{-1}q;\ q \in Q\}$, the above inequality reads $\mathrm{d_H}(\mathrm{conv}(R), \mathrm{B}(0,1)) \leqslant 2\eta$.

We have thus constructed a polyhedron that is within Hausdorff distance 2η of the d-dimensional unit ball. The last remark of [BI75] gives the following lower bound on the number of vertices of such a polyhedron:

$$|Q| = |R| \geqslant 2\sqrt{d}(8\eta)^{-\frac{d-1}{2}}$$

To conclude the proof, we simply replace η by its expression in term of the number of points N and the dimension d, i.e. $\eta^2 := d(\log(d) + C_\kappa)/N$ to obtain

$$|Q| \geqslant 2\sqrt{d}\left(\frac{N}{64d(\log(d)+C_\kappa)}\right)^{\frac{d-1}{4}}$$

□

5 Proof of the main theorem

We start this section by showing a simple expression for the support function of the halving polyhedra of a symmetric point set. For a point set P, and $N := |P|$, one has:

$$h(L_N^d(P), u) = \sum_{p\in P} |p \cdot u|\,. \tag{5.11}$$

Proofof Equation (5.11). Let $S = P \cup \{-P\}$ and recall that by definition, the support function $\mathrm{h}(L_N^d(P), u)$ is equal to $\max \sum_{i=1}^N u \cdot p_i$, where the maximum is taken on sets of N distinct points in S. Choosing $\varepsilon(p) = \pm 1$ such that for every point in p in P, $\varepsilon(p) u \cdot p \geqslant 0$, one easily sees that this maximum is attained for $\{p_1, \ldots, p_N\} = \{\varepsilon(p)p; p \in P\}$. □

This computation motivates the following definition.

Definition 5.1. We let $\lambda_{1,u}^d$ be the measure on $[0,1]$ given by the distribution of $|u \cdot X|$ where X is a uniformly distributed random vector on the $(d-1)$-dimensional unit sphere, and u is a fixed unit vector. This measure turns out not to depend on u, so we will denote it by λ_1^d.

The random polyhedron L_N^d is defined as $L_N^d(P)$ where the point set P is obtained by drawing N independent points on the unit sphere. Equation (5.11) implies that for any unit vector u the distribution of values of $\mathrm{h}(L_N^d, u)$ is given by the formula:

$$\frac{1}{N} \sum_{i=1}^N Y_i, \tag{5.12}$$

where the (Y_i) are N independent random variables with distribution λ_1^d.

Lemma 5.1. *The measure λ_1^d has the following properties:*

(i) λ_1^d is absolutely continuous with respect to the Lebesgue measure, with density

$$f_d(t) := c_d (1-t^2)^{\frac{d-2}{2}}$$

the constant c_d being chosen so that f_d is the density of a probability measure, i.e. $c_d \int_0^1 (1-t^2)^{\frac{d-2}{2}} \mathrm{d}t = 1$.

(ii) the mean of λ_1^d is given by $m_d := \frac{c_d}{d}$. Moreover, m_d is equivalent to $\sqrt{2/(\pi d)}$ as $d \to \infty$.

(iii) The variance σ_d^2 of λ_1^d is equivalent to $(1-2/\pi)/d$ as $d \to \infty$.

Proof. (i) Using Pythagoras theorem, one checks that the intersection of the hyperplane $\{x \in \mathbb{R}^d;\ u \cdot x = t\}$ with the unit sphere is a $(d-2)$-dimensional sphere with squared radius $(1-t^2)$. This implies the formula for f_d, for a certain constant c_d. To compute this constant one uses the fact that λ_1^d has unit mass, i.e.

$$c_d \int_0^1 (1-t^2)^{\frac{d-2}{2}} \mathrm{d}t = 1.$$

Note that a formula of Wallis asserts that

$$\lim_{d \to \infty} \sqrt{d} \int_0^1 (1-t^2)^{\frac{d}{2}} = \sqrt{\pi/2}.$$

This implies that $c_d \sim \sqrt{2d/\pi}$.
(ii) The function $t \mapsto t f_d(t)$ admits an explicit primitive:

$$g_d(t) := -\frac{c_d}{d}(1-t^2)^{\frac{d}{2}},$$

so that $m_d = g_d(1) - g_d(0)$ is as in the statement of the lemma. Thus, we get $m_d \simeq \sqrt{2/(\pi d)}$.
(iii) Integrating the following inequality between 0 and 1

$$t^2(1-t^2)^{\frac{d-2}{2}} = (1-t^2)^{\frac{d-2}{2}} - (1-t^2)(1-t^2)^{\frac{d-2}{2}},$$

one gets the following formula for the second moment of λ_1^d:

$$c_d \int_0^1 t^2(1-t^2)^{\frac{d-2}{2}} \mathrm{d}t = 1 - \frac{c_d}{c_{d+2}}.$$

Moreover, an integration by part gives

$$c_d \int_0^1 t^2(1-t^2)^{\frac{d-2}{2}} \mathrm{d}t = \frac{c_d}{d} \int_0^1 (1-t^2)^{\frac{d}{2}} \mathrm{d}t = \frac{1}{d}\frac{c_d}{c_{d+2}}$$

These two equalities imply that $c_d/c_{d+2} = d/(d+1)$, and using the formula for the mean given above, we get:

$$\sigma_d^2 = \frac{1}{d} - \frac{c_d^2}{d^2} \sim \frac{1-2/\pi}{d}.$$

□

Lemma 5.2. *There exists a universal constant $c > 0$ such that for any dimension d, any $N > 0$ and any set of directions U in $\mathcal{S}^{d-1}$, one has*

$$\mathbb{P}\left(\max_{u \in U} \left|\mathrm{h}(L_N^d, u) - m_d\right| \geqslant \eta m_d\right) \leqslant 2\,|U| \exp\left(-c \cdot N\eta^2\right).$$

Proof. Consider N random variables $Y_1, \ldots, Y_N$ with distribution λ_1^d. These random variable are bounded by 1 and their variance is σ_d^2. Applying Bernstein's inequality gives:

$$\mathbb{P}\left(\left|\frac{1}{N}\sum_{i=1}^N Y_i - m_d\right| \geqslant \varepsilon\right) \leqslant 2\exp\left(\frac{-N\varepsilon^2}{2\sigma_d^2 + 2\varepsilon/3}\right)$$

This implies that for a fixed direction u,

$$\mathbb{P}\left(\left|\mathrm{h}(L_N^d, u) - m_d\right| \geqslant \eta m_d\right) \leqslant 2\exp\left(\frac{-N\eta^2 m_d^2}{2\sigma_d^2 + 2\eta m_d/3}\right)$$

Rescaling everything by a certain constant $\kappa > 0$, one gets

$$\begin{aligned}\mathbb{P}\left(\left|\mathrm{h}(L_N^d, u) - m_d\right| \geqslant \eta m_d\right) &= \mathbb{P}\left(\left|\mathrm{h}(\kappa L_N^d, u) - \kappa m_d\right| \geqslant \kappa\eta m_d\right) \\ &\leqslant 2\exp\left(\frac{-N\kappa^2\eta^2 m_d^2}{2\kappa^2\sigma_d^2 + 2\eta\kappa m_d/3}\right)\end{aligned}$$

Letting κ go to infinity, we obtain the following bound

$$\mathbb{P}\left(\left|\mathrm{h}(L_N^d, u) - m_d\right| \geqslant \eta m_d\right) \leqslant 2\exp\left(-\frac{1}{2}N\eta^2\frac{m_d^2}{\sigma_d^2}\right)$$

This gives the desired estimate for a single direction using the two estimates $m_d = \mathrm{O}(d^{-1/2})$ and $\sigma_d^2 = \mathrm{O}(d^{-1})$. The conclusion of the lemma is obtained by a simple application of the union bound. □

Lemma 5.3. *If K is contained in the ball $\mathrm{B}(0,r)$, the support function $\mathrm{h}(K,.)$ is r-Lipschitz.*

Proof. Consider u in the unit sphere, and x in K such that $\mathrm{h}(K,u) = u \cdot x$. For any vector v in the unit sphere,

$$\begin{aligned}\mathrm{h}(K,v) = \max_{y\in K} v\cdot y \geqslant v\cdot x &= u\cdot x + (v-u)\cdot x \\ &\geqslant \mathrm{h}(K,u) - \|u-v\|\,\|x\| \\ &\geqslant \mathrm{h}(K,u) - r\,\|u-v\|\,.\end{aligned}$$

Swapping u and v gives the Lipschitz bound. □

A subset U of the unit sphere $\mathcal{S}^{d-1}$ is called a *δ-sample* or a δ-covering if the union of the Euclidean balls of radius δ centered at points of U cover the unit sphere.

Lemma 5.4. *Consider a convex set K contained in the unit ball $\mathcal{B}(0,1)$, and two numbers $\lambda, \eta \in (0,1)$. Moreover, suppose*

$$\max_{u\in U} |\mathrm{h}(K,u) - \lambda| \leqslant \eta\lambda, \tag{5.13}$$

where U is a δ-sample of the unit sphere, with $\delta := \min(\lambda,\eta)$. Then, the Hausdorff distance between $\frac{1}{\lambda}K$ and the unit ball is at most 5η.

Proof. Note that, almost by definition, a convex set K is included in the ball $\mathcal{B}(0,r)$ if and only if its support function satisfies $\|\mathrm{h}(K,.)\|_\infty \leqslant r$. Assuming that the convex set K is contained in some ball $\mathcal{B}(0,r)$, we have:

$$\begin{aligned}\|\mathrm{h}(K,.) - \lambda\|_\infty &= \max_{v\in\mathcal{S}^{d-1}} |\mathrm{h}(K,v) - \lambda| \\ &\leqslant \max_{v\in\mathcal{S}^{d-1}} \left[\min_{u\in U} |\mathrm{h}(K,u) - \mathrm{h}(K,v)| + |\mathrm{h}(K,u) - \lambda|\right] \\ &\leqslant r\min(\lambda,\eta) + \eta\lambda. \end{aligned}\tag{5.14}$$

The first inequality is obtained by applying the triangle inequality, while the second one follows from the Lipschitz estimation of Lemma 5.3, the fact that U is a δ-sample of the unit sphere and from Equation 5.13. Applying

Inequality (5.14) with $r = 1$, $\eta < 1$ and using the triangle inequality we get $\|\mathrm{h}(K,.) - \lambda\|_\infty \leqslant 2\lambda$. This implies that K is contained in the sphere $\mathcal{B}(0, 3\lambda)$ and allows us to apply the same inequality (5.14) again with the smaller radius $r = 3\lambda$, implying

$$\|\mathrm{h}(K,.) - \lambda\|_\infty \leqslant 3\lambda\eta + 2\lambda\eta = 5\lambda\eta.$$

Dividing this last inequality by λ, and using Equation (2.4) implies the conclusion of the Lemma. □

Proof of Theorem 3.1 Consider $\delta = \min(\eta, m_d)$, and let U be a δ-sample of the unit sphere with minimal cardinality. The cardinality of such a sample is bounded by $|U| \leqslant (\mathrm{const}/\delta)^d$, where the constant is absolute. Applying Lemma 5.4 first and then Lemma 5.2 gives us the following inequalities:

$$\begin{aligned}\mathbb{P}\left(\mathrm{d_H}\left(\frac{1}{m_d}K_N^d, \mathcal{B}(0,1)\right) \geqslant 5\eta\right) &\leqslant \mathbb{P}\left(\max_{u\in U}\left|\mathrm{h}(L_N^d, u) - m_d\right| \geqslant \eta m_d\right)\\ &\leqslant 2\,|U|\exp\left(-c\cdot N\eta^2\right).\end{aligned}$$

We conclude the proof by applying the upper bound on the cardinality of the δ-sample U stated above, and by using the equivalent $m_d \sim \sqrt{2/(\pi d)}$.

6 Extension to other values of $\frac{k}{N}$

It is possible to extend some of the results to cases where ratio k/N is different from one half. Let us consider the random polytope $M_{N,k}^d = K_k^d(P)$, where P is a set of N random points sampled uniformly and independently on the $(d-1)$-dimensional unit sphere, and k is between 1 and N. Lemma 5.2 can be partially extended to this more general family of random polytopes. Note however that the statement below does not provide any estimate for the radius $r(d, N, k)$ as the number of points N grows to infinity with k/N remaining constant. In particular, it is not precise enough to generalize the lower bounds of Theorem 3.1.

Lemma 6.1. *For any dimension d, and any numbers N and k, there exist a value $r := r(d, N, k)$ such that for any set of directions U in $\mathcal{S}^{d-1}$, and $N > 0$ one has*

$$\mathbb{P}\left(\max_{u\in U}\left|\mathrm{h}(M_{N,k}^d, u) - r\right| \geqslant \eta r\right) \leqslant 2\,|U|\exp\left(-\frac{1}{4}\frac{k^2}{N}\eta^2 r^2\right).$$

This lemma follows from a version of Chernoff's inequality adapted to Lipschitz functions. Consider a function F from the cube $[-1,1]^N$ to $\mathbb{R}$,

which is α-Lipschitz with respect to the ℓ^1 norm on the cube. Then, for any family $Y_1, \ldots, Y_N$ of i.i.d. random variables taking values in $[-1,1]$ one has:

$$\mathbb{P}(|F(Y_1, \ldots, Y_N) - \mathbb{E}F| \geqslant \varepsilon) \leqslant 2\exp\left(\frac{-\varepsilon^2}{4\alpha^2 N}\right). \tag{6.15}$$

Proofof Lemma 6.1. We consider the map F_N^k from the cube $[-1,1]^N$ to $\mathbb{R}$ defined by:

$$F_N^k(x) := \max\left\{\frac{1}{k}\sum_{j=1}^{k} x_{i_j};\ 1 \leqslant i_1 < \ldots < i_k \leqslant N\right\},$$

where x_i denotes the ith coordinate of x. We also consider the measure μ obtained by pushing forward the $(d-1)$-area measure on the unit sphere of $\mathbb{R}^d$ by the map $x \mapsto u \cdot x$, for some direction u in the unit sphere. As in the halving case (cf (5.12)), the distribution of $\mathrm{h}(M_{N,k}^d, u)$ is given by the distribution of $F_N^k(Y_1, \ldots, Y_N)$, where $Y_1, \ldots Y_N$ are i.i.d random variables with distribution μ.

Moreover, the map F_N^k is $\frac{1}{k}$-Lipschitz with respect to the ℓ^1 norm on the cube $[-1,1]^N$. Indeed, given two points x, y in $[-1,1]^N$ and a set of indices $i_1 < \ldots < i_k$ corresponding to the maximum in the definition of $F_N^k(x)$, one has

$$\begin{aligned} kF_N^k(x) = x_{i_1} + \ldots + x_{i_k} &\leqslant y_{i_1} + \ldots + y_{i_k} + \|x-y\|_{\ell^1} \\ &\leqslant kF_N^k(y) + \|x-y\|_{\ell^1} \end{aligned}$$

Thus, we can apply Chernoff's inequality (6.15):

$$\mathbb{P}\left(\left|F_N^k(Y_1, \ldots, Y_N) - \mathbb{E}F_N^k\right| \geqslant \varepsilon\right) \leqslant 2\exp\left(-\frac{\varepsilon^2 k^2}{4N}\right)$$

where $m = k/N$. The Lemma follows by setting $r := \mathbb{E}F_N^k$, $\varepsilon = \eta r$ in the equation and by using the union bound. □

Acknowledgements This work was partially funded by French ANR grant GIGA ANR-09-BLAN-0331-01.

References

[AAFM06] S. Artstein-Avidan, O. Friedland, and V. Milman, *Geometric applications of Chernoff-type estimates and a zigzag approximation for balls*, Proceedings of the American Mathematical Society **134** (2006), no. 6, 1735–1742.

[BI75] EM Bronshteyn and LD Ivanov, *The approximation of convex sets by polyhedra*, Siberian Mathematical Journal **16** (1975), no. 5, 852–853.

[BS94] I. Bárány and W. Steiger, *On the expected number of k-sets*, Discrete & Computational Geometry **11** (1994), no. 1, 243–263.

[CCSM11] F. Chazal, D. Cohen-Steiner, and Q. Mérigot, *Geometric inference for probability measures*, Foundations of Computational Mathematics **11** (2011), 733–751.

[Ede92] Herbert Edelsbrunner, *Weighted alpha-shapes*, Tech. Report UIUCDCS-R-92-1760, Department of Computer Science, University of Illinois, 1992.

[ELSS73] P. Erdős, L. Lovász, A. Simmons, and EG Straus, *Dissection graphs of planar point sets*, 1973, pp. 139–154.

[EVW97] H. Edelsbrunner, P. Valtr, and E. Welzl, *Cutting dense point sets in half*, Discrete & Computational Geometry **17** (1997), no. 3, 243–255.

[EW85] H. Edelsbrunner and E. Welzl, *On the number of line separations of a finite set in the plane*, Journal of Combinatorial Theory, Series A **38** (1985), no. 1, 15–29.

[GMM12] L. Guibas, Q. Mérigot, and D. Morozov, *Witnessed k-distance*, Discrete and Computational Geometry (2012), To appear.

[HPK12] Sariel Har-Peled and Nirman Kumar, *Down the rabbit hole: Robust proximity search and density estimation in sublinear space*, 53rd Annual IEEE Symposium on Foundations of Computer Science, 2012, 2012, pp. 430–439.

[Sch93] Rolf Schneider, *Convex bodies: the brunn-minkowski theory*, vol. 44, Cambridge University Press, 1993.

[Sei87] R. Seidel, Personal communication. Reported in H. Edelsbrunner, Algorithms in Combinatorial Geometry, Springer-Verlag, 1987.

[Tót01] G. Tóth, *Point sets with many k-sets*, Discrete & Computational Geometry **26** (2001), no. 2, 187–194.